

AAPOR

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

Abstract Book



76th

Annual Conference

**Data Collection, Measurement,
and Public Opinion During
a Pandemic**

May 11-14, 2021

www.aapor.org #aapor   

Table of Contents

Tuesday, May 11, 2021 Day 1 of Conference	2
5/11/2021, 11:30 AM - 1:30 PM	2
5/11/2021, 12:00 PM - 1:30 PM	24
5/11/2021, 2:00 PM - 3:00 PM	34
5/11/2021, 2:30 PM - 4:30 PM	35
5/11/2021, 3:00 PM - 4:30 PM	59
Wednesday, May 12, 2021 Day 2 of Conference	68
5/12/2021, 10:00 AM - 11:00 AM	68
5/12/2021, 10:30 AM - 12:30 PM	70
5/12/2021, 2:00 PM - 4:00 PM	101
5/12/2021, 2:30 PM - 4:00 PM	127
5/12/2021, 4:00 PM - 5:00 PM	135
Thursday, May 13, 2021 Day 3 of Conference	138
5/13/2021, 9:30 AM - 11:30 AM	138
5/13/2021, 10:00 AM - 11:30 AM	161
5/13/2021, 1:00 PM - 3:00 PM	170
5/13/2021, 1:30 PM - 3:00 PM	196
5/13/2021, 3:00 PM - 5:00 PM	202
5/13/2021, 3:30 PM - 5:00 PM	224
Friday, May 14, 2021 Day 4 of Conference	232
5/14/2021, 11:30 AM - 1:30 PM	232
5/14/2021, 12:00 PM - 1:30 PM	254
5/14/2021, 1:30 PM - 3:30 PM	261
5/14/2021, 2:00 PM - 3:30 PM	281

Tuesday, May 11, 2021 | Day 1 of Conference

5/11/2021, 11:30 AM - 1:30 PM

Concurrent Session A: Prerecorded Sessions

Mini-Conference: American Community Survey Data Collection Challenges During the 2020 Census and the Coronavirus Pandemic

American Community Survey Data Collection Challenges During the 2020 Census and the Coronavirus Pandemic

Dorothy Barth, *U.S. Census Bureau*

2020 was a challenging year for data collection for the American Community Survey. This panel will outline the challenges and discuss how the Census Bureau adjusted to obtain survey response despite the obstacles. The presentations will discuss the changes made to the self-response mail materials due to the 2020 Census and the coronavirus pandemic. The panel will also include a presentation on how response to the survey was affected throughout the year.

Data Collection for the American Community Survey During the 2020 Pandemic

Dameka Reese, *U.S. Census Bureau*

During the 2020 American Community Survey (ACS) data collection period, the coronavirus pandemic led to mandatory local stay-at-home-orders across the United States and Puerto Rico. These stay-at-home orders created substantial disruptions resulting in radical changes to the ACS data collection operations in 2020. Due to the swift change in our environment, the ACS program revised messaging, altered mailout strategies, and made significant sampling adjustments to accommodate staffing limitations at the assembly and mailout facilities. The computer-assisted personal interview (CAPI) operation did not experience any downtime, but the operation was limited to conducting only telephone interviews for a few months. The decision to conduct telephone only interviews greatly impacted the Group Quarters operation which relies on personal visit to collect data. This presentation addresses how the ACS modified the data collection operations during the rapidly changing environment.

Response to the American Community Survey During the 2020 Pandemic

Stephanie Baumgardner, *U.S. Census Bureau*

The COVID-19 global pandemic significantly altered data collection for the 2020 American Community Survey (ACS). The ACS typically sends up to five mailings to sampled addresses but, due to mandatory stay-at-home orders, the Census Bureau's National Processing Center (NPC) was not able to send out any mailings from mid-March through June. In July, with reduced staff available at NPC, mailouts started again but could only send up to two mailings to sample addresses. As resources allowed, three mailings were sent to sample addresses starting in October 2020. The stay-at-home orders also affected the computer-assisted personal interview (CAPI) operation, which restricted field representatives to complete interviews by only telephone for a few months. Gradually, areas opened for in-person interviewing and by September, all areas allowed in-person interviewing. This presentation will show how these changes to data collection affected self-response and CAPI response to the ACS in 2020.

Designing New Self-Response Letters for the American Community Survey During the Coronavirus Pandemic

Dorothy Barth, *U.S. Census Bureau*

The American Community Survey (ACS) sends up to five mailings to each sampled address during the self-response phase each month: two packages, two pressure seal mailers, and one postcard. During the height of the pandemic, from mid-March to June, we were not able to send out any mailings due to the shutdown. By July, we were ready to start sending out mailings on a modified schedule, with modified

mailings. Because we had little time and reduced staff to assemble the packages, we chose to replace some packages with newly designed pressure seal mailers. We also redesigned the final reminder pressure seal letter to reflect the changes to our data collection strategy during the pandemic. In the designs, we were able to incorporate lessons learned from recent cognitive testing and field-testing of ACS mail materials. This presentation will include a detailed look at the redesigned letters and a brief overview of ACS self-response in 2020.

Findings From Qualitative Research on the ACS 2020 Specialized Mail Materials

Karen Stein, *Westat*

Martha Stapleton, *Westat*

Sarah Heimel, *U.S. Census Bureau*

Lindsay Longsine, *U.S. Census Bureau*

During the 2010 Decennial Census, ACS response rates were higher than usual in the first few months of the year but were lower than usual in the spring and summer months. The increase early in the year may have occurred because the decennial census communications campaign brought attention to the U.S. Census Bureau. The decrease later in the year may have been because recipients thought the ACS was the 2020 Census and disregarded the ACS questionnaire. To mitigate these issues in 2020, the Census Bureau developed a set of modified mail materials, for the first three mailings, with language that directly addressed the difference between the ACS and the 2020 Census. Westat conducted interviews with over three dozen ACS self-respondents to assess recipients' reactions to receiving both the ACS and the 2020 Census and their impressions of the revised messaging.

The ACS initial mailing, pressure seal reminder letter, and questionnaire package were re-designed following recommendations from an evaluation conducted by Chesnut and Davis (2011) after the 2010 Census. Beginning in March, the ACS mail materials included references to the 2020 Census and how it was different from the ACS. With the emergence of the novel coronavirus (COVID-19) in February 2020 and its unprecedented impact on Census Bureau operations, modifications were made to the ACS contact strategy, including reducing the number of contacts respondents received. In addition, the cognitive testing mode was converted from in-person to telephone.

This paper will describe the methodological issues with adapting cognitive interviews to pandemic restrictions and present key findings from the interviews. Initial analysis shows that respondents felt they were getting a lot of mail from the Census Bureau, were unable to readily distinguish ACS from decennial mail, and were most persuaded by the statement that their response is required by law.

Impact of Revised Mail Materials on Self-Response to the American Community Survey During the 2020 Census

Samantha Spiers, *U.S. Census Bureau*

Sarah Heimel, *U.S. Census Bureau*

The U.S. Census Bureau contacts a sample of housing units by mail to obtain their response to the American Community Survey (ACS). Historically, ACS response rates have decreased in the months when the decennial census is also collecting data. For 2020, the Census Bureau developed revised ACS mail materials; the goal was to distinguish the ACS from the 2020 Census and mitigate negative impacts on response rates. We will discuss whether the revised wording was beneficial in obtaining response from addresses in the ACS sample and also compare the timing of when these addresses responded to the ACS and to the 2020 Census.

International Public Opinion Surveys

Social Values in Citizens' Deliberation on Solar Futures: Comparing Deliberative Pollings in Tokyo and Hong Kong

Victor Lam, *University of British Columbia*

Daphne Ngar-Yin Mah, *Hong Kong Baptist University*

Darren Man-wai Cheung, *Hong Kong Baptist University*

Alice Siu, *Stanford University*

Benjamin McLellan, *Kyoto University*

Shinya Wakao, *College of the Mainland*

Catherine Luo, *Stanford University*

Smart energy transition studies have drawn attention to the limits of government-led, top-down approaches and have called for citizen-centric transitions that explore the roles of citizens in contributing to deep decarbonization goals. Yet the literature has been limited in addressing the social aspects of energy transitions from the perspective of citizens. This paper explores the contribution of social values and demographic factors to public acceptance of solar futures scenarios in two global cities in East Asia, Tokyo and Hong Kong. Based on qualitative and quantitative data on participants from the Tokyo (n = 136) and Hong Kong (n = 174) online Deliberative Polls (DPs), preliminary findings suggests that participants in both DPs highly supported moderate solar scenarios before and after deliberation, followed by increasing support for business-as-usual solar scenarios and decreasing support for ambitious solar scenarios. Furthermore, public acceptance of moderate and ambitious solar scenarios was linked with support for energy autonomy and one's willingness to get involved in energy transitions, as well as the perceptions of one's capabilities to get involved, feasibility of solar in the future, and technical viability of solar given adequate infrastructure support. Moreover, demographic characteristics such as household size, age group, and household electricity expenses affected public acceptance to the promotion of solar energy technology.

Afghan Futures: Peace Negotiations With the Taliban and Beyond

Matthew Warshaw, *D3: Designs, Data, Decisions*

Through the presentation of data from our on-going "Afghan Futures" polling series, we propose to present a current view of conditions in Afghanistan related to the peace process with the Taliban and conditions for average Afghans related to future of their country. We will look at the results from our recent national surveys and discuss what comes next for Afghanistan.

ACSOR Surveys and D3: Designs, Data, Decisions will present survey data from our most national polls in Afghanistan, to examine the above mentioned issues as well as other topics such as migration and the economy in Afghanistan. The goal of our "Afghan Futures" project is to continue to enrich public understanding of Afghanistan, as well as to share insights into the methodological and operational challenges of conflict-zone research. The project began in 2010 and includes results from nine national face-to-face surveys. We will have new data from a wave of the project to present at AAPOR. Interviews are conducted in Dari and Pashto among a random national sample of an average of 2,000 respondents in each wave covering all thirty-four provinces of Afghanistan. Examples of our work are available at www.acsor-surveys.com.

A Mind Is Not Closed for No Reason Why Do Some Chinese Dispute the Apollo Moon Landing?

Liwei Shen, *University of Wisconsin-Madison*

This study uses the Chinese Netizen Survey data (N = 2,379) to explore how nationalism and authoritarianism orientations, media use, and attitudes toward the US among Chinese netizens are related to their belief in a conspiracy theory that disputes a well-known historical fact: Apollo Moon Landing. We found that stronger nationalist sentiment and authoritarian orientation predict individuals' belief that Moon Landing is a hoax and the relationships are mediated by their attitudes toward the US. In addition, the need for closure is found to strengthen the relationship between each of the orientations and attitudes toward the US. The findings lend a new lens for us to study the misinformation and conspiratorial beliefs in the present context and to further our understanding of the spread of infodemic.

Women's Leadership, Political Authority and Autonomy: Challenging Gender Norms

Noora Lari, *Qatar University*

Mohammad Hassan M A Al-Ansari, *Qatar University*

Engi Assaad Ahmed Elmaghraby, *Qatar University*

The objective of this paper is to develop an understanding of the factors that contribute to the continuing representation of women in domestic responsibilities and examine its impact on their participation in leadership and political authority in patriarchal settings. National public opinion data were drawn from a 2019 survey conducted in Qatar. The findings are in accordance with the prevailing patriarchal beliefs present in Qatari society. The respondents strongly agreed with a question that favored men holding key leadership positions and positions of authority. Education had a favorable and significant association with attitudes favoring women. Compared to respondents who never attended school, those who had completed secondary school (OR 2.87; 95% CI: 0.28, 5.45) and those with higher education (OR 3.27; 95% CI: 0.72, 5.83) favored women's authority. We explore these findings in light of the theoretical expectations of 'hegemonic masculinity' which explains gender disparities in political power and leadership positions in Qatar. Future studies should explore ways to change stereotypes and traditional views of women's roles and help increase women's representation in top management positions.

Financing of Public Higher Education in Afghanistan: Public Opinion and Information Effects

Rafiuddin Najam, *Oregon State University*

The Afghan public higher education (HE) system absorbs less than one-fifth of government education spending, and per-student spending declines while the student annual enrollment is rising. This contrast may trigger challenges and inevitable tradeoffs for the public universities to encounter. Policies concerning the financing stream's diversification need attention; however, charging tuition in public institutions may confront robust public opposition due to its constitutional and historical features. We execute a randomized control trial (RCT) experiment to examine whether an information campaign contributes to lessening the opposition between citizen's views and the contents of government policies concerning the financing of public HE. We utilize the linear probability model (LPM) to shed light on information effects empirically. Our result suggests that introducing tuition fees with supplemented information to citizens concerning charging tuition could mitigate citizens' resistance. Besides, nighttime school tuition narrative as information treatment interventions concerning charging tuition fees for daytime undergraduate programs might be more effective compared to other forms of information concerning financing policies.

Scaling Up: Cannabis, Newcomers and Non-Profits in the New Economy

Arundati Dandapani, *Generation1.ca*

This presentation will draw parallels in the disruptive growth stories of the innocuously unconnected categories of legal cannabis, immigrant newcomers and non-profit associations to unravel a story of productivity, prosperity and sustainability using a mix of primary and secondary research in public opinion research, case studies, and interviews with key opinion leaders at important moments in time pre and post the pandemic.

This is the mosaic of a new economy and a new society post-COVID19, and will draw on influences from the following trends:

- Cannabis legalization takes its many forms in North America: four waves of acceptance and what they mean to society, consumers and the workplace using research from Vividata's Canadian Cannabis Study (waves 1.0 to 3.0).
- Newcomers at the forefront of change and what this means for society at the cusp of change and new opportunities.
- Data monetization, business intelligence and research storytelling are all changing how non-profits are being run, administered and marketed to appeal to new and diverse customers and stakeholders.
- The themes covered above will all converge in a vision for long term success, satisfaction and stability in expectations as the collective aspirations of a society grows.

The Crowding-out Effect of Financial Incentives for Households to Recycle Waste

Jiri Remr, *Institute for Evaluations and Social Analyses*

The willingness of individuals to recycle their household waste is driven by a whole range of factors. These may be distinguished as extrinsic, e.g. the influence of the primary social groups or opinion leaders, and intrinsic, for example the perceived importance of recycling or trust in the waste management system. In order to increase the participation rate (e.g. the share of households that recycle) or the amount of recycled waste, different policy measures and interventions are used. Typical interventions usually take the form of social campaigns, environmental education or financial incentives. In this respect, the crowding-out effect might occur when financial incentives reduce the effect of intrinsic drivers.

The presented paper evaluates the impact of financial incentives on recycling behavior and compares the effectiveness of these incentives with non-financial policy measures and interventions. Besides the "crowding-out hypothesis", a rival concept was also tested. Such an alternative hypothesis considers financial incentives as having an additive effect when said financial incentives encourage different individuals and different segments of the population. The presented data are based on a nationwide survey (n=1.579) that was recently conducted in Czechia.

The paper provides conclusions that are relevant for other situations where public policies (or intervention programs) are to be implemented. The proposed paper could be appreciated for tackling an important element of the research and for its innovative approach to measuring the attitudes toward the environment.

Exploring Self-Response: Paper Versus Web

Modeled Preference: Would Sending Paper Questionnaires to Paper-Sensitive Respondents Increase Response Rates?

Tzu-Jou Wan, *American Institutes for Research*

Michael Jackson, *American Institutes for Research*

Mahi Megra, *American Institutes for Research*

Rebecca Medway, *American Institutes for Research*

Zoe Padgett, *American Institutes for Research*

Jiashan Cui, *American Institutes for Research*

Being offered a survey via the preferred mode has been found to boost participation. Moreover, paper questionnaires tend to have higher response rates than web surveys, but also tend to be more expensive than web. These competing forces could be addressed by using paper only for cases where it is expected to have the strongest positive impact on response. However, mode preference is not typically available on sampling frames. The 2019 administration of the National Household Education Survey (NHES) included an experiment where a modeling approach was used to make predictions of mode preference (“modeled paper-only condition”). Using data from prior NHES administrations, each household was assigned a paper-sensitivity score indicating how likely it was to respond to a paper questionnaire, relative to a web survey. Within the modeled paper-only condition, the top 15 percent of households (“paper-sensitive group”) were assigned to receive paper questionnaires and the rest were assigned to receive a web-push protocol. Randomly assigned web-push and paper-only groups were used as comparisons. In these conditions, all cases received the same mode protocol, regardless of their paper-sensitivity score. The effects were evaluated at the screening phase of the NHES.

The results showed a much higher screener response rate for those who received paper-only, than those receiving web-push. However, while the predictive model successfully differentiated paper-sensitive from non-paper-sensitive households, it was seen primarily in early-stage screener response rates; the benefit of sending paper to the paper-sensitive group dwindled once the paper-sensitive web cases had also been sent paper. The modeled condition increased early-stage participation but had only a marginal impact on overall participation relative to a sequential mixed-mode protocol.

The Effects of Mail, Web, Push-to-Web and Mode Choice on Response Rates and Survey Cost

Lindsey Witt-Swanson, *University of Nebraska-Lincoln*

Amanda Ganshert, *University of Nebraska-Lincoln*

Jolene D. Smyth, *University of Nebraska-Lincoln*

Surveys are increasingly moving to self-administered and mixed-mode approaches to fight increased coverage and nonresponse issues with telephone surveys, keep costs under control, and improve data quality (Olson et al. 2019). Early research in the early 2000s on using mail and web as modes for participation for general population surveys indicated that mail-only surveys produced the highest response rates, that offering web concurrently with mail reduced response rates, and that web-push methods could be effective (see Dillman et al. 2014). However, with over 90% of adults having home internet access (Pew Research Center, 2019), it is worth reconsidering whether the early findings about whether mail modes alone continue to hold response rate advantages over designs involving web, and whether early advantages of sequential mixed-mode designs still hold over concurrent mixed-mode designs.

In this paper, we will address these questions by replicating early mode experiments. Specifically, we will report the results of an experiment embedded in the 2020 Winter Nebraska Annual Social Indicators Survey in which sampled addresses are randomly assigned to a web only condition, a mail only condition, a web and mail concurrent mixed-mode (i.e., choice) condition, or one of two web push (i.e., sequential mixed-mode) conditions that vary whether the mail is offered at the second or third follow-up mailing (n=1,687, 28.1% AAPOR RR2, January 20-April 29, 2020). All conditions use a four-contact approach, including a survey invitation, postcard reminder, and two survey reminder mailings, varying only in which modes are offered to respondents at different contact attempts. We will examine differences in response

rates, respondent characteristics, and costs across experimental conditions. We will conclude with implications for survey design and recruitment of general population surveys.

Everything But the Kitchen Sink and \$1: An Effort to Get 19-25 Year Olds to Respond to a Survey

Amanda Ganshert, *University of Nebraska-Lincoln*

Lindsey Witt-Swanson, *University of Nebraska-Lincoln*

Mindy Anderson-Knott, *Schmeeckle Research*

Push-to-web surveys have shown to be a cost-effective way to collect data when only mailing addresses are available (Smyth et al., 2010). Push-to-web surveys also have the ability to increase response for younger people by providing a web option (Smyth, Olson, & Millar, 2014). However, when asking respondents to respond via web, researchers should decrease burden as much as possible to increase response rates (Dillman et al., 2014). This leaves researchers the responsibility of determining the easiest way for the respondent to access the web survey from information printed on the letter they receive. Also, while push-to-web benefits from offering a different mode of response later, which has been shown to increase response rates (Millar & Dillman, 2011) and gather responses from different types of people (Smyth et al., 2010), the literature is not clear on the best time to offer the paper survey. Many push-to-web methodologies offer a paper survey as a final contact, but little research has been done on if this is the best time to offer the paper version.

In this study, we report results of the 2020 Nebraska Young Adult Alcohol Opinion Survey (n=15,018), a push-to-web mail survey of 19-25 year olds in Nebraska. Sample members were provided with three ways of accessing the web survey, including a shortened URL, a QR code, and texting a number to then receive the web survey link texted back. In addition, sample members were randomly assigned to either receive the paper survey in the third mailing or fourth mailing. We examine response rates across the survey access options, the demographic makeup of respondents, and timing of the paper survey. Preliminary results indicate higher response rates via shortened URL and when the paper survey was provided in the third mailing. We will conclude with recommendations for push-to-web methodology.

Re-inventing the Classics: Are Postal Surveys Possible in Turkey?

Tugba Adali, *Hacettepe University Institute of Population Studies*

Alanur Cavlin, *Hacettepe University Institute of Population Studies*

Pelin Cagatay, *Hacettepe University Institute of Population Studies*

Faruk Keskin, *Hacettepe University Institute of Population Studies*

Ahmet Sinan Türkyilmaz, *Hacettepe University Institute of Population Studies*

The COVID-19 pandemic had major design implications for many institutions relying on face-to-face interviewing for household surveys. As an organization experienced in such surveys, we felt the need to widen our spectrum of modes. Postal surveys have never been popular in Turkey, nor has been used in any major data collection attempt. Therefore, we conducted a mail survey experiment to evaluate it for future research. We mailed out 3739 envelopes in 37 clusters of around 100 households each in Ankara, using the 2018 Turkey Demographic and Health Survey (2018 TDHS) second stage sample frame. The contents of the envelopes were as follows: 1) a letter, 2) a stamped envelope, 3) a short questionnaire card. We provided 4 different channels of returning the questionnaire: 1) using the stamped envelope, 2) through sending the photo of the form using WhatsApp, 3) through calling the Institute, 4) through an online form on our website.

Individuals from 114 households responded between 20 October 2020 and 5 November 2021, with most of the responses being received within the first week following mailing. 42 envelopes were returned due to address/postal service related issues. A second round of mail was sent as a reminder to selected 200 households in 3 clusters on 5 November 2020, the results of which are yet to be observed. We will conduct a multilevel zero inflated binary logistic regression to see the odds of households responding to the mail questionnaire. Cluster level variables will be obtained from the 2018 TDHS as: average wealth index score, average years of education for adult HH members, mean household size, mean proportion of interviews and mean proportion of rejections. At the household level, the variables will be as follows: being sampled for the TDHS, being interviewed in the TDHS and being sent a reminder letter.

Who Are You? Looking at the Characteristics of Web Respondents in an IRS Household Survey

Rizwan Javaid, *Internal Revenue Service*
Brenda Schafer, *Internal Revenue Service*
Pat Langetieg, *Internal Revenue Service*
Scott Leary, *Internal Revenue Service*
Jocelyn Newsome, *Westat*
Kerry Levin, *Westat*

Given the option of completing a survey via paper or web, who chooses the web option? Similarly, when the same survey is fielded in a web-only format, who is likely to reply and who isn't? Researchers have found that surveys with a paper option garner higher response rates than those with a web-only option (Hohwu et al., 2013; Ebert et al., 2018). Preference for paper, however, is not universal across all demographic groups. McMaster et al. (2017), found that young military spouses had significantly higher response rates to web-only surveys than paper. Additionally, Newsome and colleagues (2019) found that taxpayers aged 35-49 years had significantly higher response rates to an IRS web-only survey compared to paper. As we increase fielding more surveys online, we need to better understand what characteristics of respondents make them more likely to respond by the web. The IRS Individual Taxpayer Burden (ITB) survey is an annual multi-mode survey sent to 20,000 individuals in the US. It measures the time and money taxpayers spend complying with their tax reporting responsibilities and is currently being fielded for the tenth consecutive year. The survey's fielding method has included either a standard invitation with a paper survey and web option or a web-push invitation that emphasizes taking the survey online. The average response rate when taxpayers receive an invite with a paper survey is around 36.0%, but only 23.8% for the web-push protocol. We will use the ITB survey data from the 2010-2019 to analyze response mode given the type of survey protocol. We will also consider the impact of the COVID-19 pandemic in 2020 on ITB 2019. Tax administration data, including age, filing status, filing method, income, gender, and geolocation, will allow us to build a model to better understand who is more likely to respond via web.

Patterns in Perceived Stigma Reported in a Mixed-Mode Survey of Behavioral Health Enrollees in Washington State

Felix Rodriguez, *Washington State Health Care Authority*
Rose L. Krebill-Prather, *Washington State University*
Danna L. Moore, *Washington State University*
Kent J. Miller, *Washington State University*

Individuals with behavioral health disorders face the challenge of dealing not only with their symptoms but also with social stigma. This study examines the patterns in perceived stigmatization reported by behavioral health enrollees in an annual statewide survey administered through a partnership between the Washington State Health Care Authority and Washington State University. Stigmatization is a very personal experience and feelings of being stigmatized due to behavioral health challenges may be reported differently according to survey mode. This study seeks to answer the following questions: (1) How does survey mode influence self-reports of stigma? (2) How does survey mode influence the reporting of stigma when other measures of enrollee well-being are taken into account? Employing a mixed-mode design with telephone, web, and mail-in options, the 2019 survey collected responses from 1,878 randomly selected adults who received publicly funded behavioral health treatment services in Washington State. There were 1,461 enrollees (77.8%) that participated by telephone, 205 (10.9%) by web, and 212 (11.3%) by mail. We evaluate differences in stigma perceptions by survey mode and examine how those differences influence responses on other constructs such as participation in treatment goals, improved functioning, social connectedness, and concern for privacy. Stigmatization can potentially impair the recovery process by exacerbating an individual's sense of isolation and inhibiting help-seeking behavior. Results may help to demonstrate how survey mode facilitates self-assessment of stigmatization and thus may offer a case for researchers and professionals to: (1) Factor in the role of modal approaches in understanding perceived stigma; and (2) Devise strategies for addressing bias, satisficing, and reporting of sensitive issues in surveys.

Qualitative Research Methods and Applications

The Influence of Religion on Latinx Spanish and English Speakers' Abortion Attitudes

Maria Montenegro, *Indiana University*

Julie Maier, *Indiana University*

Alejandra Kaplan, *Indiana University*

Kristen N. Jozkowski, *Indiana University*

Brandon L. Crawford, *Indiana University*

Ronna Turner, *University of Arkansas*

Wen-Juo Lo, *University of Arkansas*

Background: A vast majority of Latinx people in the US identify with a religious group and report some religiosity level; both religious denomination and religiosity are linked to people's attitudes toward abortion. For instance, scholars have found that higher degrees of religiosity and identifying as Catholic and evangelical are associated with less support for abortion. While there is research on the relationship between religiosity and abortion attitudes in the general population, there is a dearth of research focusing specifically on the US Latinx population. This study examines the nuance of whether and how religion influences Latinx's views on abortion.

Method: This analysis is part of a larger cognitive interview study examining Spanish (n = 25) and English-speaking (n = 73) participants' understanding of a salient belief elicitation survey about abortion. Participants answered open-ended survey questions related to abortion, followed by an interview assessing how they arrived at their answers. We conducted a thematic analysis on Latinx participants' interviews conducted in English and Spanish (n = 32), examining the role of religion in shaping Latinx participants' abortion attitudes.

Results: Participants described varying degrees of religious beliefs, which they indicated, directly and indirectly, affected their abortion views. Many participants identified as religious and therefore believed abortion is killing a life given by God. Alternatively, a significant minority indicated that although they do not identify as religious, religion permeates the Latinx culture and indirectly influences their thoughts on abortion. For example, although they support abortion, they see it as something shameful.

Conclusion: While religion informs participants' attitudes toward abortion, some participants' views on abortion are nuanced, complex, and sometimes contradictory. More research is needed to examine how communities largely influenced by religious beliefs and norms (e.g., Latinx community) conceptualize abortion and reconcile their religious beliefs with their abortion attitudes.

Defining Modernity in Developing India: How Peoples' Conceptual Understanding of "Modernity" and "Modernization" Affects Survey Research

Kelsey Starr, *Pew Research Center*

Ariana Monique Salazar, *Pew Research Center*

Neha Sahgal, *Pew Research Center*

Jonathan Evans, *Pew Research Center*

Manolo Corichi Gomez, *Pew Research Center*

Over the last two decades, Indian society has rapidly developed, with increased financial, technological and cultural standing in the world. Given these changes, survey researchers aptly want to measure how modernization has affected people's lives and attitudes toward a variety of social, political and religious issues. But how do Indians conceptualize modernity and how can survey researchers frame valid questions around this concept? In 2018, Pew Research Center conducted qualitative interviews and focus groups which asked Indian adults what being "modern" means to them and how modernization interacts with religion. Depending on who you ask in India, "modernity" can mean using a smart phone, being well-rounded, wearing jeans, being casual towards religious ritual and many other definitions. And most people interviewed did not see conflict between modernity and religion – in part because of their definitions of being "modern," but also because they are often seen as interconnected. This paper draws from 56 cognitive interviews and 22 focus groups conducted with a variety of religious and demographic

groups across 12 Indian states. The data allows researchers to better understand what being “modern” means in India, and how surveys might frame better questions about modernity and modernization in developing countries.

Using Focus Groups to Improve Surveys for Users of American Sign Language

Marcus Berger, *U.S. Census Bureau*

Angela O'Brien, *Census Bureau*

Betsarí Otero Class, *Census Bureau*

We know of no national data source on the number of users of American Sign Language (ASL) in the United States. However, relatively modest adjustments to typical questions about language spoken in a household would make it possible to estimate numbers of ASL users. This talk details part of a larger project aimed at investigating how major nationally representative surveys could collect data on, and produce estimates for ASL users. This project involves three focus groups with ASL experts (advocates, linguists, instructors) which will be held virtually due to COVID-19. The data will be collected in December 2020-January 2021. Experts will be identified through contacts from the National Association of the Deaf and Gallaudet University.

This presentation will report experts' thoughts on how Deaf people respond to surveys, information about culturally appropriate terminology for survey questions about ASL users, and the technology currently available for Deaf respondents' use in responding to surveys. Because there is no reliable count that we know of for the number of ASL users in the U.S., we plan to ask the experts about how important this information is to their research or work, and what sorts of data they use in the absence of good data on the topic. Finally, we will report on feedback on draft questions designed to better elicit ASL use. This talk will focus on major points and any consensus that comes out of these focus groups. Focus group results will inform cognitive interviews with ASL users in the second phase of the project.

Understanding Privacy Attitudes among Nonrespondents: Results of Qualitative Interviews With Nonrespondents to a National Household Survey

Paula Dias, *American Institutes for Research*

Melissa Scardaville, *American Institutes for Research*

Given persistent declines in survey response rates in recent decades (Brick and Williams 2013), researchers have devoted considerable attention to understanding the factors that contribute to survey nonresponse. Some researchers have argued that concerns about privacy and confidentiality can be a driver of nonresponse, particularly to government-sponsored surveys or surveys by any entity that is believed to share data or not protect respondents' privacy (Singer and Presser 2008). This presentation analyzes privacy attitudes among nonrespondents to the National Household Education Survey (NHES). It draws on findings from an in-depth qualitative study by the National Center for Education Statistics focused on better understanding the drivers of nonresponse to the NHES. The study included over 80 in-person, qualitative interviews with households that did not respond to the first three NHES:2019 survey mailings. These 90-minute interviews took place in sample members' homes at four locations across the country.

In this presentation we explore nonrespondents' definitions of privacy, reported privacy protection measures, and degrees of concern about privacy. We also report on sub-group analysis results to help understand variation in privacy attitudes. Our findings suggest that nonrespondents generally defined privacy in two ways: (1) protecting personal information (i.e. confidentiality); and/or (2) maintaining distance or boundaries between themselves and others. Additionally, some respondents believed that there was no such thing as privacy because their information was already freely available to the government or corporations. One in five participants were extremely concerned about privacy and took various measures to protect it, including not using social media or cell phones, not using banks and/or credit cards, and/or burning their mail. We conclude by discussing the implications of these findings for nonresponse. Though the drivers of nonresponse are complex, understanding privacy attitudes could help inform aspects of survey design, such as contact materials and cover letter content.

Comparing Phone versus E-mail Recruitment for Qualitative Research

Kenneth Herrell, *National Agricultural Statistics Service*

Kathy Ott, *National Agricultural Statistics Service*

While response rates to surveys have declined steadily over time (Sterrett, Malato, Benz, Tompson, & English, 2017), it has likewise been difficult for government survey methodologists to find willing participants to aide in the improvement of said surveys through cognitive interviews and other qualitative research. Recruiters for this research typically contact participants via phone calls at their phone number on record. Several factors contribute to difficulties reaching potential respondents by phone, including calls placed during the participant's working hours, increased use of Caller ID (Sakshaug, Yan, & Tourangeau, 2010), and an increasing use of e-mail for business communication. Traditionally, email has been limited for recruitment for participants in government qualitative research, due to several factors including regulations on what can be disclosed in email.

In recruiting farm operations for cognitive interviews for a livestock survey, qualitative researchers from the National Agricultural Statistics Service sought to compare whether email or phone would be a more successful method of recruiting prospective participants. With 292 participants available to contact and a target of 20 interviews to be completed, the researchers randomly assigned prospective participants who have e-mail addresses and phone numbers on file to be recruited by email or phone; in addition, a separate group of prospective participants only had phone numbers on file. Researchers compared success rates in recruiting participants by e-mail or phone, including whether they agreed to participate and whether the interview was completed successfully. In addition, the researchers investigated whether recruiting on particular days of the week or times of day was more likely to yield an agreement to participate. Finally, the researchers discuss the unique characteristics of USDA survey respondents that may make these findings unique in government business research.

Audio-Recorded Diaries in Kenya: Using a Socially Distanced Approach to Data Collection in a Low-Income Setting

Lila Rabinovich, *University of Southern California*

The COVID-19 pandemic has restricted opportunities for traditional qualitative field research. In an ongoing study led by a team at USC's Center for Economic and Social Research, we have been studying the personal financial management practices and preferences of low-income women in Kenya, through in-depth interviews and focus groups. Recognizing the significant economic challenges now facing households with low and uncertain incomes globally, we added a new module to our study to examine how women are coping financially on a day-to-day basis with the health and economic emergency unleashed by the pandemic. In order to comply with social distancing guidelines, we deployed a solicited diaries approach to data collection. Solicited diaries aim to elicit every-day experiences on particular themes, and allow the capture of in-depth personal perspectives in the context of day to day activities, emotions and interactions.

Specifically, we used "audio recorded diaries", an adapted diary approach to account for literacy constraints and limited access to personal computers. A sample of twenty of our existing sample of women were asked to submit entries every day for a total of one week. The diaries helped us understand what financial services women still use or rely on during this time, what barriers they may experience to accessing financial services and products, and what supporting intervention they may benefit from. Two follow-up interviews provided additional opportunity to delve into diary entries in greater depth. In addition to outlining our findings, the presentation will discuss the audio-recorded diary approach, challenges and opportunities faced in its deployment in our study setting, adaptations, and lessons for use in Kenya and elsewhere.

A Mixed Methods Approach to Evaluating an Age of Disability Onset Survey Question for Adults With an Intellectual or Developmental Disability

Ann MacFadyen, *National Center for Health Statistics*

The prevalence of adults with intellectual or developmental disabilities (IDD) is not sufficiently captured through current population survey questionnaires (Krahn, 2019). In order to address this gap in prevalence data, the Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER) is conducting a question evaluation study to assess questions related to key identifiers of IDD. The evaluation uses qualitative data from cognitive interviewing to explore constructs captured, particularly among those with IDD and their proxies; and, the study uses quantitative data from the National Center for Health Statistics' Research and Development Survey (RANDS) – which utilizes a probability-sampled commercial survey panel – to inform our understanding of measurement error, particularly among those without IDD. This presentation will focus on the performance of the age of disability onset question, an essential identifier for IDD.

The age of onset question evaluated here aims to identify respondents with IDD, by determining that the age of their disability onset was before 22, meaning that onset occurred during development. Age of onset is a critical criterion used for identifying people with IDD. When coupled with other identifiers such as learning ability and independent living skills, it provides a way to differentiate between those with IDD and those with other conditions that impact areas of life activity but are not developmental, such as Alzheimer's disease or stroke (Havercamp et al. 2019). Following a series of questions about functional difficulties in day to day life, respondents to both cognitive interviews and the RANDS were asked the onset question: "You said you had difficulty with [Fill from prior questions]. Did [this difficulty/any of these difficulties] begin before age 22?" The RANDS design included targeted, embedded probes, expanding our cognitive interview findings to a wider sample.

Multi-Level and Geographic Modeling Applications

A Poll-Based Bayesian Hierarchical Model for American Presidential Election

Brittany Alexander, *Texas A&M University*

In this presentation, a Bayesian Hierarchical polling aggregation model for American Presidential Elections is presented. It considers the support for the Democratic and Republican candidates relative to each other. It has multiple variations that control the uncertainty, the weight of the polling data, and the number of polls to include. It is not a forecast and attempts to aggregate the polls at various points in election cycle. At the end of an election cycle it can be used to predict the outcome. It uses previous elections results and poll data from other states to create partially pooled estimates of support for a candidate even when no polling data exists in a state. This model is computationally efficient and can be run on a laptop in minutes. This model was used to predict the 2020 election, and was tested on data from the 2008, 2012, and 2016 elections. This method outperforms a polling average and provides reasonable estimates of uncertainty in each state. It also produces highly similar estimates to the FiveThirtyEight and Economist model while not taking hours to run. This method could also be adapted to predict Senate or House races.

Bayesian Estimation of Program Specific Impacts in the HPOG Program

Stanislav Kolenikov, *Abt Associates*

David Judkins, *Abt Associates*

Local Health Profession Opportunity Grants (HPOG) programs, funded by the Administration for Children and Families within the U.S. Department of Health and Human Services, provide education, training, and support services (including financial and other assistance) to help transition low-income adults into healthcare occupations. ACF's Office of Planning, Research, and Evaluation funded an evaluation to assess the success of these HPOG programs and to provide local, program-specific estimates of impact. Direct estimation of the local average treatment effect (LATE) consists of simply comparing the means of outcomes for the local treatment and control groups to estimate local impacts. Unfortunately, most programs serve too few students to support estimation of local impacts. To overcome those issues, we developed a complementary set of Bayesian estimates of local impacts based on mixed effect models with random effects defined at the program level, and random slopes for the treatment indicator. Before preparing Bayesian estimates of local program effects for the second round of grants (HPOG 2.0), we demonstrated the techniques on the previous round (HPOG 1.0) evaluation data. In addition to allowing methods revisions without fear of accusations of p-hacking, demonstrating the techniques on HPOG 1.0 first allowed us to use the posterior distributions for components of variance as priors for components of variance for HPOG 2.0. As expected from general methodological considerations, Bayesian estimates of impact exhibited less variability than direct estimates did. Bayesian credible intervals were shorter, often by a factor of about 2 to 4, than the confidence intervals at the same coverage level. At the same time, a frequentist empirical Bayes (EB) analysis of the same mixed models produced confidence intervals that were half as long as Bayesian intervals, still, which highlights the importance of properly accounting for the uncertainty in the variance component estimation that EB methods cannot fully incorporate.

A Methodological Primer on Using Adaptive Bandwidth Kernel Density for Aggregating Spatial Data: Applications to Social Science Research

Peter Herman, *NORC at the University of Chicago*

Adam F. Benson, *Schroeder Institute at Truth Initiative*

Minal Patel, *Schroeder Institute at Truth Initiative*

Chang Zhao, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

Barbara S. Schillo, *Schroeder Institute at Truth Initiative*

Researchers often need to compare two datasets at different spatial resolutions, such as point data and data aggregated to Census boundaries. While it is possible to directly sum the number of points within geographically bounded regions, this method risks failing to capture the true location of data due to an aggregation bias known as the Modifiable Areal Unit Problem (MAUP). Adaptive bandwidth kernel density estimation (KDE) is a nonparametric method of extrapolating spatially-distributed point location data over an area by calculating the density of the point locations using a specified bandwidth. Adaptive bandwidth KDE allows for the comparison of two spatial datasets at varying levels of resolution while avoiding the MAUP by producing a continuous raster surface in which each cell is assigned a density value regardless of administrative boundaries. This technique is particularly useful in helping to control for underlying differences in population density across a country or area of interest as bandwidths are allowed to vary until a certain population threshold is reached. For example, by using a complementary raster surface of population count and limiting the influence of each point to a maximum surrounding population of 1000 people, adaptive bandwidth KDE will yield cells in which the value of each cell is equal to the count of points per 1000 people (i.e. features per 1000 persons). These density values can then be aggregated to any administrative level via zonal statistics. The application of these methodologies can be used to examine a variety of issues, including understanding the density of storefronts that may impact health behaviors. Using examples from the Truth Initiative and NORC at the University of Chicago's tobacco retailer density study, this methodological overview details the why and how of adaptive bandwidth KDE using examples written in R with the raster package.

Consuming the News

New Data, Old Practices: Perceptions of Public Opinion in the News

Ozan Kuru, *National University of Singapore*

Shannon McGregor, *University of North Carolina*

With the growing centrality of social media in the public sphere, journalists have also begun to report on public opinion using anecdotes and statistical data from social media. This new sourcing for journalists mostly includes tweets being quoted as vox pop in the stories or aggregate/statistical data about Twitter buzz such as Twitter trends, popular hashtags or data journalists' own analyses of Twitter data around an issue. These new data sources communicate a strong social message about public opinion, yet one that may further undermine the public's shaky trust in the news media and public opinion industry. In this study, we aim to determine two crucial questions: to what extent journalists' use of social media data as well as traditional forms in news reports impact people's perceptions of public opinion as well as how credible people judge that content to be. In January 2020, using a non-probability Qualtrics sample, we conducted an online survey experiment in the U.S. to test how individuals evaluated and reacted to messages about different forms of data reported in news stories as evidence of public opinion. Respondents were shown one of the six news stories showing American public opinion on one polarizing (government spending on healthcare) and one non-polarized (marijuana legalization) issue in the form of either a poll, a vox pop, combination of poll and vox pop, tweet quotes, a Twitter trend analysis, and a combination of a tweet quotes and Twitter trend analysis. Initial analyses showed that there were little differences in public opinion perceptions between each type of data, but perceived credibility of messages differed, controlling for a variety of other factors. We discuss the results in relation to changing dynamics in public opinion coverage at a time of challenges in measuring public opinion effectively and public perceptions of public opinion metrics.

Using Moderated and Unmoderated Usability Testing to Assess Interactive BLS News Releases

Struther Van Horn, *U.S. Bureau of Labor Statistics*

Jean Fox, *U.S. Bureau of Labor Statistics*

David Biagas, *U.S. Bureau of Labor Statistics*

The Bureau of Labor Statistics (BLS) has designed prototypes for new and improved news releases. The "Next Generation News Releases" incorporate new charts and other interactive features, rather than just presenting static text. These new features are intended to help data users understand the data more easily and to allow them to tailor the information to their own needs. Prior to implementing these changes, we wanted to be sure that data users could understand and use the new interactive features easily. Observing users completing tasks with the news releases in a usability test allowed us to identify problems and areas that could be improved.

To evaluate the usability of the news release prototypes, we used three different approaches: (1) moderated, in-person testing, (2) moderated, remote testing, and (3) unmoderated testing. In the two moderated approaches, a researcher interacted directly with the participant during the session. We conducted the in-person sessions in our usability lab and the remote sessions using the web conferencing service Zoom. We conducted the unmoderated tests with TryMyUI, a service that provides remote, unmoderated usability testing. In the unmoderated sessions, participants who met the screening requirements completed the usability test on their own, following prompts in TryMyUI.

In this presentation, we will discuss the benefits and drawbacks of utilizing both moderated and unmoderated usability testing for evaluating the interactive features of the news release prototypes. We will also share lessons learned from each of the approaches.

Exploring the Over and Underreporting of News and Reality Television Exposure in the Israeli Context

Danit Shalev, *University of Haifa*

Yariv Tsfati, *University of Haifa*

While there is ample research about why people overreport their exposure to news programs, few studies have explored this issue outside the US. Furthermore, scholars debate to what extent social desirability is the main reason for this overreporting. We investigated whether overreporting would be reduced or even underreporting will be evidenced when respondents were asked about watching less socially desirable television content, such as reality TV. Two studies demonstrated that exposure to both news and reality TV programs is overreported in Israel. However, the facts that reported exposure to both genres was correlated with the perceived social desirability of the shows, and that those who regarded reality programs as socially undesirable underreported watching them, strengthen the social desirability explanation.

#INSTAPOLITICS: How Politicians in Post-Yugoslav Countries Use Instagram in Order to Self-Present Themselves?

Milos Moskovljevic, *City University of Hong Kong*

Masood Muhammad, *City University of Hong Kong*

Without understanding politicians and how they communicate with the public, we would not completely comprehend the realm of politics. Previous decades have been marked by an uprising of politicians using different social media platforms, and scholars argued that these phenomena could be approached as an indicator of increasing personalization in the political promotion. However, it is quite unusual that there is a lack of studies dealing with personalized political communication in post-communist countries, and more precisely, in former Yugoslavia states. This area of Europe is quite interesting because certain post-Yugoslav countries are experiencing regression from democracy towards autocratic regimes. One of the consequences of that trend is that traditional media channels are being captured by leading political parties, while opposition parties must rely exclusively on social media in order to reach citizens. One of the social media platforms that have shown to be a powerful platform during campaigns is Instagram (cite a scholar or news author). Therefore, this paper aims to investigate whether members from leading and opposition political parties use Instagram in different manners and unpack different underlying patterns. For this purpose, a qualitative analysis of 18 Instagram profiles of Serbia, Montenegro, and Croatia politicians will be conducted. In particular, six from each country and three from leading and three from opposition parties. We believe that findings will provide novel insights about political elites' social media use from a non-mainstream political context.

Measuring News Consumption in a Digital Era

Michael Barthel, *Pew Research Center*

Dorene Asare-Marfo, *Pew Research Center*

Kirsten Worden, *Pew Research Center*

The news media's transition to digital has brought major upheaval to the industry – including to those who seek to measure the public's consumption of news. In the mid-twentieth century, when media research came into its own, this task was more straightforward. There were only a few different ways to get news, and all were clearly distinct – print publications, radio or television. But over the past decades, in addition to a plethora of new forms of news (from 24-hour news channels to news websites), many news outlets produce content across multiple platforms. And there is an industry-wide concern that news consumption habits are overestimated in surveys where respondents self-report their behavior.

We will present the main findings of a multi-method investigation into measuring news consumption, incorporating cognitive interviews, survey experiments, and passive tracking of online activity, in addition to a nationally representative survey of U.S. adults. This survey finds mixed evidence about the public's understanding of newer forms of media and news – which impacts the topics survey researchers can reasonably ask about. The survey experiments, meanwhile, reveal that while there is no “silver bullet” for

perfect survey measures of news consumption, a series of refinements (including item wording and response order) could drive marginal improvements around the goal of reducing overreporting. And an exploration of the potential to use passive data as a direct measurement of the public's digital news habits, free of the concerns with self-reporting inherent to surveys, shows some promise. Yet there are still too many pitfalls to rely on it for a complete portrait of Americans' digital news consumption. Estimates coming in from passive data are systematically higher than those from survey questions, with inadequate coverage of devices one apparent culprit – even as aspirational reporting continues to inflate survey responses.

Political Participation among Newsjunkies: Intrinsic Need for Orientation and Voter Registration, Vote Propensity, and Voter Conscientiousness

Justin Martin, *Northwestern University*

Krishna Sharma, *Northwestern University*

This study examined the newsjunkie characteristic—the intrinsic need for orientation (INFO)—as a predictor of voter registration, vote propensity, and voter conscientiousness, while controlling for news consumption, social media use, measures of political partisanship, and demographic factors. The study is based on a nationally representative sample of U.S. adults (N=2,059) collected by The Harris Poll in September 2020 via its Harris On Demand omnibus survey.

The intrinsic-need-for-orientation scale is a four-item measure with questions on the extent to which people access news in their downtime, feel discomfort when they are cut off from news, check news among the first things they do each day, and believe that keeping up with news better connects them with other people. Introduced in 2019, the INFO measure is rooted in the uses-and-gratifications of mass media paradigm and self-determination theory, and has been found associated with a number of attitudinal and behavioral outcomes. In one example, the newsjunkie characteristic was associated with intention to vote for the Democratic candidate in the 2020 U.S. presidential election, after controlling for social media use and demographic factors.

The three dependent variables in this study are being registered to vote, respondents' estimate of the likelihood that they would vote in the 2020 U.S. presidential election, and voter conscientiousness (an individual's feeling of guilt when an election passes and they did not vote). Intrinsic need for orientation positively and significantly predicted all three outcome variables, even after controlling for news consumption, social media use, political partisanship, and demographics. Moreover, INFO was the single strongest predictor of voter conscientiousness. Consuming news about politics was also positively correlated with all three of the dependent variables. This research highlights the usefulness of the INFO measure in research on political behaviors.

Comparing Estimates of News Consumption From Survey and Passively Collected Behavioral Data

David Rothschild, *Microsoft*

Tobias Konitzer, *PredictWise*

Stephanie Eckman, *RTI International*

Surveys are a vital tool in understanding public opinion and knowledge, but behaviors can be misrepresented in surveys. We explore a popular and important behavior frequently measured in public opinion surveys: news consumption. Previous studies have shown that television news consumption is consistently overreported in surveys relative to passively collected behavioral data. We validate these earlier findings, demonstrate the value of the relative flexibility of passively collected data, and extend the literature further by showing that online and social media news consumption, both levels and trends, also differ dramatically from passively collected data on behavior. We outline a framework for using a mix of passively collected behavioral and survey-generated attitudinal data to accurately estimate consumption of news and related effects on public opinion and knowledge, conditional on media consumption.

Focusing in one set of findings: we show that survey estimates of news within social media are two to three times greater than estimates from passively collected data. Approximately 28% of Facebook users and 11% of YouTube users consumed any news on the platform in the month of August 2018, according to the passive data—as opposed to the survey-based measures of 43% and 21%, respectively. Further, in looking at results from 2016 and 2017, we document that Facebook news consumption, conditional on being on the site, dropped from 38% to 28%, which was entirely missed by the survey data. Conversely, although repeated cross-sectional surveys show a dramatic increase in news consumption on YouTube, we do not see this reflected in our passively collected data, for this timeframe.

Response Options and Response Scales

Comparing the Performance of an Unbalanced Versus Balanced Response Scale on Self-Rated Health in the United States

Kristen Cibelli Hibben, *National Center for Health Statistics*

Rebecca Hu, *National Center for Health Statistics*

Morgan Earp, *National Center for Health Statistics*

Kristen Miller, *National Center for Health Statistics*

Self-rated health (SRH) is a widely used measure of respondents' subjective evaluation of their health status. Defined as an individual's perceived overall health, extensive research has demonstrated its utility and led to its widespread use in medical, social, and behavioral science research using survey data. Included in major surveys in the United States and around the world, SRH is frequently used comparatively across key respondent characteristics (e.g., race, ethnicity, education, etc.) within countries and across countries. Despite its prevalence, no universally agreed upon response scale exists for SRH, making it challenging to compare SRH across surveys. Two main versions exist: the unbalanced scale ("Would you say your health in general is excellent, very good, good, fair, or poor?"), used predominately in the US and the balanced scale ("Would you say your health in general is very good, good, fair, bad, or very bad?"), recommended by the World Health Organization (WHO) and used widely in Europe. Some have argued that the balanced scale, endorsed by the WHO, may be preferable given the obvious balance in the categories – two positive categories (very good, good), one neutral category (fair), and two negative categories (bad, very bad) (Jurges, Avendano, & Mackenbach, 2008). Yet, little empirical research has compared the two versions apart from a study carried out across five European countries (Jurges et al., 2008) and to date no studies have taken place in the US. This study will compare the two different versions of the response scale using a US sample collected in the third round of the Research and Development Survey (RANDS), a recruited web panel, specifically, NORC's Amerispeak panel, which uses the framework of probability sampling. The performance of the two response scales will be evaluated using descriptive and regression statistical methods as well as Item Response Theory (IRT).

The More the Merrier? Comparing the Number of Response Options in a Rating Scale and its Effect on Data Quality

Eva Aizpurua, *European Social Survey HQ - City, University of London*

Ki H. Park, *Center for Social & Behavioral Research, University of Northern Iowa*

Erin O. Heiden, *Center for Social & Behavioral Research, University of Northern Iowa*

Mary E. Losch, *Center for Social & Behavioral Research, University of Northern Iowa*

Rating scales are widely used in survey research. The number of response options is known to influence data quality and is a key consideration in scale development. Despite the abundance of studies on this topic, the evidence is mixed, and the findings vary depending on multiple factors, including the mode of administration and the indicators used to measure data quality. In this study, we compare the performance of a 5- and a 10-point unipolar scale measuring self-reported distraction. The dataset comes from a dual-frame random-digit-dial telephone survey of adults in a Midwestern state who were queried on topics related to awareness of and attitudes toward STEM education. Respondents were randomly assigned to one of two conditions that manipulated the format of the rating scale. The first condition used a 5-point, fully labelled scale while the second used a 10-point scale (1 anchored with "not distracted at all" and 10 with "extremely distracted"). The label order and the end-point labels were consistent in both groups (incremental). We compare the proportion of item nonresponse, the distributions, floor and ceiling effects, and the convergent validity across the two scales. The implications of these findings and how they fit in with previous work will be discussed.

Discriminating Measures: Response Formats in Measuring Discrimination

Randall K. Thomas, *Ipsos Public Affairs*

Jasmine Stephens, *Ipsos Public Affairs*

Beatrice Abiero, *Ipsos Public Affairs*

Frances M. Barlas, *Ipsos Public Affairs*

Measurement of perceived discrimination has received empirical attention across many fields, although its measurement has varied based on different types of discrimination (e.g., race-ethnicity or gender discrimination). As a result, comparing the nature and extent of various forms of discrimination has been problematic. This study focused on developing consistent measurement for self-reported discrimination across different forms of discrimination. We conducted an experiment to examine how response formats affect prevalence reports of discrimination. Dichotomous yes-no formats (YNF) have been typically found to yield higher prevalence results than multiple response formats (MRF – “Select all”) in such varied research areas as product ownership, employment, and health condition. Generally, more salient events (stronger, more recent, etc.) are selected for MRF while both more salient and less salient events were selected for YNF. In our first study, we had over 6,043 online respondents from a non-probability-based KnowledgePanel and 5,829 from an opt-in sample. We randomly assigned respondents to two different response formats (YNF or MRF) and various recall periods (Study 1 – “Ever” or “Past 5 years”; Study 2 – “Ever”, “Past 5 years”, or “Past year”). We asked about three specific forms of discrimination, randomly selected from a set of seven types (e.g., racial, gender, age, etc.). We then presented a series of events where discrimination may have occurred (employment, education, housing, etc.). In Study 2, if a person indicated they experienced any discrimination, we asked how negative of an impact it had on them (measuring salience). We found that reports of discrimination were higher for the YNF than the MRF and, in line with other research on these formats, that discrimination events were rated as more negatively impactful (salient) in the MRF.

Do 5-Point, 7-Point or 11-Point Scales Produce More Accurate Measurement of Attitudinal Attributes?

Randal Ries, *IBM*

William Dusch, *IBM*

Annette Tassone, *IBM*

Jon Temple, *IBM*

Established psychometric research suggests that giving survey respondents more response options produces better quality data and reduces measurement error (Nunnally 1978). With more response options, respondents are better able to select a category that reflects their sentiment. A 2010 study published in *Measuring U* found that in some instances respondents using 5-point scales were forced to interpolate their answers due to the limited choices while none of the respondents in the study exhibited this behavior when using 7-point scales (Sauro 2010). A study published in the *International Journal of Market Research* comparing 5-point, 7-point and 10-point scales finds that after re-scaling, 5-point and 7-point scales produce higher means and data with a more negative skew (Dawes 2008).

This research set out to determine if information is lost with a simple 5-point Likert scale or if the quality of data can be improved with a 7- or 11-point scale? For the purposes of this research, an experiment was designed whereby respondents in the same survey were randomly assigned to one of three groups with each group answering the same four attitudinal questions but on either 5-point, 7-point or 11-point scales. The research was conducted among the employees of a large global enterprise using an instrument translated into nine languages with a sample size of over 5,000.

The findings indicate that while 7-point and 11-point scales produce data with greater variance, the data collected on these scales is more skewed and has greater kurtosis than data collected using 5-point scales. This analysis also indicates the respondents used a greater variety of scale points when using 7-point or 11-point scales, in other words they were less likely to limit choices to one or two categories. We find that 7-point and 11-point scales yield improved discriminability through greater variance.

Response Style or Response Patterns? Effects of Race and Ethnicity on Response Selection

Omar Pedraza, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

Some researchers have indicated that there may be differences in scale usage as a result of racial or ethnic background or country of residence. In the U.S., we often encounter anecdotes from other researchers who believe that Black and Hispanic respondents are more likely to use more positive response categories, though there has only been mixed research supporting this assertion. In this study, we examined response patterns across race and ethnic backgrounds using two separate online surveys. We examined response patterns for both bipolar and unipolar formats across a wide range of topics and with response format varying in the number of response categories. We specifically looked at patterns reflecting positivity of response, extreme responding, and middling response. In addition, many of the items we examined had measures of prior experience or familiarity with the topics which we believed based on prior research was a moderator of response patterns. Across studies, we found strong and significant differences in scale category usage as a result of ethnic or racial background. However, these differences were item-specific and were not consistent with the hypothesis that one group would be more likely to use more positive categories on a scale or be more extreme or middling in their responses. Our analyses indicated that, when differences occurred, they were more likely due to differences in familiarity or other topic-related issues and not due to differences in race or ethnicity. These results lend support to the view that there can be response patterns due to situational factors rather than response styles due to people's characteristics that cause them to respond differently regardless of topic.

The Measurement of Self-Rated Mental Health: Relationships With Mental Health Outcomes and Response Option Order Effects

Rachel Stenger, *University of Nebraska-Lincoln*

Christina Falci, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

Jolene D. Smyth, *University of Nebraska-Lincoln*

Self-rated health (SRH) items, wherein respondents are asked to rate their health with the common response options of "excellent," "very good," "good," "fair," and "poor" are often used in health-related surveys (Ahmad et al. 2014). Research has examined the predictive validity (e.g., Idler and Benyamini 1997) and the measurement properties (e.g., primacy effects) (Garbarski, Shaeffer, and Dykema 2015, 2016, 2019) of these items. Comparatively, little research has examined the validity and measurement properties of self-rated mental health (SRMH). Little is known about how the SRMH item is associated with mental health outcomes, like anxiety and alcohol consumption, with stressors that can reduce mental health, or with protective factors such as personal and social resources that may improve mental health (Pearlin 1999).

This paper will explore the convergent validity and measurement properties of the SRMH measure by examining the following research questions: (1) How is SRMH associated with specific mental health outcomes like depressive symptoms, anxiety, paranoia, and alcohol abuse? (2) Is the pattern of associations for stressors and resources with SRMH consistent with those for specific mental health outcomes? (3) Does the order of response options in the SRMH question affect responses and does response option order moderate the above associations? Data for the paper come from the 2014 National Health, Wellbeing, and Perspectives Study (NHWPS), a nationally representative, address-based sample, self-administered survey (N=1,002; AAPOR RR1=16.7%). NHWPS included a number of mental health-related measures, an SRMH question, and an experiment on the order of response options for the SRMH item. Preliminary analyses show that responses are sensitive to response option order, but order has little effect on the relationships between SRMH and other mental health measures. This paper will conclude with implications for the measurement of SRMH and its utility as an alternative measure for dimension-specific mental health outcomes.

Satisfied With Your Satisfaction Questions? Validity Evidence for Bipolar, Branching, and Unipolar Question Designs

Aaron Sedley, *Google, Inc.*

Yongwei Yang, *Google, Inc.*

Asaf Beasley, *Google, Inc.*

The survey and consumer research communities have seen considerable discussion about optimal questionnaire design for measuring consumer and user attitudes. Specifically for satisfaction, which has historically been treated as a bipolar construct (Thurstone, 1931; Likert, 1932), some argue that it is actually two separate unipolar constructs, which may yield signals with separable and interactive dynamics (Cacioppo & Berntson, 1994).

Related research has explored whether attitude measurement validity can be optimized with a branching design that involves two questions: a question about the direction of an attitude (e.g., positive, negative) followed by a question using a unipolar scale according to the direction selected in the previous question (Krosnick & Berent, 1993). The current experiment gathers data to evaluate differences across a variety of question designs for in-product contextual satisfaction surveys (Sedley & Müller, 2016). Specifically, we randomly assign respondents into one of the following question designs:

1. Traditional 5-point bipolar satisfaction scale (fully labeled)
2. Branched: Initial directional question (satisfied, neither satisfied nor dissatisfied, dissatisfied), followed by a unipolar question for degree in the corresponding direction (5-point scale from “not at all” to “extremely,” fully labeled)
3. Unipolar satisfaction scale, followed by a unipolar dissatisfaction scale (both use 5-point scale from “not at all” to “extremely,” fully labeled)
4. Unipolar dissatisfaction scale, followed by a unipolar satisfaction scale both use 5-point scale from “not at all” to “extremely,” fully labeled)

The current experiment meaningfully adds to the attitude question design literature in that it compares question designs based on criterion validity evidence; namely the relationship with user behaviors linked to survey responses. Additionally, we consider response latency, response rates, break-off rates, as well as relationships with other constructs measured in the same survey.

COVID-19 Social Desirability Wording Experiment

Leanna Moron, *University of Maryland, Joint Program in Survey Methodology*

Samantha Chiu, *University of Maryland, Joint Program in Survey Methodology*

Xin (Rosallynn) Yang, *University of Maryland, Joint Program in Survey Methodology*

Stanley Presser, *University of Maryland, Joint Program in Survey Methodology*

Frauke Kreuter, *University of Maryland, Joint Program in Survey Methodology*

Social desirability bias can be described as survey respondents' tendency to present themselves in a favorable light by underreporting socially undesirable behaviors or attributes, and overreporting socially desirable ones. Our research explores whether this bias in the measurement of behaviors related to the COVID-19 pandemic (e.g., social distancing) can be reduced by changes in question wording. Using a web-based survey that recruited participants via Facebook advertisements, we compared conventional wording and response categories with those that mentioned extenuating circumstances to make it easier for respondents to report socially undesirable behaviors. The results may have implications not only for understanding the public's response to the COVID-19 pandemic, but for the measurement of the many others behaviors that are seen as socially undesirable.

5/11/2021, 12:00 PM - 1:30 PM
Concurrent Session A: Live Sessions

Mini-Conference: Live Video Survey Interviews when In-Person Data Collection is Impossible: Challenges for Recruitment and Participation

Live Video Survey Interviews When In-Person Data Collection is Impossible: Challenges for Recruitment and Participation

Andrew Hupp, *University of Michigan*

In the pandemic era, surveys that have long relied on in-person data collection are being forced to consider alternative modes. While online self-administered or phone surveys may be an option for some, the face-to-face component may be particularly critical in other cases. This session brings together studies on the feasibility of conducting live video interviews as an alternative to in-person, in an era when many more members of the public (at least those with access to the technology) have become accustomed to using video in daily communication--but concerns about data privacy and security in a much more virtual world may also be rising. The papers presented here focus on new evidence about the recruitment and participation challenges that live video data collection introduces. Two papers concern what researchers have learned in recent high-visibility production surveys in which they have chosen to explore live video as an alternative to in-person data collection during the pandemic, one in Europe (European Social Survey) and the other in the US (American National Election Studies). Two other papers present experimental evidence about recruitment and participation in live video interviews, one about sample members' willingness to participate in live video surveys relative to other modes and what might predict their preferences during a pandemic era, and the other comparing recruitment into and breakoff rates in live video vs. self-administered online modes to which sample members have been randomly assigned. The evidence presented in this panel, from studies that range substantially in their approaches to recruitment, should provide useful information for practitioners and methodologists who are considering live video as a mode for survey data collection.

Predictors of Willingness to Participate in Live Video Survey Interviews

Shlomit Okon, *The New School*

Michael F. Schober, *The New School*

Frederick G. Conrad, *University of Michigan*

Andrew L. Hupp, *University of Michigan*

Ai Rene Ong, *University of Michigan*

Kallan M. Larsen, *University of Michigan*

Given the necessity of considering alternatives to in-person data collection in the COVID moment, when and how people will be willing to participate in standardized surveys that inform policy has become critically important. The study presented here tests hypotheses about four potential predictors of respondents' willingness to participate in surveys via the live video that has become ubiquitous for many members of the public in the COVID era--at least for those who have access and sufficient connectivity. Three potential predictors (selecting factors that have predicted technology acceptance and behavioral intention in other communication modes) are participants' perceptions about using live video for communication in contexts other than surveys, in particular (1) how easy they find live video to use, (2) how useful they find live video, and (3) how much they enjoy using live video. A fourth hypothesized (survey-specific) predictor is the extent to which participants report that they (4) would be uncomfortable answering a particular sensitive question in live video relative to other survey modes.

In the study, online participants rate their willingness to take part in a hypothetical live video survey that might ask about personal information in comparison with their willingness to take part in four other survey modes that differ from live video in theoretically and practically important ways: in-person, phone, a traditional (text-only) web survey, and a "prerecorded-video" web survey in which respondents play embedded video of interviewers reading questions. The modes vary in whether they are interviewer- or self-administered, in the social presence of the interviewer (which in video could be either intrusive or rapport-building), and in the interactive nature of the survey dialogue. Findings will be informative about the feasibility of live video interviewing for large-scale data collection moving forward

The European Social Survey During COVID-19: Using Video Interviews and Other Innovations

Tim Hanson, *City, University of London (European Social Survey HQ)*

The COVID-19 pandemic has presented many challenges for surveys that usually rely on in-person (face-to-face) data collection. This includes the European Social Survey (ESS), which has been conducted biennially across over 30 countries since 2002, with all countries delivering in-person samples at each round based on a central specification. Following consideration of alternative approaches, it was decided to continue to plan for in-person data collection at Round 10 of ESS (originally due to be carried out roughly in the last quarter of 2020), but to expand the fieldwork period to the end of 2021. The decision to retain an in-person approach was based on the importance of a mode-consistent time series (especially in the context of understanding the impact of the pandemic on attitudes and behaviour), and a view that no alternative methodology could be delivered consistently by all 30+ countries to an equivalent standard.

It was however acknowledged that the pandemic may impact on the participation of certain groups in an in-person interview, even after a return to relative normality. We have therefore adapted the ESS specification to allow video interviews to be conducted in cases where target respondents are unable or unwilling to be interviewed in-person. We expect video interviews to be implemented by 10-15 countries at ESS Round 10, as a complementary method to the usriefly describe other innovations on ESS, including development of self-completion modes, linked both to the COVID-19 pandemic and longer-term concerns about the continued feasibility of in-person fieldwork in some countries. This paper will share key features of our approach to video interviewing, initial findings on the prevalence and success of video interviews in countries, and future learning based on our experience to date. We will also briefly describe other innovations on ESS, including development of self-completion modes, linked both to the COVID-19 pandemic and longer-term concerns about the continued feasibility of in-person fieldwork in some countries.

Recruitment and Participation in Video Interviews

Kallan Larsen, *University of Michigan*
Andrew L. Hupp, *University of Michigan*
Frederick G. Conrad, *University of Michigan*
Michael F. Schober, *The New School*
Ai Rene Ong, *University of Michigan*
Brady T. West, *University of Michigan*
Tianheao Wang, *University of Michigan*

The COVID-19 pandemic has prompted survey researchers to turn to video technology as an alternative to in-person data collection. Despite a widespread increase in familiarity with video technology, the novelty of video as a survey mode leaves many questions unanswered regarding best recruitment practices and patterns of participation. Live video interviews, while much like phone interviews, require a recruitment process more similar to in-person interviewing: respondents must schedule a time at which to meet with interviewers, and “cold calling” is not a feature of most video platforms. Self-administered recorded video, where questions are asked via pre-recorded videos of interviewers embedded in a web survey (rather than written text), requires respondents to have audio capabilities and can restrict the speed at which an interview is completed relative to a traditional web survey (because of the time it takes the videos to play). From August 2019 to March 2020, we conducted a mode comparison experiment using both an address-based sample and two online non-probability panels as recruitment sources ($n = 1083$ completed cases in total). We compare recruitment and participation patterns (including breakoffs) for each sample source across three modes: live video, recorded video, and traditional (text-only) web. We examine the apparent impacts of device type and respondent age on completion rates. We did not find ABS to be an effective recruitment method for live video interviews. Recruitment was most difficult for live video in online panels as well, and breakoffs were most common in the recorded video mode and particularly infrequent in live video. As with any mode, there are advantages and disadvantages to video modes, and we will briefly discuss these tradeoffs as they relate to data quality.

Live Video Interviewing in the ANES 2020 Time Series Study

Lauren Guggenheim, *University of Michigan*
Natalya Maisel, *Stanford*
David Howell, *University of Michigan*
Michelle Amsbary, *Westat*
Ted Brader, *University of Michigan*
Matthew DeBell, *Stanford*
Cindy Good, *Westat*
D. Sunshine Hillygus, *Duke University*

Since 1948, the presidential time series conducted by the American National Election Studies (ANES) has included a face-to-face component. In 2020, however, challenges related to the COVID-19 pandemic led to the replacement of in-person interviewing with a contactless, mixed-mode design that included surveys self-administered by web and surveys administered by interviewers via live video or phone. Each protocol used the same nationally representative sample, selected using probability methods from an address-based frame. Sampled households were mailed hard-copy invitations and asked to complete an online screener, after which a respondent was randomly selected from the household and subsequently assigned to participate by Zoom video conferencing or by a self-administered web questionnaire. Telephone interviews were presented as an option later in some protocols to allow the participation of respondents uncomfortable with, or unable to access, web-based or video interviewing. Several interventions were implemented across all modes to increase response rates, including escalating incentives, and there were also interventions specific to video such as the use of a specially trained refusal conversion team.

Our focus here is specifically on the impact of the use of video on recruitment and participation, including how potential respondents and interviewers reacted to and perceived video interviews. To do so, we employ several types of data collected over the duration of the survey, including interviewer observations, differential participation by demographics and political attributes, help desk logs, participant emails, and a

survey of interviewers. We supplement this by comparison to data previously collected by ANES and against groups that participated by a mode other than video in the current study. We additionally discuss design decisions such as whether to schedule appointments or have interviewers standing by, and their implications for participation. We further reflect on how the pandemic may have influenced participation across the different modes.

Interviewer Effects in Live Video and Recorded-Video Interviewing

Ai Rene Ong, *University of Michigan*

Brady T. West, *University of Michigan*

Frederick G. Conrad, *University of Michigan*

Michael F. Schober, *The New School*

Kallan Larsen, *University of Michigan*

Andrew L. Hupp, *University of Michigan*

Video communication technologies (e.g., Zoom, WebEx, Skype) have the potential to provide survey researchers with all of the benefits of in-person interviewing, while at the same time greatly reducing data collection costs, given that interviewers do not need to travel and make in-person visits to sampled households. The COVID-19 pandemic has had a significant impact on in-person data collection operations, forcing survey researchers to explore alternative data collection modes, such as video interviewing, that do not require in-person interaction but still yield high-quality data. Given the potential benefits of these technologies, the operational and methodological aspects of video interviewing have started to receive research attention from survey methodologists. Video interviewing still involves respondent-interviewer interaction, however, which introduces the possibility of interviewer effects, and no research to date has evaluated this potential threat to the quality of the data collected in video interviews. We will present an evaluation of interviewer effects in a recent experimental study of alternative approaches to video interviewing, including "live video" interviewing and the use of pre-recorded videos of the same interviewers asking questions in a web survey format (recorded-video interviewing). We find that significant interviewer effects were rarely present when using these two approaches, which is a promising result for live video interviewing. We also find that when interviewer effects were present, they tended to be slightly larger in the live video mode, as would be expected. We will conclude by discussing the implications of these findings for future research using video interviewing.

Mini-Conference: Attitudes Towards COVID-19 Vaccination

COVID-19 and Vaccine Hesitancy: Confidence on a Knife Edge

Theebika Shanmugarasa, *ORB International*

Jonathan Heald, *ORB International*

Across the world, people and businesses are hanging their hopes on a COVID-19 vaccine which will enable them to return to 'normal' – but what if too few people are prepared to take a vaccine? Epidemiologists estimate a minimum uptake of 65% is needed to achieve herd immunity, yet our polling consistently shows that willingness to take the vaccine varies hugely by country, and is particularly low among certain subgroups. Could vaccine hesitancy - identified by the WHO in 2019 as one of the top 10 threats to health worldwide - prevent the world from breaking free of this pandemic?

ORB International has teamed with the London School of Hygiene and Tropical Medicine (Vaccine Confidence Project) since 2014, monitoring confidence in vaccines in more than fifty countries across the globe. Even before the outbreak of COVID-19, anti-vaccine narratives were undermining confidence in vaccination and research shows that dips in confidence are swiftly followed by declines in uptake, and outbreaks of disease. Last year alone 145,000 died of measles – a preventable disease via vaccination. We will present global data from recent studies to show how the pandemic is impacting confidence in vaccines, look at the impact disinformation has on take-up, and reveal which demographic groups are most susceptible to vaccine hesitancy.

Likelihood of Early Adoption of a COVID-19 Vaccine: Results From Six National Surveys

Rachel Kinder, *ICF*

John Boyle, *ICF*

James Dayton, *ICF*

Thomas Brassell, *ICF*

The spread of the novel SARS-CoV-2 virus has led to more than 200,000 deaths and over 7 million known COVID-19 cases in the U.S. by October 2020. Since containment is no longer possible, herd immunity is essential to impeding transmission of the virus and enabling a return pre-pandemic business, education and life in American. COVID-19 vaccination is expected to play a critical role in achieving herd immunity. However, multiple public opinion surveys suggest only moderate willingness among the public in getting a COVID-19 vaccine should one become available. This research finds less than a third of the public at the end of September were very likely to be early adopters of an FDA-approved COVID-19 vaccine (i.e., willing to get the vaccine as soon as it is broadly available). Further, the proportion who were likely to be early adopters has declined across five national surveys conducted between May and September 2020. While these surveys indicated that larger proportions of the public are willing to accept the vaccine within six months after release or after a second wave of infections, the findings suggest that immunization rates in the short run will fall far short of that needed to achieve herd immunity goals.

The study found a relatively strong relationship between likelihood of early vaccine acceptance and personal perception of risk of the disease as well as trust in government and public health agencies. There were also notable differences in likelihood of early vaccination acceptance based on respondent demographic characteristics. These findings suggest that multifaceted communication and education efforts responsive to the concerns of different groups will be needed to achieve immunization levels to achieve high levels of vaccination coverage with a safe and effective COVID-19 vaccine.

COVID-19 Vaccine Hesitancy: Skepticism Across Race/Ethnicity

Lunna Lopes, *Kaiser Family Foundation*

Liz Hamel, *Kaiser Family Foundation*

Mollyann Brodie, *Kaiser Family Foundation*

As the U.S. prepares for the approval and distribution of a COVID-19 vaccine, national surveys find a declining share of U.S. adults saying they would be willing to get vaccinated, presenting a unique challenge to policymakers and public health officials. Vaccination hesitancy among people of color is particularly important because they are bearing a disproportionate burden of the disease, and because of a range of barriers to vaccination among people of color, many of which are rooted in a historic legacy of abuse and mistreatment by the medical system and ongoing racism and discrimination. Our analysis of a series of probability-based national public opinion surveys, including one with a large oversample of Black adults, looks at levels of vaccine skepticism and distrust among U.S. adults across race and ethnicity and explores the different reasons driving COVID-19 vaccine hesitancy among Black, Hispanic, and White adults. Our surveys find that vaccine hesitancy among Black Americans centers largely around safety concerns and a distrust of the medical system, while many White adults who are hesitant to get vaccinated say that they simply don't want to get vaccinated or feel they are not at risk. As public health officials expand efforts to communicate the need for COVID-19 vaccinations, separate communication strategies will be needed to convince these groups to get vaccinated. This presentation will draw on the results of these 2020 surveys as well as forthcoming surveys to be conducted in 2021 to summarize the current state of COVID-19 vaccine hesitancy among U.S. adults by race/ethnicity and to identify communication challenges and potential solutions for public health officials and others seeking to increase COVID-19 vaccine acceptance in the United States.

Public Opinion on Coronavirus Vaccination 1: A Majority of Americans Would Take the Existing Russian Vaccine and Believe That They Ought to be Allowed to Buy It

Jonathan Kelley, *International Survey Center*

Mariah Evans, *University of Nevada, Reno*

Charlotte Corday, *International Survey Center*

In the US new vaccines are banned until shown to be safe and effective. But the approval process is slow and cautious. The faster Russian system produced an approved coronavirus vaccine months more quickly, leaving Americans at risk of dying for months longer than Russians. A Chinese vaccine was produced almost equally quickly. Our data from two national surveys in September show that a majority of Americans would willingly take the existing Russian vaccine and that a two-to-one majority – rich and poor, young and old, Democrat and Republican alike – believe that they ought to be allowed to do so. So too for the Chinese vaccine. We estimate that making the Russian or Chinese vaccines immediately available would save approximately 60 to 100 American lives each day after the first month and many more subsequently. To put the matter bluntly, current US government policy will kill some 60 to 100 people each day for a considerable period later this year and early next. To put those deaths in context, all American murderers combined kill only 45 people each day – not a record the US government should wish to emulate. In a pandemic, caution and protecting one's institutional reputation carry a high cost.

Coronavirus Vaccination Hesitancy in the Black and Latinx Communities

Steven Sparks, *Langer Research Associates*

Gary Langer, *Langer Research Associates*

Allison De Jong, *Langer Research Associates*

Christine Filer, *Langer Research Associates*

Surveys reveal disproportionate disinclination among Black people in the United States to be vaccinated against the novel coronavirus – a cause of substantial concern for public health professionals given the Black population’s high rates of infection, hospitalization and death from COVID-19. This study delves deeply into these compunctions, relying on an unusually large, random national sample of 1,050 Black Americans and a comparison sample of 258 Latinx adults. Supported by a literature review and consultation with vaccine uptake and health equity specialists, we investigate the roles of demographics, health care access and quality, racial and ethnic identities, experiences of discrimination and psychosocial factors relating to vaccines and the coronavirus. In statistical modeling, we identify key predictors of intended coronavirus vaccine uptake among Black and Latinx adults, finding that trust in the safety of the vaccine is by far the strongest predictor for both groups. Accordingly, we model trust in vaccine safety, finding that the leading predictor is confidence in the institutions involved in the vaccine production and delivery process. Finally, we analyze predictors of trust in the vaccine process; here, the two strongest factors are trust in the government to look out for the interests of Black or Latinx people and perceived social pressure to get vaccinated.

Mini-Conference: Using Data Science Methods for Public Opinion Research to Understand the Pandemic

The Impact of COVID-19 on the School-to-Work Transition: Evidence From Graduating College Seniors

Hannah McQueen, *North Carolina State University*

Sierra Satterfield, *North Carolina State University*

Due to constraints and uncertainties imposed by the COVID-19 pandemic, many individuals have recently faced an unusually turbulent job market. College graduates seeking to make the school-to-work transition in the first half of 2020 faced a shrinking number of job opportunities, coupled with other financial, social, and health-related concerns. This study examines the perceptions of May 2020 college graduates regarding the impact of COVID-19 on their post-college plans. In order to capture these narratives, we modified an existing annual survey that is administered to all college seniors at a large, public university in the weeks surrounding the May commencement period to include an open-ended question asking participants how, if at all, COVID-19 has impacted their future plans for work or study. Taking a two-pronged approach to analyzing the data, we first use computational text analysis methods in order to unearth common themes or challenges that graduates experienced. We find that the overwhelming majority of graduates who answered the question were negatively impacted and that many graduates personally dealt with retracted job offers or interviews. Second, we incorporate these narrative themes into a logistic regression model predicting likelihood of having confirmed plans for future work as opposed to those still actively seeking employment, controlling for field of study and demographic characteristics. Overall, this study illuminates the many pressures and frustrations that May 2020 college graduates experienced as they attempted to transition from school to work in the midst of a crisis. Furthermore, by augmenting a well-established survey and combining the survey data with administrative records, we demonstrate that even small changes to existing survey procedures can lead us to a more robust understanding of how job seekers fare in periods of economic uncertainty.

Explaining Consumer Expectations Using Big Data

Z. Tuba Suzer-Gurtekin, *University of Michigan*

Yingjia Fu, *University of Michigan*

Cali Li, *University of Michigan*

James Lepkowski, *University of Michigan*

Richard Curtin, *University of Michigan*

In this study, we empirically investigate properties of repeated cross-sectional estimators based on the conceptualization of twitter data as sampling of opinions over a period of time as opposed to conventional survey sampling of opinions from a selected respondent. University of Michigan's Surveys of Consumers are monthly telephone surveys using random digit dialing (RDD) cell phone frame. As a part of Surveys of Consumers, we have been collecting data on the most concern impact of covid-19 using a closed-end question since April 2020. While consumers consistently report impact on health as the most concern, we observed a growing concern of impact of social isolation. We apply the same codeframe to twitter data in order to track and test the change in opinions over time at national level. As part of our study, we test alternative weighted estimators on a sample of tweets.

Comparing Sensor-Based Pedestrian Flows With Survey Estimates of Pandemic-Preventing Behavior

Sven Alexander Brocker, *University of Duisburg-Essen*

Jonas Klingwort, *University of Duisburg-Essen*

With the outbreak of the corona pandemic, surveys were launched to measure the population's attitudes and behavior concerning visiting public places during the corona pandemic. However, such questions are likely biased due to social desirability. As survey estimates can be used to evaluate pandemic-preventing measures, their accuracy is of high importance.

In this application and as a robust start, we aim to analyze correlates in survey and sensor-based time series of people's attitudes and behavior concerning visiting public places. We focus on periods in which new measures were introduced to limit the spread of the virus. This is demonstrated using data of five different web surveys measuring population's attitudes and behavior in Germany. The development of these estimates will be compared with time-series of sensor-based measures of pedestrian flows. Sensors are installed in over 60 different cities at over 130 highly-frequented public places and streets. These automated systems count the incidence of pedestrians with 99% accuracy. Such data is typically used for investors, retailers, or city planners but was so far rarely used to obtain insight into population behavior concerning compliance with pandemic measures. The sensor and survey observations can be linked on a daily-level using the timestamp. Preliminary results show observed trends being similar in both sources. However, since the pandemic and data collection continues, results will be continuously updated.

From this research, we conclude that linking survey and sensor data of macro-phenomena has the potential of being a promising tool to confirm trends in survey estimates. Further, sensors measuring pedestrian flows provide data on a population's behavior much faster and more frequent than surveys. Therefore, it could be considered as an additional data source in critical situations, such as a pandemic.

Using Digital Trace Data to Take the Pulse of Parenting During the COVID-19 Pandemic

Claire Kelley, *Child Trends*

Sarah Kelley, *Child Trends*

Isai Garcia-Baza, *Child Trends*

During the early months of 2020, the COVID-19 pandemic presented a rapid, ever changing challenge to parents. Parents worried about increases in screen time, brainstormed strategies to help their children with online learning and wondered about the risks of teens vaping. As part of a NSF funded communications science project, we used digital trace data (from Facebook, Twitter and Google Trends) to identify trends in public opinion in the US on topics surrounding COVID and parenting, among both English and Spanish speaking parents. This research builds on a growing body of research which uses digital trace data (especially from Google Trends which provides anonymous data on search term frequency, and Twitter) to study such hard-to-measure topics such as overt racism, child abuse, and intimacy (Chae 2018; Stephens-Davidowitz, 2013; Stephens-Davidowitz, 2017). Such digital trace data may offer unique advantages to investigate questions plagued by social desirability bias, or requiring near real time measurement. We primarily draw on quantitative analysis of Google Trends data, qualitative review of Facebook parenting groups, and a natural language processing analysis of Twitter to identify the top concerns parents have relating to parenting during the pandemic. Drawing on results in both English and Spanish, we find that parents had sustained interest in COVID and parenting, with certain topics (such as vaping, and online learning showing significant variation over the course of the study). In addition the quantitative data collected from Twitter revealed interesting differences between the concerns of Spanish-speaking and English speaking parents, and a shift in parental concerns over time. The results of this study were used to help design educational videos that were shared with parents as part of a series of tv programs on positive parenting.

Hope or Despair? Sentiment Analyses of Tweets and User Engagement Generated by US companies Developing COVID-19 Vaccine

Priyanka Khandelwal, *West Texas A&M University*

COVID-19 has evoked an unprecedented public health crisis throughout the world. Major investments are made across the globe to develop effective vaccine against SARS-CoV-2 virus, with several candidates appearing in Phase III clinical trial. Four US pharmaceutical and biotechnology companies have made promising progress in the development of this vaccine. These organizations have routinely used Twitter to keep the public informed about the status of their product and shared general scientific information about COVID-19. Some of the tweets generated by these companies garnered high level of user engagement – as measured through retweets and favorites. However, given that no vaccine has been formally approved by FDA (as of November 2020), it is not surprising that a pair of repeated surveys conducted by Pew Research Center (PRC) found that US population is becoming increasingly skeptical about COVID-19 vaccine.

This study examines how PRC's findings are reflected in the Twitter communications of the foregoing four US companies. In particular, matching the period covered by PRC's polling, we examine the evolution of sentiments associated with the tweets generated by these organizations during the April-May 2020 and September-October 2020 to understand if the rhetorical structures of the tweets show significant difference between these two periods. Additionally, we analyze the association between user engagement and sentimental category of the tweets to assess whether there is a significant change in the association structure as well.

5/11/2021, 2:00 PM - 3:00 PM
Round Tables: Live Sessions

Inclusive Language and Practice in Cross-Cultural and Multilingual Research

Inclusive Language and Practice in CrossCultural and Multilingual Research

Mandy Sha, *www.mandysha.com*

Patricia Goerman, *U.S. Census Bureau*

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

Cross-cultural research in the United States and around the world contributes to a greater understanding of behaviors and opinions among people. AAPOR dedicates the 3MC track to the growing area of multinational, multiregional and multicultural research and the Cross Cultural and Multilingual Research Affinity group cultivates the space for researchers and practitioners interested in these topics. We will hold a roundtable discussion during the virtual 2021 AAPOR conference to (1) facilitate a dialogue about improved methods, measurement, and research & publication opportunities; and (2) provide direct interaction opportunities to advance common interest in cross-cultural and multilingual research. All AAPOR attendees are welcome to be part of this conversation and join the affinity group.

'New to AAPOR' Round Table: Welcome to AAPOR! How Has COVID-19 Affected Your Research, Work, or Graduate Study?

'New to AAPOR' Round Table: Welcome to AAPOR! How Has COVID-19 Affected Your Research, Work, or Graduate Study?

Holly Hagerty, *NORC at the University of Chicago*

Benjamin Schapiro, *NORC at the University of Chicago*

Erin Fordyce, *U.S. Census Bureau*

Jonathan Katz, *U.S. Census Bureau*

Are you new-ish to AAPOR? Welcome! Join with a mix of seasoned and newer AAPOR members for an informal roundtable discussion, structured around the following discussion questions:

1. What brings you to AAPOR? -- Introductions of round table leaders and participants.
2. Introduction to AAPOR – Overview of AAPOR's structure, functions, membership, committees, and activities.
3. New Normal – How has COVID-19 affected your research, work, and/or graduate study? From helping projects move to remote data collection, re-orienting academic research plans, or using Zoom to teach classes or conduct cognitive interviews, there's no doubt that COVID-19 has disrupted our world in myriad ways. Let's talk about what we've done and what we've learned so far.
4. How has COVID-19 affected your networking and/or job search, or your current work?
5. Participating in the virtual conference – Do you have any questions about AAPOR membership and benefits, networking with other AAPOR members, volunteering with AAPOR, or feedback so far on the virtual conference?
6. Open discussion

5/11/2021, 2:30 PM - 4:30 PM
Concurrent Session B: Prerecorded Sessions

Mini-Conference: Panel Survey Design and Implementation During the Pandemic

The Mannheim Corona Study - Design, Implementation and Data Quality

Ulrich Krieger, *University of Mannheim*
Carina Cornesse, *SFB 884, University of Mannheim*
The MCS Research Group, *SFB 884, University of Mannheim*

The outbreak of COVID-19 has sparked a sudden demand for fast, frequent, and accurate data on the societal impact of the pandemic. To meet this demand quickly and efficiently, within days of the first containment measures in Germany in March 2020, we set up the Mannheim Corona Study (MCS), a rotating panel survey with daily data collection on the basis of the long-standing probability-based online panel infrastructure of the German Internet Panel (GIP). In a team effort, our research group was able to inform political decision makers and the general public with key information to understand the social and economic developments from as early as March 2020 as well as advance social scientific knowledge through in-depth interdisciplinary research.

This presentation gives insights into the MCS methodology and study design. We will provide a detailed account of how we adapted the GIP to create the MCS and describe the daily data collection, processing, and communication routines that were the cornerstones of our MCS methodology. In addition, we will provide insights into the necessary preconditions that allowed us to react so quickly and set up the MCS so early in the pandemic. Furthermore, we will discuss the quality of the MCS data in terms of the development of response rates as well as sample representativeness across the course of the MCS study period.

Setting Up a Mixed-Mode Probability-Based Panel Over the COVID-19 Pandemic

Ricardo Gonzalez, *Universidad Adolfo Ibañez*

Chile has a strong reliance on face-to-face (F2F) surveys, even for electoral polls. They usually provide high-quality data, but face challenges in terms of coverage, increasing nonresponse rates and costs. Importantly, they are not possible to field when events disrupt the regular functioning of major cities, such as the social crisis in October 2019 and the COVID-19 pandemic since March 2020. Chile has been particularly affected by the pandemic. In August 2020, Chile was top 5 in terms of cases per million across the globe. To curb the contagion, the government set night-time curfews and targeted and dynamic lockdowns at the municipality (comuna) level, among other policies. In July 2020, roughly 60 percent of the population was on lockdown. Since early August lockdowns have been over in the center of the country, but new lockdowns have been put in place in the northern and southern regions. Thus, it has been impossible to field nationally representative F2F surveys to this day.

This presentation describes the processes, challenges and outcomes of setting up a mixed-mode probability-based panel in Chile over a pandemic. It presents the major deviations in terms of design from established longitudinal surveys in developed countries. Respondents are selected through probability sampling by random-digit dialing (RDD) of cellphone numbers across all regions and are asked to join the panel at the end of the interview. Online and offline households are included in the panel. Subsequent waves are conducted in a mixed-mode fashion, i.e. by web and phone. Recruitment, response and retention rates are analyzed under the two different incentive schemes: gift cards conditional on response and after collecting a certain number of points. The representativeness of the panel is assessed in each step. Differences across modes in substantive topics are discussed and lessons for developing countries are drawn.

Accelerating Mixed-Mode Recruitment in a Television Measurement Panel

Hani Zainulbhai, *Nielsen*

Lauren Walton, *Nielsen*

Leah Christian, *Nielsen*

Mixed-mode contact and recruitment approaches have become increasingly critical in reaching people to participate in surveys and research panels. In particular, the use of multiple contacts and different modes of contact increases the likelihood of response by improving the number of people who receive and attend to the information request. Offering multiple modes in sequence for people to respond to a survey can improve response over a single mode. Lastly, multiple modes of contact and recruitment improve the representation of the sample, bringing in groups more traditionally hard to reach and thus often under-represented (e.g., young people, racial and ethnic minorities).

Nielsen has designed a phased sequential multi-mode methodology to recruit homes through a combination of web, phone and in-person/proximity methods for its core TV panels that have relied exclusively on in-person recruitment. This work has been accelerated due to the COVID-19 pandemic. This paper will present the results of a web recruitment test where Nielsen mails invitations to prospective panelists, directs them to complete a registration survey online, and follows up over the telephone for Nielsen meter sign up. Addresses were selected using a multi-stage area probability design in five markets: Chicago, Jacksonville, Atlanta, New York and Oklahoma City. The test period is from October 2020 to January 2021. Each selected address will be contacted by mail, with multiple contacts sent to each address to maximize the possibility of reaching potential panelists. These contacts include a: (1) Brand awareness postcard, (2) Large mailer with web invite and small cash incentive, (3) Mail reminder for those who did not complete after initial invite, and a (4) Final reminder for hard to reach demographics. The paper will discuss cooperation rates, number of completes per mailed invitation, market level differences, and paradata used for survey monitoring.

Summing Up the Ads... Exploring Social Media Recruitment Efforts for a COVID Surveillance Panel Study

Trent Buskirk, *Bowling Green State University*

Lorene Nelson, *Stanford University*

Jessica Hinman, *Stanford University*

Barbara Barclay Topol, *Stanford University*

Patrick Arensdorf, *Stealth Biotech Startup*

Survey response rates continue to decline as survey costs continue to increase. With the emergence of the fourth era of survey research we are seeing a broader appeal for recruitment methods that extend the fit for purpose paradigm with more rapid, cost effective and targeted methods. Social media recruitment continues to be explored as one such alternative, either alone or as a supplement to probability-based sample designs. During a global pandemic, measures that afford researchers safe and effective means of extending probability-based recruitment and minimize the need for field staff interactions either with respondents or with other workers are particularly relevant.

Much of the work to date on social media recruitment has focused on cost and yield. In this study we extend this focus by investigating the performance of several targeted ads placed on Facebook and Instagram for recruitment and retention into the Community Alliance to Test Coronavirus at Home (CATCH) Study launched in Fall 2020. This panel study uses a stratified address-based sampling design to select a probability sample of households within the greater San Francisco Bay area. The core probability sample is augmented by a parallel non-probability convenience sample shaped using social media ads.

We report the results of testing multiple versions of targeted Facebook ads varying images and key components of messaging in order to recruit frequently underrepresented male, Hispanic and older populations. In the presentation, we explore recruitment costs together with click through, enrollment, participation, and attrition rates for the non-probability arm of the study. This research contributes to our

growing understanding of social media recruitment by examining not only advertisement yields but also the corresponding attrition rates within a broader probability and non-probability combined panel study.

Building a Probability-Based Online Panel in the Time of COVID-19

Lee Hargraves, *University of Massachusetts - Boston*

Recruiting participants for a probability-based online panel can follow multiple paths. In a probability-based survey asking about the effects of the COVID-19 pandemic, we asked respondents to join an online panel to conduct additional surveys. We expected that the pandemic would make the survey salient to city residents and we would see relatively high response rates and enthusiasm for joining. Using an address-based sample stratified within 25 Boston neighborhoods, we mailed an 8-page questionnaire asking respondents to join a panel and receive additional surveys. The initial survey included questions about COVID-19, employment, and health. To encourage response, we included \$2 with the invitation and information about city government partners. In neighborhoods where more than one-third of residents were Hispanic, we sent surveys and other materials in English and Spanish.

After sending a reminder letter with a URL and QR code that linked to a web-based survey and, two weeks later, another mailing of questionnaires, 1370 residents responded (1208 completed a paper survey and 162 completed online). Neighborhood-level response rates ranged from 17% to 40% with a citywide response rate of 27%. The last page in the questionnaire described the panel and stated that participants would receive a \$5 gift card after completing future surveys, 586 people provided their email addresses to join the panel, a cooperation rate of 43% for the panel. We explore variation in volunteering by resident's age, gender, education, and ethnicity. Moreover, we examine the extent to which people who are worried about coronavirus infection, have financial stress, engage in coronavirus mitigation efforts, and have been infected signed up for the online panel. Future surveys will use paper and web-based surveys to track the effects of COVID-19 on people's lives.

Polling in the Time of COVID

Jenny Marlak, *Gallup*

Jeff Jones, *Gallup*

For much of 2020, many Americans could be found at home avoiding the coronavirus, and workplaces rushed to implement work from home and other safety policies. At the same time, the pandemic, the election, and social justice movement produced a heightened interest in social issues in the United States. These events have presented unique data collection challenges for many organizations. Organizations have had to re-evaluate how to execute survey work in this environment, and whether teams, such as interviewers or mailroom teams, can safely execute work. In this paper, we will briefly share some the COVID related data collection challenges Gallup experienced in 2020 and how these challenges were addressed.

Researchers have also been concerned about non-response bias. With more individuals at home because of stay-at home orders or remote work, people may be more likely to answer the phone or open the mail. Alternatively, many Americans are reporting increased stress and mental fatigue, despite more time at home, which could negatively affect willingness to participate in a survey. Further, for some groups, events as social justice movements or the election, could increase interest in sharing opinions about these issues. In March 2020, Gallup saw an increase in telephone response rates, achieving a rate not seen in close to 10 years. Since then, response rates have remained elevated, averaging nearly two percentage points higher than in 2019. The increase is driven mostly by an increase in cooperation rates, though contact rates are up slightly and refusal rates down slightly from 2019. We will also explore the unweighted demographics of participants. Further, we will share response rate data from our Gallup Panel web surveys, which have also, overall, had strong participation. Finally, we will share data from our ABS mail push to web surveys, which have seen slight declines in participation.

Call Us! Recruiting by Phone for an In-Person Longitudinal Survey in Response to the Coronavirus Pandemic

Megan Bjorgo, *NORC at the University of Chicago*
Nathaniel Poland, *NORC at the University of Chicago*
Alanah Raykovich, *NORC at the University of Chicago*
Jennifer Vanicek, *NORC at the University of Chicago*

The Medicare Current Beneficiary Survey (MCBS) serves as the leading source of information on the Medicare program and its impact on beneficiaries. As a continuous, multipurpose survey of a nationally representative sample of the Medicare population, a new panel of beneficiaries is selected to join the survey every fall. For nearly 30 years, respondent outreach to recruit the new panel has been conducted entirely in-person as the sample is drawn from Medicare administrative data that contains beneficiary location information, but not phone numbers. The survey itself is conducted via computer-assisted personal interviewing. In response to the coronavirus (COVID-19) pandemic, MCBS rapidly shifted from in-person data collection to phone data collection beginning in March of 2020. Immediately, strategic planning began for outreach to new panel members set to join the survey in July. This presentation will provide an overview and evaluation of the strategic processes implemented to successfully recruit a new panel of participants in the fall of 2020 entirely by phone.

Prior to 2020, outreach to newly sampled beneficiaries consisted of an advance mailing and in-person visits to attempt to conduct the interview. Having pivoted to phone-only data collection, the team rapidly developed and implemented a number of mail strategies to recruit the new panel. Enhanced mailings to respondents included reminder, last chance and refusal letter mailings; the language utilized in the letters prompted panel members to call the survey toll-free number for more information and to schedule an interview. Ultimately, the recruitment effort was largely successful in meeting targets. This presentation will describe the respondent outreach protocol, analyze the results of these new mail interventions, present qualitative feedback from interviewers, and explore ways of incorporating new modes of respondent outreach for in-person surveys in the future.

Changing Response Patterns in Lockdown: Analyzing a Natural Experiment During Panel Recruitment

Jennifer Su, *SSRS*
Josh Pasek, *University of Michigan*
Kenneth Winneg, *The Annenberg Public Policy Center of the University of Pennsylvania*
Chintan Turakhia, *SSRS*
Kathleen Hall Jamieson, *The Annenberg Public Policy Center of the University of Pennsylvania*

As the coronavirus pandemic first spread in the United States in March of 2020, many Americans were implored to stay home. Anecdotally, researchers have noted that individuals may have been more willing to participate in survey research during this time. Yet, survey researchers have also hypothesized that the individuals who were predisposed to participate during this period may have differed in some systematic ways from those who did not do so or from those who participated before the mid-March pandemic lockdowns or restrictions.

In this study, we examine how two phases of recruitment into a large address-based sample differed before and after the pandemic hit to understand the scope and implications of these changing response rates. Five sets of invitations were mailed to individuals recruited to a large online panel study between November of 2019 and March of 2020. In this study, we compare two samples of respondents recruited using identical techniques with mailers sent either on Wednesday, February 12, 2020 or on Wednesday, March 18, 2020. We compare the two samples in terms of response rates based on empirical data, as well as the demographic attributes of respondents to identify notable differences in recruitment composition.

Political Participation and Opinion Across Regions and Battleground States

Investigating Mail-In Ballot Effect on Split Ticket Voting

Edward Johnson, *Dynata*

Justine D'Elia-Kueper, *Catalist*

Richard Cohen, *Dynata*

The 2020 election was significantly affected by the pandemic. One obvious way the pandemic affected the election was through the voting process. Early and mail-in voting was historically increasing, but still under 25% in previous elections. The mail-in ballot voting is coming in much higher, maybe even over 50% once all the results come in.

Being able to complete your ballot at home at your convenience is likely a much different experience than at a polling location. Voters are given more time to complete and less pressure to decide. This environment could lead to more research before voting and less use of simplifying rules such as party identification in the voting process. This could result in an increase in split ticket voting which might explain the Democrats underachieving at the local level in the 2020 election.

We will interview 1,000 voters in 2020 from Dynata's Voter Matched panel connected to Catalist's Voter File. In the survey we will investigate the following items:

1. Their voting environment and experience
2. The research they did before or during the voting process
3. Compare stated to observed method of voting in the 2020 general election
4. Did they submit a straight party ticket or split party ticket
5. Factors that contributed to their voting method decision
6. Satisfaction with voting method
7. Which voting method they intend to use in the future assuming it is available.

We hope that those attending this session will be able to take away insights into how the voting experience itself might have altered the voter's decision making process from research all the way through satisfaction with the voting process. We will then provide suggestions in how online voter polls might better mimic the process the voters will go through when deciding how they will vote.

NY Times Upshot/Siena College Research Institute Polls: Counting the Calls - A Time Series Analysis of Polling Productivity in Battleground States

Travis Brodbeck, *Siena College Research Institute*

Don Levy, *Siena College Research Institute*

Meghann Crawford, *Siena College Research Institute*

How much more difficult is it to poll in battleground states today than it was in 2016 or 2018? The New York Times/Siena College in conjunction with its partner call centers project called thousands of phone numbers across nine weeks to talk to more than 30,000 likely voters preceding the 2020 Presidential Election. The cost of live interview polling was already under pressure before the pandemic, yet now with strict capacity constraints, PPE requirements, and remote challenges, organizations are facing a variety of new factors that need to be addressed since they cut into the overall project's budget. Polling productivity may vary for a variety of reasons, but over time we can see how much weight those factors carry. To understand tomorrow's environment, we will look at past performance and utilize that data in strategically planning for the future.

Many of the NY Times Upshot/Siena College Research Institute polls were conducted with the help of multiple call centers which allowed calling to continue throughout the 9-week period. The collaborating and coordinating call centers maximized the performance of the project and allowed it to be so successful in reaching so many voters in such a short amount of time. By understanding the past performance, we can begin planning for the Georgia runoff elections, midterm elections, and the next presidential elections. This constant monitoring is necessary so pollsters can budget enough time and money to adequately reach the electorate in these key states.

2020 Presidential Election Polling in Counties in Northern New York – Weighting Adjustments to Adequately Reflect Trump Support

Joel LaLone, *Center for Community Studies at Jefferson Community College*

The Center for Community Studies at SUNY Jefferson has completed over 150 community-based studies with wide variety of research goals for agencies and organizations in rural Northern NY since 1999. The Center was accepted into the AAPOR Transparency Initiative in 2019. The typical research goals of clients of the Center do not often include political polling, however, in General Election years the Center includes election-related questions in their annual quality-of-life adult surveys as a means of assessing the sampling methodology and mathematical modeling currently employed. In the final week of October 2020 the Center used a mixed-mode methodology (online surveys and live telephone interviews from a virtual remote call center) to survey approximately 400 LV's in each of four counties. Using the standard weighting algorithms from recent-past Center surveys (weight for age, gender, education, party, and modality) the Presidential Election predictions in each county all fell well within the polling Margin of Error, however, all consistently slightly underestimated Trump support. The survey instrument used included party affiliation, voting likelihood, and racehorse questions, along with four potential items that could be considered proxy measures of distrust and denial of government, media, science, and polling. Post-election analysis of, and further weighting for, the underrepresentation of these low-trusters provided remarkable results. The presentation will overview: adjusted estimates which better-reflected the prevalence of this group of apparently underrepresented individuals; the mathematical methods used to model these new estimates; that which was learned from this post-mortem polling exercise; and how these results will impact the next four years of sampling methodology studying non-voting-related community topics for organizations in Northern NY (e.g. education, healthcare, economic development, strategic planning, program impact). Trump supporters in NNY do not appear to be shy, but appear to have very low likelihood to believe in and agree to complete a survey.

2020 Election Dynamics: Danger in Trying to Typecast Women Voters in Battleground States

Debbie Borie-Holtz, *Rutgers University*

Women participated in the 2020 presidential election at levels higher than men and national exit polls estimated a 15-point gender gap favoring Joe Biden. While female participation was higher in the battleground states, support for the Democratic ticket varied by region, race, ethnicity, and a family's financial situation. To study the intersection gender and race played, we surveyed roughly 600 women in five battleground states (Florida, Georgia, Nevada, North Carolina and Pennsylvania) in the 10 days leading up to the election. The questions we asked centered on the historic nature of the democratic vice-presidential selection, the ideal traits for a president, the need for a smooth transfer of power, and limitations on political rhetoric.

We found an equal proportion of women said Biden's choice for vice president influenced their decision to vote for the Democratic ticket this year compared to those who said it made no difference. Twice as many women voted for the Democratic ticket as compared to those who said Biden's choice made them less likely to do so. The difference among Black women in North Carolina and Georgia was even more lopsided in favor of the Biden-Harris slate. Among the traits women were looking for in a president, a leader who was compassionate, hardworking, empathetic, and honest topped the list. Overall, respondents said Joe Biden best matched this profile.

We also assessed the impact the pandemic, ranked as the most important issue facing the country, was having on families and personal finances, as well as the length of time those we surveyed thought it would take their state to get back to normal. A comparison between a panel of women voters in 2016 and 2020 is planned for trend questions asked of our panelists concerning sexual harassment, racial discrimination, immigration, and race relations.

Partisanship and Hot-Button Issues

Generations Y and Z: Twenty Years of Political Attitude Trends Among Young Adults

Chase H. Harrison, *Harvard University*

John Della Volpe, *Harvard Kennedy School Institute of Politics*

Justin Tseng, *Harvard College*

Public opinion scholars often focus on importance of the generational cohorts in shaping public opinion. Scholars of political socialization find that attitudes formed in early childhood, and shaped during young adulthood, remain largely stable and enduring over subsequent years, suggesting that long-term change in attitudes is determined more strongly by intergenerational replacement of values and attitudes rather than short-term changes. Thus, understanding the distinctive attitudes of early generational cohorts can provide a sort of view to the future dimensions of ideology and attitudes. This paper will summarize important trends from twenty years of surveys conducted with young Americans between the ages of 18 and 29, beginning when the oldest “millennials” (Generation Y) entered adulthood (2000), and including data from the youngest adult members of the generational cohort commonly called Generation Z.

Data will be drawn from 20 years of biannual national surveys conducted by the Harvard Kennedy School Institute of Politics Youth Poll between 2000 and the spring of 2021 using the IPSOS KnowledgePanel (since 2009), HarrisOnline (2006-2008), and RDD telephone surveys (2000-2005). Supplemental data from national focus groups, conducted both online (post-COVID) and in-person (pre-COVID) across a variety of states will be used to discuss nuances in the findings. We will focus on four key sets of areas where we have observed significant changes in attitudes among young adults over twenty years: the role of government, climate policy and issues, violence and gun regulations, and political engagement.

Attitudes in 2020 Towards Medical and Recreational Marijuana in Prohibitionist Nebraska

Patrick Habecker, *University of Nebraska-Lincoln*

Rick Bevins, *University of Nebraska-Lincoln*

The legal status of marijuana is rapidly changing in the United States. After the 2020 election, all but two states have legalized some form of medical or recreational use of the substance. In this presentation, we examine public opinion about marijuana in Nebraska, one of the last states to maintain full prohibition of marijuana. Using an addressed-based sample of Nebraskans, we examine attitudes about medical and recreational marijuana laws at both the state-wide and sub-state levels. Within the state, we compare differences between the two largest Nebraskan cities (Omaha and Lincoln) and also between multiple rural areas that match the behavioral health regions of the state. Our subdivision means we can compare different rural areas instead of considering all rural areas to be alike. This is particularly important as Nebraska shares a border with Colorado. In 2012, Colorado became the second state to legalize recreational sales and use of marijuana, and until the 2020 election, was the only state to border Nebraska with recreational marijuana laws. The shared border with Colorado may be associated with differences in legal attitudes in Nebraska. We also consider how a respondent’s own personal history with marijuana and other substances, and their attitudes towards people who use drugs in general are associated with legal attitudes around marijuana. In addition, we also examine the associations with age, gender, political affiliation, and education. Despite 48 states minimally passing medical use laws, and 15 states and the District of Columbia passing recreational use laws, marijuana remains illegal to use federally and remains a Schedule 1 drug for the FDA and DEA. National legalization of marijuana is not a guaranteed outcome. Our focus is on differences in attitude in a prohibition state, but these attitudes are likely to be indicative of other Midwestern plains residents.

To Tax or Not to Tax: Conservatives and Liberals' Differing Attitudes on the Rich

Michelle Io-Low, *Stony Brook University*

Economic inequality has quickly become a hot button issue over the recent decades. Alongside this discourse, scholars have consistently documented rising economic inequality. This documentation of economic inequality is accompanied by the public's expressed concerns about it. Although people have reported concerns about inequality, their support for redistribution remains largely unaligned to these expressed concerns. This is also commonly referred to as the inequality paradox. In the present work, I argue that one reason we continue to observe the inequality paradox is because conservatives and liberals are split on their opinion to redistribute from the rich. This split in opinion to redistribute from the rich arises because conservatives and liberals are guided by different forces when they form opinions of the rich. Conservatives are more likely to be guided by stereotypes of competence and warmth – how capable/incapable the rich are and how warm/cold the rich are) whereas liberals are unaffected by these stereotypes. I utilize a series of observational datasets along with an experiment to demonstrate this theoretical argument. These findings contribute to the extant literature by showing that attitudes towards the rich is yet another important component to understanding the public's redistributive preferences and policies that mitigate growing inequality.

Angeleno Opinions on Climate Change by Partisanship

Alejandra Alarcon, *Center for the Study of Los Angeles*

In the past year, Los Angeles has seen record-setting heat waves, massive wildfires, poor air quality, and disappearing coastlines all indicative of a larger trend: climate change is affecting everyone around the world. However, not all residents agree on its impact and it has even become a polarized issue. While 94 percent of liberal Angelenos believe that climate change will affect future generations of people, only 68 percent of conservative Angelenos believe it to be the case, according to the 2020 Los Angeles Public Opinion Survey. This research examines Angeleno opinions on the effects of climate change leading up to the 2020 presidential election cycle. Specifically, this research gauges whether or not Angelenos think climate change will affect them personally, people in the Los Angeles region, people in the U.S., or future generations with a focus on differences in opinion by political ideology before and after the 2020 presidential election. Data from the 2020 Los Angeles Public Opinion Survey of 2,000 adult Los Angeles County residents conducted by the Center for the Study of Los Angeles at Loyola Marymount University are analyzed. The annual survey, conducted every January, is a mixed-mode telephone and online survey available in English, Spanish, Mandarin, and Korean. The goal of this research is to understand differences in opinion on this important issue.

Partisan Cues, Framing, and Public Support for NATO

Kirby Goidel, *Texas A&M University*

Kyung Suk Lee, *Texas A&M University*

Considerable research has been devoted to understanding how Americans think about international affairs. While not well informed, Americans beliefs about international relations are structured around a broader set of values, including general orientations about whether the United States should play an active role in the world, the use of military force as a tool of statecraft, and unilateral versus multilateral approaches to international affairs. Attitudes toward specific international organizations (NATO) reflect these general orientations as well as elite framing of the issues at stake and partisan cues signaling support or opposition. In this paper, we ask how President Donald Trump's "America First" rhetoric has reshaped the structure of American foreign policy beliefs. Trump's emphasis on isolationism, militarism, and unilateralism presents a sharp challenge to existing partisan divisions while his transactional approach to foreign policy reframes U.S. alliances away from the security benefits of these alliances provide and toward the financial costs these alliances incur. We hypothesize these changes alter the foundations of public support for the NATO alliance. To test this hypothesis, we use two sources of data. First, we use survey data from the Chicago Council Survey of American Public Opinion and US Foreign Policy from 2010-2019 to model public support for NATO. This allows to test how the foundations of public support have shifted over time. Second, we conduct an experiment to capture how the reframing the U.S.- NATO alliance to emphasize costs and partisan cues alter public support for American

allies. Overall, we find that while public opinion, framing effects, rather than partisan cues, better explain this shift.

When Mass Opinion Goes to the Ballot Box: A National Assessment of State Level Issue Opinion and Ballot Initiative Results

Jonathan Robinson, *Catalist*

Chris Warshaw, *The George Washington University*

John Sides, *Vanderbilt University*

In this paper, we will develop estimates of public opinion in each state on a host of salient policy issues using multilevel regression and post-stratification (MRP) models to downscale the results of a large scale collection of national surveys. We will estimate state-level public opinion for all adults and for registered voters on each issue. We will then compare public opinion on individual policy issues to the results of ballot initiatives on these issues (e.g., minimum wage, background checks for gun purchases, same-sex marriage amendments, etc). First, we will evaluate the accuracy of issue polling on the issues put on the ballot in predicting ballot initiative results, for example, comparing opinion on minimum wage increases in a state to the results of a specific minimum wage initiative. This will provide a novel tool to assess the mean error in issue polling. Second, we will assess patterns in polling errors on policy issues. We hope to provide an assessment of whether there is a liberal or conservative bias in issue polls compared to initiative results. We also hope to develop theories of how bias in issue opinion compared to ballot initiative results varies based on factors such as when in the broader election cycle an initiative takes place. For instance, we expect there to be a liberal-bias in issue opinion in years when there is a Democrat in the White House (and thus Republicans have a turnout advantage) and a conservative bias in years when there is a Republican in the White House (and thus Democrats have a turnout advantage).

Prototypicality of Alleged Sexual Harassment Victim and Perpetrator Advocates

Jessica Roden, *University of Michigan*

Josh Pasek, *University of Michigan*

Social psychological research has demonstrated that in-group members perceive themselves to be more prototypical of a positively valued superordinate group than out-group members. Our survey experiment, using a two-wave panel study on a broad national sample of Americans (N = 706), tests this in-group projection effect with the politically-relevant yet less politically-sorted issue of sexual harassment allegations to determine the consequences of this phenomenon in a political communication context. Participants were randomly assigned to see a mock news article with a quote from either a Democrat or Republican that either defended alleged female victims or male perpetrators. Among other measures, participants were asked their partisanship, if they view themselves as a supporter of women who come forward with sexual harassment claims or of men who have been accused of sexual harassment, and how prototypical of these supporters they found the person quoted in the article. Controlling for wave one attention and attitudes toward sexual harassment allegations, prototypicality was significantly influenced by the interaction between whether the partisanship condition matched the participant's partisanship and whether the defending alleged victim or perpetrator condition matched the participant's relative identification with supporting alleged victims and perpetrators. Those who saw a quote from someone who either shared both or neither of their political party and attitudes about alleged victims and perpetrators were evaluated as highly prototypical speakers. In contrast, those who saw a quote from someone who possessed only one of these two traits in common with the participant were rated as less prototypical speakers. These results highlight how individuals see those in the supporting alleged victims or perpetrators superordinate groups as typical of that group only when they possess relevant characteristics that the individual also possesses, demonstrating the relevance of in-group projection for the political issue of sexual harassment allegations.

A Vulnerable Moving Target: Increased Polarization in US Foreign Policy Opinions Since 9/11

Ryan Tully, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

US government agencies have warned that future confrontations between the US and major powers will increasingly involve the use of hybrid warfare tactics, including political persuasion and disinformation campaigns. These tactics aim to diminish US global influence by inflaming internal political discord and, thereby, creating incoherent governmental responses to international developments. Democracies are most susceptible to these tactics when there are elevated levels of internal political polarization.

Following World War II, US foreign policy was long viewed as the ‘water’s edge’ of partisan politics. While there were disagreements over specific policies, polling data regularly showed overwhelming bipartisan support for general US foreign policy goals. Recent research suggests that ongoing media fragmentation may be producing increasingly polarized partisan views of US foreign policy. This study examines if partisan foreign policy views have changed significantly since the September 11 terrorist attacks and if these changes have resulted in more polarized partisan views toward foreign policy issues.

To accomplish this, we analyzed data from annual surveys conducted by the Chicago Council on Global Affairs from 2002 through 2020, which provide repeated measures of attitudes toward general foreign policy goals and support for specific policies. These yearly cross-sectional surveys utilized samples from KnowledgePanel®, the largest probability-based online panel in the U.S. Our analyses evaluate both current and longitudinal trends in foreign policy views across a variety of characteristics, including political ideology, party affiliation, and party extremity. The results show that partisan views on general foreign policy goals and specific policies have changed dramatically since 2002. The analyses further demonstrate increasingly polarized partisan views toward foreign policy issues in recent years. We discuss our results in the context of US vulnerability to foreign political persuasion and disinformation campaigns and how partisan sorting may be leading to increasingly polarized foreign policy attitudes.

Experiments to Optimize Incentive Use

What You See Is What You Get: Evaluating the Use of Visible Incentives in the California Health Interview Survey

Susan Sherr, *SSRS*

Brian Wells, *UCLA Center for Health Policy Research*

Research has demonstrated that the inclusion of pre-incentives with survey recruitment mailings is beneficial in increasing response rates. There is significant expense associated with the use of pre-incentives, particularly for large-sample surveys that send out hundreds of thousands of invitation letters with small cash incentives. If the respondent is not motivated to open the envelope and, consequently, see the incentive, these resources are wasted. DeBell et al. (2019) showed that a windowed envelope with a visible \$5 cash incentive increased response to a non-response follow-up (NRFU) survey relative to a standard envelope. In order to explore avenues to increase the rate of envelope opening, the California Health Interview Survey (CHIS) will conduct a randomized experiment testing the use of a visible cash incentive to increase response rates with an envelope that has a small window that will show the cash incentive but no other envelope contents. This experiment will provide data about effects on a fresh sample as opposed to a NRFU. For this experiment, half of the households invited to participate in the first two waves of data collection in 2021 will receive their initial invitations in an envelope with a window on the front, allowing respondents to see that there is a two-dollar bill contained therein. The other half of the sample will receive a standard envelope with no cash visible. All subsequent reminder mailings and outbound phone protocols will remain the same between the two groups. This paper will present the results of the experiment as well as the rate of postal returns to assess possible mishandling of mail. This will contribute to the body of knowledge evaluating the utility of visible cash incentives in mailed survey recruitment letters.

Two Bucks or Six? A Push-to-Web Survey Incentive Experiment

Susan Hewitt, *Health District of Northern Larimer County*

Suman J. Mathur, *Health District of Northern Larimer County*

F. Jay Breidt, *Colorado State University*

David A Brown, *Colorado State University*

Since 1995, the Health District of Northern Larimer County (Colorado) has surveyed the adult population of the county every three years about health status, needs, and behaviors. The ninth iteration of the triennial Larimer County Community Health Survey (LCCHS) was fielded in fall of 2019, using an address-based random sample (ABS) and a six-wave, push-to-web approach. Just as the survey landscape has changed over time, the LCCHS design has been modified to maintain validity and control costs. For example, sample recruitment transitioned from an initial RDD with mail follow-up to a full ABS design in 2010. An online completion option was added in 2013. One thing that remained constant is the inclusion of a \$2 bill in the first postal mailing as a pre-completion incentive. We have seen a general decline in the response rates over time, and in 2016 the response rate was 28%. Only 29% of surveys were completed online in 2013 and 2016, using our passive web-as-an-option approach. With the adoption of a push-to-web approach in 2019, we designed an experiment comparing two different survey incentives. Of the 12,000 randomly selected addresses, 8,000 were mailed the standard \$2 bill in the introductory letter and 4,000 were mailed a \$1 bill with the promise of \$5 more should they complete the survey online. This experiment examined the effectiveness of prepaid vs. promised incentives to increase the number of surveys completed by web and overall, and the cost per completed survey of each approach. We found that the \$1 pre/\$5 post incentive increased survey response rates and yielded a more representative sample, while reducing costs by pushing more respondents online (79%online, 21% mail). Presenters will share recommendations for other survey practitioners interested in the potential impact of pre/post survey incentives for mixed mail/web designs.

Two Decades of Data Collection Experimentation on the National Longitudinal Survey of Youth, 1997 Cohort

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

Quentin Brummet, *NORC at the University of Chicago*

Vicki Wilmer, *NORC at the University of Chicago*

The National Longitudinal Survey of Youth, 1997 Cohort (NLSY97) is one of a series of longitudinal surveys sponsored by the Bureau of Labor Statistics, U.S. Department of Labor, studying predictors, correlates and consequences of labor force participation. The NLSY97 is a nationally-representative survey representing individuals born 1980 to 1984 and living in the U.S. in 1997 that both provides comparisons to prior generations, and serves as an unparalleled data source in understanding life course experiences of the millennial generation. The NLSY97 has continued in the tradition of its predecessors in maintaining exceptional response rates, with almost 80 percent of living original sample members participating in the 2019 interview. The roughly 9,000 person sample includes oversamples of Black and Hispanic individuals, and has been fielded 19 times since 1997, currently on a biennial schedule as a primarily phone survey with limited in-person follow-up.

Over its two decade history, the NLSY97 has incorporated several programs of rigorous experimentation into data collection activities. These include investigations of: mode transitions, operational issues such as gaining consent and collecting locating information, responses to data collection interventions such as email or mail prompts during data collection, variation in interim communications strategies, incentives, and questionnaire content to increase respondent engagement and subsequent participation. Example insights include the feasibility of decreasing respondent incentives in a panel context without damaging cooperation rates, different models of rapport-building throughout the course of an interview (such as for consent or locating information), and the limited value of many verification or confirmation questions within the questionnaire. This paper brings together two decades of evidence to contribute specific experimental results to the literature on panel surveys. We also share insights on the challenges and opportunities specific to undertaking experimental research within a long-running longitudinal survey.

Maximizing Response Rates for Physician Surveys: Testing Timing for Sequential Prepaid Incentives

Erin Glancey, *SSRS*

Robyn Rapoport, *SSRS*

Declining response rates in physician surveys have prompted researchers to experiment with different methods of incentivizing response among these sought-after professionals. Numerous studies have demonstrated that monetary prepaid incentives help induce cooperation and have been associated with meaningfully higher response rates than promised incentives (e.g., Delnevo et al 2004). Previous research also (e.g., Rapoport et al 2017) indicated that the incentive increases response in the mode that the incentive accompanies (i.e., online via weblink or paper). This paper contributes to research on maximizing responses rates in physician surveys by presenting findings from an experiment testing the most effective use of sequential prepaid incentives. The experiment was included in a multi-mode survey fielded in 2020 with OBGYNs and Family Medicine (FM) doctors.

The experiment tested two conditions: Half of the sample was sent (1) a \$75 check in the pre-notification mailing with a web link and (2) a \$2 cash prepaid incentive with a paper survey in the second mailing. Half of the sample was sent (1) a \$2 cash incentive in the pre-notification mailing with a web link and (2) a \$75 check with a paper survey in the second mailing. The following findings will be discussed: Overall yield was improved among those physicians that received the \$2 cash pre-incentive with the pre-notification letter and the \$75 check incentive with the paper survey. While OBGYNs had an overall better yield compared to FMs, the boost in overall yield was consistent among OBGYNs and FMs. Consistent with past research, the increase in yield for the experimental condition (\$2 first and \$75 second) appears to be driven by disproportionately more respondents completing the paper survey when they received the \$75 check with the paper survey. For both groups, a similar number of respondents completed the survey online.

Effects of Front-Loaded and Escalating Incentives on Response Rates and Response Quality in Election Surveys

Deji Suolang, *Johns Hopkins University*

Surveys on voting, public opinion, and political participation need high-quality data to inform explanations of election outcomes. Incentives have long been used in surveys to encourage sampled units to participate in the surveys and provide optimal responses. However, survey researchers are struggling to determine effective incentive strategies. This study sought to compare the effects of two different types of monetary incentives on election survey data quality: front-loaded incentive versus escalating incentive, and consider how these effects vary across respondents with different characteristics to inform adaptive designs for future election surveys.

Data analyzed in this study is from the American National Election Studies (ANES) 2016 Recruitment Pre-test Study. ANES is a nationally representative survey designed to collect data on voting and public opinion in the United States, conducted before and after every presidential election. Its pre-test study is an experiment to prepare for the internet component of the ANES 2016 Time Series Study. The respondents were randomly assigned to different conditions with different types of incentive and invitation methods. In total, there were 696 respondents who completed the screener and were invited to the main survey online. The data quality is evaluated by two measures: non-response error and measurement error. Results indicate that the front-loaded incentive is more effective in increasing the overall response rates, but escalating incentive causes lower rates of survey satisficing behaviors such as speeding and straight-lining. The propensity model is employed to evaluate the effect of different types of incentives across different subgroups, which identifies certain subgroups that are associated with lower response propensity and the higher probability of survey satisficing. To encourage the high-quality participation of the sample units, future election surveys could use adaptive survey design to reduce the overall survey budget and achieve higher cooperation.

Can Prepaid PayPal Incentives Be As Effective As Prepaid Cash Incentives?

Antje Kirchner, *RTI International*

Emilia Peytechva, *RTI International*

Nicole Tate, *RTI International*

Cash prepaid incentives have been shown to significantly increase response rates by increasing the perceived legitimacy of the survey request and invoking norms of reciprocity. However, mailing prepaid cash incentives is not always an option. For example, increased concerns over mail theft or increased mobility of sample members may render prepaid cash incentives suboptimal. Alternative forms of prepaid incentives, such as prepaid PayPal incentives, however, are often not as successful as cash prepaid incentives. One reason that prepaid PayPal incentives may not be as successful as cash incentives may be that they are not as tangible as dollar bills. The 'prepaid PayPal' information may not stand out enough in the invitation letter leaving sample members unaware. The goal of this study is to investigate whether it is possible to increase the perceived saliency of prepaid PayPal incentives by manipulating the communication and design of the notification. Ultimately, we are interested in whether we can achieve similar response rates and sample composition as a comparable prepaid cash incentive.

We implemented an experiment with prepaid incentives as part of the second follow-up of sample members from a large, nationally representative survey of recent college graduates. Due to the longitudinal nature of the study we had access to email addresses from prior responses and tracing efforts which can be used to send prepaid PayPal incentives. We randomly assigned sample members (n=3,130) to either the cash condition, mailing a \$2 bill, or the PayPal condition, announcing the prepaid PayPal incentive on a separate index card attached to the letter similar in size to a dollar bill. Preliminary results suggest that overall response rates do not differ significantly, but that there is a significant interaction effect of age with the PayPal condition resulting in a significantly younger respondent sample in the PayPal condition.

Burden, Breakoff, and Nonresponse Bias

Is Shorter Better? Survey Length and Response Rates in a US Survey of Nonprofits

Scott Leary, *Internal Revenue Service*
Rizwan Javaid, *Internal Revenue Service*
Patrick Langetieg, *Internal Revenue Service*
Brenda Schafer, *Internal Revenue Service*
Martha Stapleton, *Westat*
Sarah Bennet-Harper, *Westat*
Kerry Levin, *Westat*

Numerous factors, including the sample population, topic, layout of the questionnaire, complexity of the items, and length have been found to play a role in respondents' perceived burden of a survey, which in turn impacts survey response rates (Dillman, Smyth, & Christian, 2014). While some research has shown that longer paper surveys are associated with lower response (Yammarino, Skinner, & Childers, 1991), other studies, particularly those examining web response, have been less conclusive (Cook, Heath, & Thompson, 2000; Sheehan 2001). The IRS Tax-Exempt Organization Burden (TEB) survey is a U.S. government mixed-mode survey that measures the time and money nonprofits and charities spend complying with U.S. federal tax law regulations. The TEB survey is sent to 15,000 organizations in the U.S. There are five contact attempts, with contacts 1, 3, and 5 including a paper survey. Respondents can complete the survey by paper or the web.

We conducted an experiment to compare the standard-length questionnaire to a shortened questionnaire. Eighty percent of respondents received the standard 12-page, 14-item survey and 20 percent received a shorter eight-page, 10-item survey. While both versions include key questions about time and money spent to comply with tax regulations, the shorter version includes fewer demographic, contextual, and follow-up questions. We will examine the impact of survey length on overall response rates, including examining any differences by mode. Furthermore, using IRS administrative data, we will look at differences in response by characteristics of the organizations (e.g., size, type of nonprofit, filing status) to determine if a shorter survey brings in organizations that would otherwise not have completed the standard-length survey.

Cooperation Rates, Respondent Burden, and Topic Salience in Bladder Health Research: Findings From the Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium

Cynthia Fok, *University of Minnesota*
Charles Cain, *University of Minnesota*
Emily S. Lukacz, *UC San Diego Health - LaJolla*
Sara Putnam, *University of Minnesota*
Eileen A. Robinson, *University of Michigan*
Todd Rockwood, *University of Minnesota Division of Health Policy and Management*
Kyle Rudser, *Biostatistical Design and Analysis Center, Clinical and Translational Science Institute*
Melissa Constantine, *Division of Biostatistics, Coordinating Centers for Biometric Research (CCBR) University of Minnesota*

Introduction: Leverage salience theory posits that response rates are determined by the perceived benefit of participation (leverage) as well as perceived topical relevance (salience). In women's bladder research survey data collection has occurred primarily in a symptomatic population (those with bladder symptoms) so there is a paucity of survey response rate estimates from non-symptomatic women. Based on leverage salience theory we, as part of the Prevention of Lower Urinary Tract Symptoms Research Consortium, hypothesize there will be differential response and cooperation rates across women with differing perceptions of bladder function.

Methods: Women from communities near consortium clinical centers were screened based on self-report of bladder related symptoms to assure symptom variance. Participants were randomized to receive either a paper or online survey containing 209-348 bladder related items (BHI), depending on branching. Participants who completed the survey were asked to complete a 3-day bladder diary measuring daily

urine output, and at a clinical visit complete clinical tests to evaluate individual bladder function. Response and cooperation rates across the different levels of participant burden were compared across self-reported degree of bladder function.

Results: There was a difference in self-reported bladder function between those who did (n=539) and did not (n=83) complete the BHI ($p < 0.05$). Women who responded to the BHI reported better bladder function and less bother from lower urinary tract symptom (LUTS) of pain/discomfort ($p < 0.05$). Women with less severe bladder symptoms were also more likely to have completed the clinical evaluation (n=337) than those who did not (n=58) ($p < 0.05$).

Conclusion: Contrary to our hypothesis, women in our study with less severe bladder symptoms were more likely to participate in research. This suggests that increased symptom severity and presence of LUTS are not the salient or leveraging factors that impact a woman's decision to participate in research with increasing levels of burden.

Time to Break and Breakoff for Computer-Assisted Survey of Bladder Health

Kyle Rudser, *University of Minnesota*

Charles Cain, *University of Minnesota*

Sara Putnam, *University of Minnesota*

Emily Lukacz, *University of California, San Diego*

Todd Rockwood, *University of Minnesota*

Melissa Constantine, *University of Minnesota*

For the Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium, *University of Minnesota*

Survey length is a well-recognized key attribute to participant burden leading to completion breaks and increase likelihood of breakoff and nonresponse. We report on the experience from participants assigned to a computer-assisted survey used in a validation study of a bladder health instrument under development by the PLUS Research Consortium.

Participants stem from two groups: an address-based sample across the United States and a convenience sample from communities and clinics near clinical research centers. Pre-notification letters, including a noncommittal \$2 bill, were mailed to 6000 households identified via stratified sampling of the delivery sequence file asking for participation by an adult female in the household. Participants were randomized 1:2 to receive a paper or online survey with \$10 incentive contingent on completion. 284 participants were recruited near research centers and randomly assigned to the same electronic bladder health survey. The survey length was 209 to 348 items, depending on branching. Using timestamps on data fields throughout completion, we used survival analyses to examine the magnitude and timing of breaks and breakoff.

A total of 428 participants initiated the survey. Breaks were identified as gaps between data entries of 20min or more. The average time to survey completion was 51min [IQR: 34, 64], among those who completed, excluding breaks. Only 57% completed the survey without a break and 16% took 2 or more breaks. Within 30min, 32% of participants took at least one break, rising to 42% by 60min. The breakoff cumulative incidence leveled off at 20min before more than doubling to 7% in total. In terms of number of items, by the 75th and 150th questions, 22% and 36% took at least one break, respectively. This data quantifies the relationship between survey length and timing of both breaks and breakoff over a computer-assisted survey of over 200 items.

Impact of Question Topics and Filter Question Formats on Web-Survey Breakoffs

Zeming Chen, *University of Manchester*

Alexandru Cernat, *University of Manchester*

Natalie Shlomo, *University of Manchester*

Stephanie Eckman, *RTI International*

Web surveys have become increasingly popular over the last decade, but they tend to suffer from more breakoffs, which take place when respondents start the survey but do not complete it. Many studies have investigated the factors impacting breakoffs, but they often ignored the breakoff timing and gave scant attention to two factors: question topics and filter question formats (grouped vs. interleaved as defined by whether filter questions are presented upfront or not). This study aims to research the effect of these two factors on the breakoff event and its timing with the help of survival analysis. By using a web survey that experimentally manipulates the filter question format and randomly orders the question topic, this study finds that presenting the filter questions in the grouped format leads to lower breakoffs at the beginning in comparison to the interleaved counterpart, but the breakoff risk in the grouped format catches up quickly when respondents realise their previous answers will trigger more questions. This study also identifies that the insurance topic has more breakoffs while demographic and income topics have fewer breakoffs despite their perceived sensitivity level. The present study demonstrates the importance of taking into account the time dimension in the study of breakoff and gives practical guidance on questionnaire design and survey breakoff mitigation.

The Evolution of Non-response Bias: A Look at Yesterday, Today and Tomorrow

Erin Caldwell, *Polco*

Michelle Kobayashi, *Polco*

Identifying and correcting non-response is becoming increasingly more difficult as the response rates of probability surveys continue to decline and the probability modes are commonly being replaced by non-probability methods. To understand the trends in non-response bias, the National Research Center at Polco examined its national database of more than 100 matched sets of probability and non-probability community surveys.

The datasets analyzed are based on the National Community Survey TM, a templated, benchmark that survey local governments use across the nation to measure quality of life, community livability, local government performance and civic participation. Probability-based surveys have been aggregated in this database since 1997 and non-probability surveys (a second survey completed alongside the probability surveys) since 2013. The database includes the data for more than 850 communities (probability samples) and 100+ matched pairs of probability and non-probability samples.

Resident demographics along with measures of civic participation and public trust were examined to identify trends over time and better understand the key discrepancies between the matched datasets. Community level data also will be examined to examine if non-response patterns are similar by geography, population size, urban/rural status and residential socio-economic compositions.

Investigating Nonresponse and Nonresponse Bias in a Nonresponse Follow-up (NRFU) Study

Paul J. Lavrakas, *Independent Consultant*

Andrew Zuckerberg, *National Center for Educational Statistics*

Mahi Megra, *Ameriicna Instiuttues for Research*

In the past two decades, it has become a best practice to incorporate nonresponse bias studies into survey designs. Such studies have started to be reported (cf. Montaquila and Olsen, 2012), including those so-called “NRFU studies” (nonresponse-followup) that attempt to gather new data from a sample of the original study’s nonresponders. However, few of these NRFU studies achieve a high response rate and to our knowledge almost none have reported about the nonresponse that resulted (an exception being Lavrakas et al., 2012).

Most NRFU studies have been quantitative in nature and have tried to gather the same type of quantitative data from nonrespondents as gathered in the original survey questionnaire. In 2019, however, the National Center for Educational Statistics planned a qualitative nonresponse investigation on a sample of nonrespondents to the 2019 National Education Household Survey (NHES). That NRFU study sampled 760 nonresponding addresses from the 2019 NHES that were distributed across seven heterogenous metropolitan areas of the U.S. In-person interviewers approached each sampled address to determine whether it was occupied, and to gather structured observational data about each address and its neighborhood. In four of the locations (n = 400 addresses), interviewers attempted to complete a 90-minute in-person qualitative interview with the address’s householder(s). Observational data were gathered from 100 percent of these 400 addresses, and an interview was completed at 21% of the eligible addresses.

We will report on the characteristics of the addresses that did, and those that did not, participate in the NRFU qualitative interview, and on the possible nonresponse bias in the NRFU study that resulted from the failure to obtain interviews from all the NRFU addresses. To do this, we will use the observational data that were gathered for each sampled NRFU address and auxiliary data that were appended to each NRFU address.

Precision of Estimates of Nonresponse Bias in Means

Stephanie Eckman, *RTI International*

Jennifer Unangst, *RTI International*

Jill Dever, *RTI International*

Chris Antoun, *University of Maryland*

Survey researchers increasingly provide estimates of nonresponse bias in several variables when they release or analyze data. Researchers understand that sample estimates of population values should be reported with appropriate measures of uncertainty, such as standard errors or confidence intervals. However, few studies acknowledge that nonresponse bias estimates are also subject to sampling variability. Using simulations, we study the sampling variability of nonresponse bias estimates in means and how that variability is affected by features such as clustering and response rates. Results show that low response rates and high clustering make nonresponse bias estimates more variable. We then evaluate three methods to estimate the sampling variance of estimates of nonresponse bias in means: a method developed by Lee (2006), jackknife replication, and linearization. We find that the Lee approach works well for simple random samples but overestimates variability for clustered samples. Linearization and replication work well with all populations, and we give an algorithm for the implementation of these approaches. We also apply the Lee and replication methods to the LISS panel and the General Social Survey, confirming the simulation results.

Sampling and Recruitment Strategies for Hard-to-Reach Populations

Two Roads Diverged in a Wood, and I Oversampled the One Less Traveled By: Evaluating Stratification Methods for Targeting Hard-To-Reach Populations

Cameron McPhee, *SSRS*

James McKinstry, *SSRS*

Arina Goyle, *SSRS*

Arifah Hasanbasri, *SSRS*

Brian Wells, *UCLA Center for Health Policy Research*

As response rates decline, it becomes challenging for survey researchers to obtain representative samples of harder-to-reach subgroups (e.g., Hispanics, African-Americans, individuals living below the poverty line, uninsured individuals). Often sample frames are stratified in order to oversample these important groups. However, this type of stratification is not always straightforward since it often relies on the appending of auxiliary data to probability frames. These auxiliary data could come from overlaying area-level information (e.g., based on census data) on the frame or from merging in commercial data based on a common identifier, such as address. If the goal is the identification of homogenous strata prior to sampling, the Census block or tract level demographic data may match the sample unit and the commercial data sources are often inaccurate (Harter et. al., 2016). Consequently, the sampling efficiency obtained from the use of these variables varies by survey and by variable (Roth, et. al., 2018). However, Dutwin (2018) demonstrated that machine learning techniques, such as random forests, could predict rare population membership better than single sampling frame indicators of group membership.

This presentation explores the utility of creating sampling strata based on several different classification methods. Specifically, we will use sample and response data from the 2019 California Health Interview Survey to compare sampling strata designated based on the same values of appended auxiliary data (either commercial or census-based) with strata formed using predictive models. The analysis will compare each method on the accuracy with which the stratification differentiates between members of the target population and non-members as well as the ease and efficiency of the process. We will evaluate these stratification methodologies for multiple subpopulations including Asians, Hispanics, African Americans, Households with Children, and Households living below the poverty threshold.

Racial/Ethnic Oversamples From a Web Panel: Comparing to Geographic Oversampling in RDD and ABS

Randal ZuWallack, *ICF International*

Matt Jans, *ICF*

Thomas Brassell, *ICF*

Priscilla Martinez, *Alcohol Research Group*

Deidre Patterson, *Alcohol Research Group*

Thomas K. Greenfield, *Alcohol Research Group*

Katherine J. Karriker-Jaffe, *RTI*

A common sampling design to increase the number of respondents from specific racial/ethnic groups is to stratify and oversample geographic areas with high corresponding population density. However, geographic sampling is often inefficient since the oversampling rates often need to be large to achieve the target number of respondents, resulting in high design effects due to differential weighting. Further, not all members of the geographic area will be of the target race/ethnicity, which results in an increase in all respondents from those areas or requires costly screening of those who are not members of the target race/ethnicity group. Additionally, members of the high-density geographic areas only represent a portion of the target race/ethnicity group, often geographically clustered, and may be different from other members of the race/ethnicity group living elsewhere. Availability of pre-recruited nonprobability web panels, where race/ethnicity of panel members is known in advance, offers a more efficient and cost-effective alternative for increasing sample size of specific race/ethnicity groups.

To meet the goal of increasing sample sizes of Black and Hispanic respondents, using data from the 2019-2020 National Alcohol Survey, we compare results of random digit dialing and address-based

sampling using geographic oversamples with an oversample from a nonprobability web panel. First, we assess whether the Black and Hispanic respondents selected from the oversampled geographic strata are different from Black and Hispanic respondents selected from non-oversampled strata. Second, we compare Black and Hispanic respondents from the nonprobability web panel with Black and Hispanic respondents from the RDD and ABS, separately. Preliminary results indicate demographic differences between RDD and web panel for both Black and Hispanic respondents, as well as differences in demographics between the ABS and web panel for Black respondents, but not Hispanic respondents. Our presentation will focus on differences in demographics, alcohol consumption, tobacco use, discrimination, and religion.

Representation in Research: Diversity and Online Samples

Frances Barlas, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

Beatrice Abiero, *Ipsos Public Affairs*

In 2020, we have seen a broader awakening to the continued systemic racism throughout all aspects of our society and renewed calls for racial justice. For the survey and market research industries, this has raised questions about how well our industry does in ensuring that our public opinion research captures the full set of diverse voices that make up the United States. These questions were reinforced in the wake of the 2020 election with the scrutiny faced by the polling industry and the role that voters of color played in the election. We conducted a study to assess how well online samples represent communities of color and their diversity. While past studies have found lower bias in probability-based samples with online panels compared to opt-in samples (MacInnis et al., 2018; Yeager et al., 2011) there has been little investigation into representativeness among subgroups of interest. In Sept. 2020, we fielded parallel studies on Ipsos' probability-based KnowledgePanel which is designed to be representative of the US and an opt-in nonprobability online sample with approximately 3,000 completes from each sample source. The questionnaire included a number of measures that could be benchmarked against gold standard surveys such as the Current Population Survey, the American Community Survey, and the National Health Interview Survey. We found that across all race/ethnicity groups KnowledgePanel had lower bias than opt-in sample. In both sample sources bias was lowest among white respondents and higher among Black and Hispanic respondents. A calibration adjustment using variables related to the self-selection bias of opt-in samples helped reduce the bias in the opt-in sample across all race/ethnicity groups. We provide recommendations to improve representativeness with online samples.

Is RDS the Right Approach? Assessing if the Number of Potential Seeds is Sufficient for Your Study

Deirdre Middleton, *ICF*

Adam Lee, *ICF*

Laurie Drabble, *Alcohol Research Group*

Karen Trocki, *Alcohol Research Group*

Amy Mericle, *Alcohol Research Group*

Tonda Hughes, *Columbia School of Nursing*

Ronaldo Iachan, *ICF*

Brian Kim, *University of Maryland Joint Program in Survey Methodology*

Sample size calculators are regularly used in survey planning to identify requisite numbers of respondents. Researchers considering respondent driven sampling (RDS) studies should also assess if the initial sample size is sufficient, but the calculations needed are different from traditional surveys. RDS is a network sampling technique where respondents recruit additional study participants. In RDS studies, including virtual adaptations of RDS, there are often limited numbers of people from the target population known to researchers (potential seeds). There is currently no standard method for determining whether the number of available potential seeds is sufficient, though it is a known RDS problem to sometimes require many seeds (a large initial sample known to researchers). The critical question becomes not how much sample should be drawn from a frame, but whether the limited known sample is sufficient to achieve the required number of respondents.

This presentation discusses two metrics that can be used to estimate if the initial number of potential seeds is sufficient to reach the target number of respondents. First, the seed productivity probability estimates whether there will be enough people in the first wave to spur sufficient recruitment at the study start. Second, the reproduction number describes the average number of recruits generated per each participant, which indicates whether recruitment will expand or diminish with each new wave of recruitment. These estimates are important because, for rare populations, the number of potential seeds may be quite limited, and programming and other costs associated with setting up an RDS study can be high. As many researchers are making modifications to adapt to COVID, calculating these numbers in advance will help researchers assess if virtual RDS is a feasible approach for their study goals.

A Study of Respondent Driven Sampling Researchers

Sunghee Lee, *University of Michigan*

Ai Rene Ong, *University of Michigan*

RJ Batas, *University of Michigan*

Jiangzhou Fu, *University of Michigan*

While respondent driven sampling (RDS) has increasingly become popular for studies of hard-to-sample populations, recruitment noncooperation in RDS makes the sample sizes essentially a random variable for which researchers have little information, making them neither predictable nor controllable. Exacerbated by the fact that operational details are rarely reported in the peer-review literature and the lack of transparency and inadequate reporting of RDS studies, when faced with recruitment noncooperation, RDS researchers will be left on their own to make design changes on the spur of the moment in hopes of making RDS “work”.

This study attempts to fill this gap by providing a more complete picture of RDS operations. For doing so, we will conduct a Web survey with RDS researchers, followed by in-depth interviews. The researcher sample comes from a list of researchers identified as those who have used or planned to use RDS from our literature review and grant search. Because individual projects may be at different stages and some studies may have considered RDS in the proposal stage but decide not to pursue, questionnaires will be tailored toward operational aspects of RDS (e.g., design details, operational challenges and solutions). Further, based on the responses to the survey, we will subsample researchers stratified by whether encountered operational challenges in RDS data collection and conduct semi-structured in-depth interviews that will focus on target population, seed selection, data collection mode, instruments, project personnel, design and use of coupons and incentive structures and amount to assess critical features leading to successful recruitment cooperation in RDS.

The Challenge of Recruiting Teens: Examining Recruitment Methods for a Nationally Representative Probability-Based Teen Panel

Lindsay Liebert, *NORC at the University of Chicago*

Ipek Bilgen, *NORC at the University of Chicago*

Erlina Hendarwan, *NORC at the University of Chicago*

Roopam Singh, *NORC at the University of Chicago*

Vicki Pineau, *NORC at the University of Chicago*

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

While there is numerous research on optimizing methods to reduce nonresponse in recruitment to probably-based web (or web mostly) panels (e.g. Schernzeel and Toepoel , 2012 ; Rao et al., 2010), there is little on the subject of recruitment of other (“secondary”) household members and much less on the recruitment of teen household members (those aged 13-17). Using a panel to recruit teens has many advantages. Recruitment is cost efficient, taking less time and effort to draw in teens; additionally, trust already exists with the enlisted household member. This paper examines the impact of various methods of recruitment of teens in the household with the help of the primary household member (the parent/legal guardian or “requestor”) who is already a member of the panel, as well as teen yield rates and teen panel composition. NORC’s AmeriSpeak Panel serves as the data source for this study. During the 2020 teen panel recruitment, panelists reported to have teens in the household were identified from household roster information obtained during the initial recruitment of the primary household member. We utilized

multiple methods to engage identified requestors and to recruit teens, including email reminders, text outreach, phone dialing, and letter mailings that incorporate instructions on accessing surveys directly using a custom URL or accessing the panel portal log-in. We will present findings from analyses examining the impact of these different outreach methods on requestor and teen recruitment. Preliminary analysis shows that 2020 recruitment of teens increased 24% over the last teen recruitment effort in 2017. Additionally, we will examine teen panel demographic composition and determine which methods succeed at recruiting a nationally representative teen panel. The results from this study aim to illuminate successful methods for recruiting and empaneling secondary teen household members utilizing existing primary panelists.

Trusted Partners: A Model for Recruiting Hard-to-Reach Survey Populations

Melissa Diliberti, *Rand Corporation*

David Grant, *RAND Corporation*

Stephanie Walsh, *RAND Corporation*

Heather Schwartz, *RAND Corporation*

The American School District Panel (ASDP) is a new effort to examine trends in the policy and practice of school districts and charter management organizations (CMOs). Over the next three years, researchers will conduct surveys and qualitative studies with district personnel that address a variety of timely topics, including districts' responses to the COVID-19 pandemic.

After several delays due to the pandemic, we invited a nationally representative sample of nearly 1,700 school districts and CMOs to join the panel in early fall 2020. School closures due to the COVID-19 pandemic as well as historical difficulties conducting research with this population forced researchers to explore strategies to recruit sampled districts into the panel. Researchers supplemented more traditional recruitment strategies (e.g., phone calls, emails, mailings) with a “trusted partner” approach. Specifically, researchers matched most districts in our sample with a state or national organization that had an existing relationship with school district personnel to act as a representative of the research team in recruitment efforts. These “trusted partner” organizations were involved in recruitment efforts to varying degrees; some played an active role in recruiting member districts by sending emails on the ASDP team’s behalf whereas others offered their reputation as an endorsement of research efforts.

Preliminary evidence from our recruitment effort finds that a “trusted partner” model may be an effective recruitment tool. Specifically, we find that districts matched to a trusted partner were consistently more likely to enroll in the panel—and more likely to complete the first survey—than districts without a trusted partner. We also find that districts matched to state-level rather than national trusted partners more likely to be recruited into the panel. Finally, we find that recruitment efforts were more successful among those trusted partners that played a more active role in recruiting member districts.

Model-Based Approaches

Do Simulations of Electoral Votes Have To Be So Complicated? A Simple Data-Driven Approach Says No

Xinyue Zhang, *SurveyMonkey*

Jack Chen, *SurveyMonkey*

Computer-intensive statistical methods like Monte Carlo Simulation have been proven an effective methodology in measuring risk and uncertainty. Nowadays, the majority of Election simulation literature employs the method of calculating candidates' state win probabilities based on the pre-election polls and uses these probabilities as an input to Monte Carlo Simulation, assuming all states are mutually independent. However, there has been increasing evidence indicating the significance of between-states correlation. For instance, midwestern swing states such as MI, WI and PA are liable to behave similarly to one another. This study will expand the existing electoral vote simulation literature by 1) incorporating a joint distribution of all fifty states to account for state-by-state covariance, and 2) applying Weighted Moving Average(WMA) to assign more weight on the recent poll data to reflect the most updated trend. We use SurveyMonkey's polls (n = 1M) conducted continuously since June 8th to obtain weekly state-by-state vote spread and construct a multivariate normal distribution with WMA adjusted mean vector and WMA adjusted covariance matrix. With 100k simulations, the multivariate normal distribution is used to generate possible values of vote spread for each state. Our simulations indicate a 91% probability of Biden winning the election, with a range of 263 to 331 and an average of 290.

Advancing While Regressing: Toward Extending Multiple Regression For Partial Data

Meng Li, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

Robert Petrin, *Ipsos Public Affairs*

Multiple regression is one of the most common techniques used to study the simultaneous influence of multiple variables in political, organizational, and market research. Most multiple regression software excludes cases that have any pairwise missing data. As a result, researchers require that all cases have complete data on all items to be included in the analysis before proceeding with regression. For exploratory studies to examine a range of influences on key dependent variables, we often include as many questions as possible during data collection. This leads to surveys with high respondent burden, especially if we insist that all respondents complete all questions. The research industry is always under pressure to reduce respondent burden to improve data accuracy. One way to do this would be to randomly assign respondents subsets of questionnaire items, in what are often called fractional or planned missingness designs. An example would be randomly assigning 10 items out of a possible 20 items to any given respondent. This allows every pair of variables to have about an equal number of completes, while cutting response burden in half. At the same time, item missingness is random and not systematic. To evaluate the performance of fractional design in conducting multiple regression analysis, we took two studies where we had full datasets and, through repeated sampling, randomly selected subsets of respondents at different rates of missingness to generate correlation matrices via which we could calculate multiple regression coefficients and their average confidence intervals. We conclude that we can significantly reduce response burden by using fractional designs and still retain the ability to conduct multiple regression analyses. We also review some limitations on sample size and robustness to violations of data missing at random, and avenues for future research.

Separating Interview Effects From Non-Random Case Assignment Via Anchoring

Michael Elliott, *University of Michigan*

Brady T. West, *University of Michigan*

Methodological studies of the effects that human interviewers can have on the quality of survey data have long been limited by a critical assumption: that interviewers in a given survey are assigned completely random subsets of the larger overall sample that is being measured (also known as interpenetrated assignment). In the absence of this type of study design, estimates of interviewer effects on survey measures of interest may simply reflect differences between interviewers in the characteristics of their assigned sample members, rather than recruitment or measurement effects specifically introduced by the interviewers. We introduce a new approach for overcoming this lack of interpenetrated assignment when estimating interviewer effects. This approach, which we refer to as an “anchoring” method, leverages correlations between observed variables that are unlikely to be affected by interviewers (“anchors”) and variables that may be prone to interviewer effects (e.g., sensitive or complex factual questions) to statistically remove the components of within-interviewer correlations that a lack of interpenetrated assignment may introduce.

The improved estimates of interviewer effects on survey measures will enable survey managers to more effectively manage a data collection in real time and intervene when particular interviewers are producing survey outcomes that vary substantially from expectations. After describing this new approach, we will demonstrate its improved performance empirically using a simulation study. We will then illustrate its ability to adjust estimates of interviewer variance components in the absence of interpenetrated assignment using real survey data from the Behavioral Risk Factor Surveillance System (BRFSS), where interviewer IDs are systematically provided on public-use data files. We will present both Bayesian and frequentist approaches to applying the new methodology, and discuss the benefits and drawbacks of these alternatives.

It’s Not Always Location, Location, Location: Disentangling Design Effects From Interviewer Effects

Michael Robbins, *Princeton University*

Matthew Gottfried, *U.S. Department of State*

Design effects consider the sample design by inflating the variance of parameter estimates to account for clustering in survey. The calculation is driven by the intra-cluster correlation, which aims to account for the degree to which respondents interviewed in the same cluster share common experiences, beliefs or attitudes relative to those in other clusters. However, in most face-to-face surveys, respondents in a cluster not only share a geographic unit, but also the same interviewer or team of interviewers. Some research, such as Schnell and Kreuter (2005), has found that interviewer effects account for a larger share of the design effect than geographic variation, while Turner et al. (2014) find that interviewer effects account for a smaller share of the design effect. However, both of these studies have the limitation of relying on evidence from a single survey carried out in an OECD country where survey research is more established and oversight of interviewers tends to be stronger.

To gain greater leverage on this question, we rely on a 3MC survey of non-OECD countries that is carried out over regular waves that includes an interpenetrated design within each cluster. This approach allows for comparisons of the same questions across time and space, including in countries that are relatively homogenous and those that are relatively heterogenous, to evaluate the degree to which the design effect is driven primarily by geographic variation or is linked with interviewer effects. The results suggest that standard design effect calculations are often biased by virtue of capturing interviewer effects, which are inherent to both SRS and non-SRS techniques. As a result, when applied to 3MC surveys, the improper application of design effect calculations could affect the equivalence of estimations across countries.

Understanding Terrorism Policy Preferences Through Bayesian Model Averaging and Multiple Imputation

Brittany Alexander, *Texas A&M University*

Arnold Vedlitz, *Texas A&M University*

We present a Bayesian approach that combines multiple imputation and Bayesian model averaging to model Terrorism policy preferences among American adults using a two-wave probability-based panel from May 2016 and November 2016. Between the two waves, multiple domestic and foreign terrorist attacks occurred, providing a natural experiment for how the occurrence of terrorist attacks affects policy preferences and the perceived likelihood of a terrorist attack. The primary response variables of interest were: the change in perceived likelihood of a terrorist attack across the two waves, the change in support for federal and local spending across the two waves, and support for new terrorism-related policies at both waves and the change between the waves. The dataset had many missing items and contained 28 questions with multiple items, and the goal was to form a parsimonious model. An ordinary least squares (OLS) model without variable selection or imputation had to ignore many observations due to missingness and it was unclear what variables were the most important and if the model was concise. In an attempt to create a better model, multiple imputation using the R package *mi* was combined with Bayesian model averaging using the R package *BMA*. These new models were able to explain better the role of social-economic-political variables and psychometric variables in an individual's terrorism policy preferences than an OLS model. We found that terrorism policy support was stable across the two waves and that policies such as increased screening at airports and stricter background checks for gun sales were supported by most of the individuals surveyed.

5/11/2021, 3:00 PM - 4:30 PM
Concurrent Session B: Live Sessions

Mini-Conference: How Has COVID-19 Shaped Equity, Views, and Values

How Has COVID-19 Shaped Equity, Views, and Values

Katherine Carman, *Rand Corporation*

These four papers bring together research about how the COVID-19 pandemic has influenced our society with a focus on the role of equity. The pandemic and resulting recession have disproportionately impacted people of color and vulnerable populations. In these papers, we examine the groups that have been affected, the public's views about who has been affected, values about how we can address inequity, and pathways forward. Each paper uses a different survey and different populations to better understand the dramatic and inequitable impacts that COVID-19 is having on our society. In "The Impact of Coronavirus on Households in Major U.S. Cities", the authors consider the extreme financial hardship brought on by COVID-19 and the resulting recession, which is having a more significant impact on the most vulnerable households. In "Tracking the Consequences of COVID-19 Pandemic: Adapting an Ongoing Survey to a Public Health Emergency", the authors document the racial and ethnic disparities in impacts. In "Experience of Pandemic Hardship and Attitudes toward Health Disparities", the authors consider the impacts of the pandemic, and also survey respondents' awareness of inequity. In "American's View of the Impact of COVID-19: Perspectives on Racial Impacts and Equity", the authors use longitudinal data to examine how views of inequity have changed over time comparing responses from the same respondents in 2018 and 2020. Together these papers tell a story of dramatically increasing inequity, coupled with slowly changing attitudes and views of inequity.

Tracking the Consequences of the COVID-19 Pandemic: Adapting an Ongoing Survey to a Public Health Emergency

Michael Karpman, *Urban Institute*

Stephen Zuckerman, *Urban Institute*

Genevieve M. Kenney, *Urban Institute*

Elaine Waxman, *Urban Institute*

Poonam Gupta, *Urban Institute*

Dulce Gonzalez, *Urban Institute*

The economic impacts of the COVID-19 pandemic emerged quickly and many families, especially families of color, continue to struggle six months into the crisis. To document the consequences of the pandemic on individual and family well-being, the Urban Institute moved quickly to change the content of the March 2020 round of its Health Reform Monitoring Survey (HRMS), a nationally representative, internet-based survey of adults ages 18-64, to include questions around the impact of the pandemic on domains such as employment and material hardship. Smaller tracking surveys followed in May and September to assess changes in well-being during the COVID-19 pandemic and to assess racial/ethnic disparities. These tracking surveys sampled from respondents to the March 2020 HRMS and included an oversample of Black and Hispanic/Latinx participants. About one in four Hispanic/Latinx adults reported that they or their spouse or partner lost or were laid off from a job since the beginning of March, shares higher than those reported by white and Black adults. Black adults were nearly 4 percentage points more likely than white adults to be in families that experienced a job loss (15.8 percent versus 12.0 percent). Over one-third of Hispanic/Latinx adults and nearly 3 in 10 Black adults reported that they were worse off financially in September than they were in March. Though household food insecurity decreased between March/April and May 2020, the share of adults reporting food insecurity in the past month edged back up from May to September, after major federal relief programs expired. Among those who said they or a spouse/partner lost a job, nearly four in ten (37.1 percent) reported food insecurity in September, up from 33.6 percent in May and roughly the same level as in March/April (37.3 percent). Absent additional steps to protect and strengthen safety-net programs, families could face worsening hardship.

The Impact of Coronavirus on Households in Major U.S. Cities: Results From New York City, Los Angeles, Chicago, and Houston

Mary Findling, *Harvard School of Public Health*

The COVID-19 pandemic has created widespread disruptions to Americans' lives, including severe impacts concentrated among a group of households in crisis. Harvard T.H. Chan School of Public Health, NPR, and the Robert Wood Johnson Foundation conducted The Impact of Coronavirus on Households in Major U.S. Cities poll, July 1 – August 3, 2020, to examine the most serious health and financial problems facing households in the four largest U.S. cities and nationally prior to the expiration of federal coronavirus support programs. We conducted a representative, probability-based, address-based sampling poll among 3,454 U.S. adults ages 18+, including 512 adults in New York City, 507 adults in Los Angeles, 529 adults in Chicago, and 447 adults in Houston. We found at least half of households New York City (53%), Los Angeles (56%), Chicago (50%), and Houston (63%) reported facing serious financial problems during the coronavirus outbreak, compared to 46% nationally. Serious financial problems were reported by majorities of Latino and Black households in these cities, as well as majorities of households whose incomes fall below \$100,000, and majorities of households with employment or wage losses during this time. In health care, 19% of New York households, 20% of Los Angeles households, 23% of Chicago households, and 27% of Houston households reported household members unable to get medical care for serious problems when they needed it (20% nationally), with a majority facing negative health consequences as a result. Widespread serious financial problems were reported despite trillions of dollars distributed in federal, state, and charitable aid during the COVID-19 pandemic, signaling a future need for additional help to vulnerable groups. Serious problems in major cities are likely to continue and worsen over time, as more than half of households with employment loss reported using up household savings by August.

Experience of Pandemic Hardship and Attitudes Toward Health Disparities

Abigail Williamson, *Trinity College*

Sarah Willen, *University of Connecticut*

Colleen Walsh, *Cleveland State University*

COVID-19 has had profound consequences for many Americans, but these effects are unevenly distributed. In this paper we ask whether personal experience of COVID-19-related hardships is associated with greater support for policies that aim to mitigate health disparities. We draw on data from the Americans' Conceptions of Health Equity Study (ARCHES), an interdisciplinary investigation of how Americans form and change their views about who deserves what in the health realm. Phase I, from 2018-2019, included 170 interviews in Greater Cleveland, Ohio. In Phase II, findings from these interviews were tested in a nationally representative survey.

Phase I suggested that personal experience of hardship promotes receptivity to structural explanations for health disparities and a sense of solidarity with those who experience similar circumstances, resulting in greater support for policies that advance health equity. This paper tests these findings by examining how COVID-19-related hardships affect understandings of health disparities and related policy preferences. Drawing on a national survey of 3,300 Americans fielded in October 2020, it examines whether household exposure to COVID-19 hardships (defined through (1) a 9-item pandemic impacts inventory and (2) county-level exposure to cases and deaths) is tied to greater recognition of the social determinants of health, a stronger sense of health solidarity, and more support for policies that mitigate health disparities. We find that self-reported pandemic hardships are strongly and consistently associated with all of these outcomes, even when controlling for demographic and ideological characteristics. Geographic exposure to COVID is not associated with these outcomes.

Previous studies examining how COVID experiences affect pandemic-related policy views have found limited effects because partisanship dominates views on the pandemic. Our findings suggest that even when holding constant party and ideology, personal experience of COVID-related hardship is shaping broader health policy attitudes in ways that warrant attention in advocacy and messaging.

American's View of the Impact of COVID-19: Perspectives on Racial Impacts and Equity

Katherine Carman, *Rand Corporation*

Anita Chandra, *RAND Corporation*

Carolyn Miller, *Robert Wood Johnson Foundation*

Christopher Nelson, *RAND Corporation*

Jhacova Williams, *RAND Corporation*

Context: The COVID-19 pandemic has had a disparate effect on African Americans and Latino groups. But it is unknown how aware the public is of these differences, and how the pandemic has changed perceptions of equity and access to health care.

Methods: We use panel data from nationally representative surveys fielded to the same respondents in 2018 and 2020 to assess views and changes in views over time. At this time, one wave of data in 2020 has been analyzed. Another wave just closed, and a third wave will be collected in early 2021. We will combine data across waves for this presentation.

Findings: Based on data from 2018 and the first wave in 2020, we found that awareness of inequity is highest among Non-Hispanic Black Respondents and higher income and higher educated groups, and that there have been only small changes in perceptions of inequity over time. However, there have been significant changes in views of the government's obligation ensure access to health care.

Conclusions: Even in the face of a deadly pandemic, that has killed disproportionately more African Americans and Latinos, many in our society continue to not recognize that there are inequities in access to health care and the impact of COVID-19 on certain groups. But policies to address inequity may be shifting. We will continue to follow these respondents to see changes in attitudes endure over time or dissipate.

Mini-Conference: Response Patterns Before and During a Pandemic

The Impact of the COVID-19 Pandemic on Response to the California Health Interview Survey

Arina Goyle, *SSRS*

Margie Engle-Bauer, *SSRS*

YuChing Yang, *UCLA*

The fielding of the second year of the 2019-2020 California Health Interview Survey (CHIS) began in early March 2020, coinciding with state's stay-at-home orders in response to the COVID-19 pandemic. While the most immediate and noticeable impact of the pandemic on CHIS 2020 response was an improved sample yield and higher response rate, we hypothesized that the response rate did not increase evenly across all demographics. Though the impact of the pandemic on people's lives was immediate and dramatic, it did not impact people the same way. For instance, while many people faced a change to their work status, either through unemployment, furloughs, working from home, or working part-time, others such as essential workers were required to report to work as usual. Without long commutes to work some people found more time on their hands, while others struggled to juggle work and childcare. Based on such observations, we predicted that the uptick in response to the CHIS survey would be attributable to certain key demographics. The current paper seeks to compare response from CHIS 2019 with CHIS 2020 and explore changes in the demographic composition of those who responded during the pandemic. Preliminary comparisons of the demographic composition of CHIS 2019 and 2020 respondents indicate differences in levels of income, education, and employment. Further analyses will explore differences based on sex, age, citizenship, race, family composition and geography. This paper will also discuss what the differential response implies for data collected in CHIS 2020.

The Impact of the COVID-19 Pandemic on RDD Cell Phone Response in the National Immunization Surveys

Jason Boim, *NORC at the University of Chicago*

Benjamin Skalland, *NORC at the University of Chicago*

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

Holly A. Hill, *Centers for Disease Control and Prevention*

Michael Chen, *Centers for Disease Control and Prevention*

Natalie Sterrett, *Centers for Disease Control and Prevention*

David Yankey, *Centers for Disease Control and Prevention*

The COVID-19 pandemic upended the lives of millions of Americans in 2020. The public health crisis, stay-at-home orders, and economic upheaval have had far-reaching impacts on the attitudes and behaviors of the public. One surprising change during 2020 is the increased willingness of respondents to participate in telephone surveys. In this presentation, we demonstrate this effect by tracing respondent participation rates over time in the National Immunization Surveys (NIS). The NIS are random-digit-dial (RDD) cell-phone surveys sponsored by the U.S. Centers for Disease Control and Prevention that are used to assess vaccination coverage in the United States among children age 19-35 months (NIS-Child) and adolescents 13-17 years (NIS-Teen), and influenza vaccination coverage among children 6 months-17 years (NIS-Flu). To examine the effect of the pandemic on survey participation, we present a weekly time series of the NIS-Child contact and cooperation rates from Spring 2019 to Spring 2020. In particular, we observe a substantial increase in respondent cooperation rate in March and April 2020 compared to March and April 2019, followed by a leveling off of the cooperation rate in the months following April 2020. We also examine whether the impact on NIS-Child cooperation rate differed among Census divisions of the country and among demographic subpopulations and whether the increase in respondent participation resulted in a more representative sample before weighting adjustments.

Impact of the Pandemic on AmeriSpeak Cooperation Rates

Min Zhu, *NORC at the University of Chicago*

Kanru Xia, *NORC at The University of Chicago*

Vicki J. Pineau, *NORC at the University of Chicago*

Facing the challenge of COVID-19 quarantine, AmeriSpeak®, a large probability-based panel designed to be representative of the U.S. household population, implemented changes to data collection procedures and experienced changes in response rates in both the panel recruitment survey and in client surveys conducted with panelists. AmeriSpeak panel recruitment is historically accomplished using a two-stage process: (i) an initial recruitment using USPS mailings and telephone contact, and (ii) a more elaborate non-response follow-up recruitment using FedEx mailings and in-person visits by NORC field interviewers. Once a panelist is recruited, s/he will participate in client surveys either by web or by phone. This study will describe the delays in mail and phone outreach and cancellation of the in-person mode of data collection during recruitment and their impacts on the recruitment survey response rate by comparing to the previous years' recruitment survey response rate. We will analyze differences in recruitment survey response rates overall and by subgroup, especially hard-to-reach populations. This study will also assess the effect of COVID-19 on subsequent survey response rates after recruitment. Overall, we are finding a 6% increase in client survey response rates during the 2020 COVID-19 lockdown, despite setbacks in outreach efforts using the phone mode for client surveys. We will also examine the COVID-19 influence by survey topic, language, panelist demographics, and geography.

Personal Visits to Telephone: Effects of a Mode Shift on Coverage and Nonresponse Due to COVID-19 for a National Panel Survey

Christine Bottini, *U.S. Census Bureau*

Beth Newman, *U.S. Census Bureau*

In March 2020, the U.S. Census Bureau temporarily shifted Field Representatives (FRs) from in-person interviews to telephone-only interviews in an effort to slow the spread of the Coronavirus (COVID-19). This was a dramatic change for many surveys, especially since most sample from addresses frames that do not include phone numbers and many were already in the middle of data collection in March. An unscheduled mode change during data collection has the potential to introduce survey error. This research examines the potential for coverage and nonresponse error after changing the data collection mode of a national panel survey.

The 2020 Survey of Income and Program Participation (SIPP) collected information from both new and returning panel members. To conduct telephone interviews for new panel members, FRs had to rely on matched administrative records to obtain a telephone number for the household. Our research aims to identify if households without a phone number, or those with a phone number that is unreachable, are different from households with an available/reachable phone number. This research also seeks to identify potential nonresponse bias. Using previous cycle data for retuning panel members and Census Planning Database block group level estimates for new panel members, we determine whether the changes to the data collection protocol affected the respondent and nonrespondent compositions compared to other data collection years. We compare demographic information for respondents and nonrespondents before and after the March 2020 mode change, and across multiple years of data collection.

Examining Response Rates Across Time in a Multinational Telephone Survey – Before and During the Coronavirus Pandemic

Martha McRoy, *Pew Research Center*

Over the past two decades, the Pew Research Center has been conducting the annual Global Attitudes Survey using interviewer-administered modes for data collection. The global pandemic had a large influence on where surveying could be done for our 2020 surveys as face-to-face interviewing was not possible during our fieldwork period and therefore only telephone interviewing could be conducted. The summer 2020 survey was conducted solely using computer-assisted telephone interviewing in 14 countries, 12 of which were surveyed annually for the project and where pre-pandemic data is available. These 12 countries comprise three countries in the Asia-Pacific region, seven countries in Europe and two countries in North America. While the impact of the coronavirus pandemic is clear on face-to-face interviewing for the project, it is unclear how this affected our telephone research.

In this presentation, we examine the potential impact of the pandemic on our 12 telephone countries across three regions of the world by examining changes to contact, cooperation, and response rates; average attempts-per-complete; and average attempts-per-contact in our Global Attitudes Survey from 2018 through the end of 2020. As most of the surveys were conducted using dual-frame designs, these rates are compared separately by landline and mobile samples as well as the combined rates. For countries where substantial differences were found regarding response rates across time, we also explore the changes to the demographic and household compositions of our respondents in an attempt to target how the pandemic may have influenced our study in terms of survey response.

Mini-Conference: Questions, Questionnaires, and Measurement: The Effect of COVID-19

Cognitive Questionnaire Testing in COVID Times... and Beyond

Benoit Allard, *Statistics Canada*

Statistics Canada's Questionnaire Design Resource Centre (QDRC) routinely conducts cognitive testing to evaluate questionnaires. This usually takes the form of one-on-one, cognitive interviews, or focus groups. These activities were put to a halt when the Covid-19 pandemic hit Canada in March 2020. In reaction, the QDRC quickly investigated ways to shift its activities towards virtual cognitive interviews and focus groups. By summer 2020, all questionnaire testing was done through videoconferencing. This has become the "new normal" for QDRC consultants, who have been teleworking full-time since then.

It is now time to look ahead to the post-pandemic situation. With reductions in government spending on the horizon, the QDRC will probably not be able to afford travelling to meet test participants in person. Also, some questionnaire testing activities have turned out to be more efficient online than in person. In short, there is no going back to "how things used to be". On the other hand, after testing some questionnaire virtually, it was evident that an in-person test would have been more efficient. It is therefore likely that questionnaire testing activities will consist of a mix of online and in-person cognitive testing.

Responding to Emerging Public Health Threats: Examples From the National Health Interview Survey

Benjamin Zablotzky, *National Center for Health Statistics*

James Dahlhamer, *National Center for Health Statistics*

Sarah Lessem, *National Center for Health Statistics*

Aaron Maitland, *National Center for Health Statistics*

Grace Medley, *National Center for Health Statistics*

Catherine Simile, *National Center for Health Statistics*

Beth Taylor, *National Center for Health Statistics*

The National Health Interview Survey (NHIS) was recently redesigned with two major goals in mind. First, to reduce respondent burden while enhancing data quality, and second, to meet the data needs and the current missions and priorities of the Department of Health and Human Services (DHHS), the Centers for Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS). The emergence of a global pandemic, like coronavirus or COVID-19, presented new challenges in continuing to fulfill these NHIS redesign goals. It required both adapting to a rapidly changing survey landscape, and quickly adopting relevant survey questions to fully capture the scope and impact of a public health threat.

This presentation will consist of two main topic areas. First, how researchers at NCHS identify and design effective questions to capture emerging public health topics areas. Second, using the coronavirus pandemic as a case study, how researchers at NCHS developed questions to capture the potential impact of this public health threat, including health care utilization and unmet health care needs, as well as mental health and social supports. The presentation concludes with lessons learned and a focus on how NCHS intends to continue to evaluate the health impact of the coronavirus pandemic moving forward.

A Scale for Measuring Social Distancing Behavior: Survey Questions and National Norms, USA 2020

Mariah Evans, *University of Nevada, Reno*
Jonathan Kelley, *International Survey Center*
Sarah Kelley, *Child Trends*

"Social distancing", a set of "nonpharmaceutical interventions" or NPIs in the medical literature, is a primary defence, perhaps the primary defence, against infectious disease, universally advocated by medical authorities in the US and throughout the world during the current coronavirus pandemic. The idea is not new. Perhaps the first government-directed quarantine system was mid-fourteenth-century Venice's "quaranta giorni", forty days of mandated isolation for incoming ships. We propose a 5-item primary scale of "social distancing" behavior (KEK-3) and a slightly extended variant (KEK-3m), developed for use during the COVID-19 epidemic (and, potentially, beyond). The candidate items all had 7 answer categories. Assessment aligns very well with the classical measurement model for multi-item scales: interitem correlations are high; alpha reliability is 0.86; correlations with criterion variables are similar across the candidate items; factor analysis (oblimin rotation) finds a single dimension with an eigenvalue over 1 and loadings in the .5 to .6 range. We provide behavioral norms for America during the 2020 pandemic and describe KEK-3's links to demographic and socioeconomic factors. Developing a replicable scale is especially important now, because many researchers are making erroneous comparisons using the same terminology to describe aspects of the epidemic which have been measured differently. To successfully assess the "...meaning of social change related to COVID-19, the newly emerging social practices due to lockdown measures..." (Esposito, Stark and Squazzoni 2020), high-quality measurements sufficiently reliable and robust to be replicated in different times as the epidemic evolves and in different settings are desperately needed: KEK-3 contributes to such a set of measures. Data: four large national sample surveys conducted April - July, 2020 during the coronavirus pandemic. Data collection was through Amazon.com's Mechanical Turk. A large majority of respondents took it on a computer with a normal-sized screen rather than on a cell phone

Assessing the Impact of Wording Changes in the General Social Survey 2020

Rene Bautista, *NORC at the University of Chicago*
Benjamin Schapiro, *NORC at the University of Chicago*
Jaesok Son, *NORC at the University of Chicago*
Michael Davern, *NORC at the University of Chicago*
Jodie Smylie, *NORC at the University of Chicago*

The General Social Survey (GSS) is a nationally representative survey of adults in the United States conducted since 1972. For nearly 50 years the GSS has systematically studied major social topics. This presentation discusses three experiments conducted in 2020, two motivated by the need to pivot from face to face to an ABS web-based survey due to the global COVID19 pandemic and one by societal changes. Experiment 1: Comparison of GSS core questions whose response options allow recording of volunteered responses. In the traditional mode of administration, interviewers can accept volunteered responses; however in the web version, such flexibility might not be entirely possible. For example, for the question "In general, do you think the courts in this area deal too harshly or not harshly enough with criminals?" the control version includes two options (Too harshly / Not harshly enough) while the experimental version used three options (Too harshly / Not harshly enough / About right). Experiment 2: Adoption of grids for opinion questions. For example, for the question series "Do you think it should be possible for a pregnant woman to obtain a legal abortion?" The control version includes a series of independently presented Yes-No statements while the experimental version assigned the statements to a grid. Experiment 3: Changes to include gender neutral wording for references (which was motivated by evolution of language in society). For example, "should he be allowed to speak?" versus "should this person be allowed to speak?" A total of 19 variables are considered for the volunteered responses, eleven variables for grids, and nineteen variables for gender neutral pronouns. The analysis explores which groups of the general population are more likely to be impacted by changes of key GSS variables in 2020.

Closing Information Gaps in a Pandemic: The National COVID-19 School Response Dashboard

Elizabeth Dean, *Qualtrics*

Gregory Laughlin, *Qualtrics*

Carol Haney, *Qualtrics*

Emily Oster, *Brown University*

Noelle Ellerson Ng, *AASA, The School Superintendents Association*

As the U.S. faced the back-to-school season in late Summer 2020, several critical information gaps around COVID-19 became clear to parents, teachers, school administrators, and public health officials. First, there was no single comprehensive public-facing data source on COVID-19 case rates in schools. Second, reporting of official COVID-19 numbers varied according to state policy and was not universally transparent. Finally, where cases of COVID-19 in school settings were publicly reported, critical contextual information was often lacking. For example, 5 cases in a 300-student elementary school in a community with less than 5 cases per 100,000 residents might be evidence that the school contributed to the spread of COVID-19. On the other hand, 5 cases in a 2000-student high school in a community with 25+ cases per 100,000 residents might instead imply that the spread of COVID-19 is coming into the school from elsewhere in the community. To address gaps in our national understanding of COVID-19 and schools, a team of researchers from Qualtrics, Brown University and AASA, the School Superintendents Association constructed the National Covid-19 School Response Dashboard, a tool that crowdsourced COVID-19 case data, in-person enrollment counts, grade levels, reopening approach (in-person, remote, or hybrid), and mitigation strategies from public and private schools in every U.S. state. The Dashboard and its associated statistical analysis environment, StatsIQ are free and available to the public, with the stated intention of being transparent and immediately useful. This presentation will document the strategies used to crowdsource data collection from schools, districts, and states, and will evaluate the overall representativeness and utility of the Dashboard as a tool for monitoring and responding to the pandemic in the context of American schools.

Wednesday, May 12, 2021 | Day 2 of Conference

5/12/2021, 10:00 AM - 11:00 AM

Round Tables: Live Sessions

Asian American, Native Hawaiian, and Pacific Islander Survey Participation: Challenges and Strategies

Asian American, Native Hawaiian, and Pacific Islander Survey Participation: Challenges and Strategies

Lisa Lee, *NORC at the University of Chicago*

Justine Bulgar-Medina, *NORC at the University of Chicago*

Angela Houghton, *AARP*

Neil Ruiz, *Pew Research Center*

This round table presents results from a joint project between AARP and NORC examining the barriers and strategies for increasing representation of Asian American/Native Hawaiian and Pacific Islander (AA/NHPI) populations in both community and national surveys. This roundtable will bring together experts who focus on these inclusion efforts to discuss barriers and obstacles to inclusion, as well as suggest paths forward. This roundtable will also highlight the unique opportunities AAPORites have to come together and support each other in this important work of ensuring adequate and responsive representation of AA/NHPI communities in all of our work. The goals of the roundtable align well with the mission of the recently formed AAPOR AAPI Research & Affinity group. Researchers from many disciplines with an interest in inclusive surveys, demographics, and smaller/rarer populations will find this lively roundtable particularly interesting.

Proposed Discussion Questions:

- What defines the AA/NHPI communities that we're talking about? How well does this acronym capture these groups?
- How well are AA/NHPI populations represented in national and community surveys?
- What are the key reasons that AA/NHPI communities are not well represented in surveys?
- What are the most effective tools/approaches for recruiting groups within the AA/NHPI communities to participate in surveys?
- What are the opportunities for researchers across AAPOR to come together to increase meaningful representation of AA/NHPI communities in our research?

Roundtable: Survey Research Teaching

Roundtable: Survey Research Teaching

Chase H. Harrison, *Harvard University*
Jennifer Dineen, *University of Connecticut*
Jolene D. Smyth, *University of Nebraska-Lincoln*
Sonja Ziniel, *Children's Hospital Colorado*

Many AAPOR members teach survey research in different settings, both inside and outside of formal academic programs. This roundtable, sponsored by the newly formed Survey Research Teaching (SRT) affinity group, will bring together AAPOR members who teach survey methods in different settings and discuss ways that the SRT group might better support and encourage survey methods teaching. Panelists will include practitioners who teach survey methods across a diverse group of settings, including undergraduate students, applied masters students, Ph.D. students, and professionals.

This panel is also designed to provide for strong participation by all interested attendees, and we hope that you will join us in helping brainstorm the many ways we can collectively work together to advance and improve our teaching.

Topics will include:

- The different settings we teach survey research
- Differences between courses, workshops, and direct advising
- Challenges and opportunities among different types of learners
- Strategies for teaching survey methods: lectures, hands-on, self-directed activities, others
- Teaching to different audiences: undergraduates, terminal masters-level, PhD, and others
- Strategies for teaching to methodologists versus end users
- Course projects, practica, client-focused research, student involvement in funded faculty research
- Challenges and opportunities of working with external clients, projects, and organizations
- Overlap and interests with different substantive specialties and traditional departments
- Necessary and complementary skills for students looking for survey research careers
- Possibilities and opportunities for sharing teaching material
- Possibilities and future directions for sustained engagement involving survey methods teaching

5/12/2021, 10:30 AM - 12:30 PM
Concurrent Session C: Prerecorded Sessions

Mini-Conference: Health Care Access and Delivery During the Pandemic

Attitudes and Work Place Experiences of Health Care Providers at an Academic Pediatric Hospital in the Midst of the First Wave COVID-19 in 2020 – Could We Have Predicted How Health Care Providers Would Feel About and Deal With Pandemic Uncertainty?

Sonja Ziniel, *University of Colorado School of Medicine/Children's Hospital Colorado*

Stacey Whiteside, *Children's Hospital Colorado*

Sara Deakyne Davies, *Children's Hospital Colorado*

Jennifer Reese, *University of Colorado School of Medicine/Children's Hospital Colorado*

Concerns about increasing stress levels and burnout of healthcare providers have reached new heights in the last decade. While individual- and system-level causes are examined and effective interventions are tested, institutions are more regularly monitoring their staff's burnout and resilience levels. Children's Hospital Colorado administers a yearly faculty experience survey and in 2017 added the validated 2-item Maslach-Burnout Index (MBI) to determine burnout levels for our faculty. Based on these data we created a "resilience index", composed of the top 7 survey questions correlating highly with the MBI, and used this index to monitor resilience levels for future years.

In 2020, when the COVID-19 pandemic swept the world, healthcare workers found themselves at the frontlines and needed resilience more than ever to do their jobs against all adversity. The faculty experience survey was fielded out of schedule in July to check in with faculty, when COVID-19 cases (including pediatric) had been continuously increasing, protective personal equipment supplies were limited, lockdowns had been imposed, and child care options became increasingly limited. We included questions related to how health care providers feel the hospital is performing due to the pandemic and how the pandemic impacts them personally.

The AAPOR RR2 response rate was 62.9% (929/1,478). Analyses will show through descriptive statistics how healthcare providers felt during the first wave of the pandemic about how they think the hospital has reacted to the pandemic and how the pandemic affected themselves. Bivariate analyses will assess what demographic and work characteristics are significantly related to positive vs. negative attitudes and experiences. In addition, on a departmental-level we will explore how well resilience and response behavior data from the 2019 survey is able to predict levels of positive/negative attitudes during the COVID-19 pandemic in 2020.

Examining Racial Differences in Willingness to Seek Healthcare During the COVID-19 Pandemic

Daniel Harwell, *American Institutes for Research*

Jen Durow, *American Institutes for Research*

Rachel Shapiro, *American Institutes for Research*

Sarah Mossburg, *American Institutes for Research*

Jazmine Orazi, *American Institutes for Research*

Krishna Winfrey, *American Institutes for Research*

Gina Sgro, *American Institutes for Research*

Alane Bristow, *American Institutes for Research*

Since early in the COVID-19 pandemic, researchers have identified racial disparities showing that the pandemic has disproportionately impacted persons of color. A Kaiser Family Foundation analysis found that Black and Hispanic patients had significantly higher rates of infection, hospitalization, and death due to COVID-19 compared to non-Hispanic, White patients. Persons of color are also more likely to have chronic conditions that put them at higher risk for complications due to COVID-19. However, persons with these same chronic conditions also require ongoing care to ensure that they are being properly treated; this ongoing care often entails in-person visits to allow for a physical examination, blood work, or other tests.

In October 2020, the American Institutes for Research administered a survey of the United States population using an online, probability-based panel to assess individuals' willingness to seek in-person healthcare during the COVID-19 pandemic for themselves, as well as children and other adults for whom they provide care. The survey also explored barriers and facilitators to seeking in-person care. We will present the findings of this survey, focusing specifically on differences between Black, Hispanic, and White respondents in seeking in-person care for themselves and the individuals for whom they provide care.

LGBTQ Disparities in Health and Social Determinants of Health in a National Monitoring Survey of Americans During the COVID-19 Pandemic

Rachel Kinder, *ICF*

Deirdre Middleton, *ICF*

Thomas Brassell, *ICF*

Matt Jans, *ICF*

John Boyle, *ICF*

James Dayton, *ICF*

Given the wide range of health disparities observed among sexual orientation and gender identity minorities, it is important to assess impacts on this population during the COVID-19 pandemic. Increased incidence of mental health issues in the general population has been reported during the COVID-19 pandemic, and raises concern about mental health impacts in the LGBTQ population.

The ICF COVID-19 Monitor Survey of U.S. Adults is a national survey that began in March of 2020 and obtained snapshots of public attitudes and behaviors over the course of the COVID-19 pandemic. It collected approximately 1,000 completed interviews a month using a non-probability, Census-balanced mobile panel. For the analysis of outcomes in LGBTQ respondents, waves are combined to increase the sample size and analytic power.

Interim results show three times as many LGBTQ respondents as the non-LGBTQ respondents reporting a diagnosis of the virus in their household. Nearly half (46%) of LGBTQ Americans reported moderate or severe symptoms of psychological distress during the two weeks prior to surveys. By contrast, only 21% of other Americans reported moderate to severe symptoms of psychological distress during the same period. Additionally, a larger percentage of LGBTQ respondents reported job and income loss, and while more than half of non-LGBTQ respondents reported "no difficulty" paying the rent or mortgage, only 36% of LGBTQ respondents could say the same. The LGBTQ population suffered from health disparities prior to the pandemic, and these data indicate this population continues to be vulnerable. These interim data will be combined with additional waves from this survey as data collection continues. Multivariate analyses will be used to assess if these differences are statistically significant when controlling for potentially confounding variables such as age. Analyses will be weighted to reflect the general US population.

Privacy Attitudes in Times of Crisis: Acceptance of Data Sharing for Public Health

Frederic Gerdon, *University of Mannheim*

Helen Nissenbaum, *Cornell Tech*

Ruben Bach, *University of Mannheim*

Frauke Kreuter, *University of Maryland, Joint Program in Survey Methodology*

Contact tracing data and health data collected from wearables and other devices are likely extremely helpful in tackling the COVID-19 pandemic. However, a privacy issue arises if data gathering systems put into place during the pandemic persist beyond the pandemic and if they ignore appropriate flows of information during non-pandemic times. We do not know how public preferences regarding the sharing of health data will develop in the future, but we can compare preferences measured before the pandemic with preferences measured during the pandemic. With such data, we answer the question whether people's acceptance to share health data for the sake of public health changed during the COVID-19 pandemic.

In 2019, we happened to conduct a survey experiment in Germany that examined acceptance of sharing data for a public good purpose. One of the scenarios described to respondents mentioned the spread of an infectious disease. We repeated the survey experiment April 2020 during the first peak of the pandemic in Germany. Data shows that respondents' acceptance to share data for public health purposes increased in response to the crisis, while acceptance to share other non-health data types remained mostly unchanged. The data presented here are a good reminder for policymakers to carefully consider the intended purpose of and appropriate limitations on data use.

Tracking Public Health, Utilization and Outcomes During a Pandemic: ICF COVID-19 Monthly Monitoring Surveys

John Boyle, *ICF*

Thomas Brassell, *ICF*

James Dayton, *ICF*

Surveys can be a critical tool in monitoring public health during emergencies. Existing surveillance systems provide timely reporting of cases and deaths associated with diseases. However, during COVID-19 they did not provide accurate information on virus-related symptoms and health behaviors in the population. In fact, the surge of potentially infected individuals seeking diagnosis, testing and treatment represented a serious but largely unmeasured dimension of the crisis. From March through December 2020, we conducted a tracking survey with approximately 1,000 completed interviews each month with national Census-balanced surveys of adults from a non-probability sample of approximately 2 million panelists. These surveys monitored physical health, mental health, employment and finances, lifestyle, mitigation, institutional confidence and future expectations across the course of the pandemic.

Respondents were questioned about their current and past COVID-related symptoms, whether they had seen a doctor, had a confirmatory test for the virus, type of test, test results, and time to get test results. Approximately 10% of adults reported current COVID-like symptoms and nearly a quarter had been sick for three days or longer with what they thought might be COVID since January. These numbers dwarf the 0.3% in April and 4.4% in December surveys with COVID positive test results. Nonetheless, many of these others sought medical care and testing during the crisis, and this contributes to both attitudes toward the risk of the disease and the importance of mitigation efforts, as well as strains on the health care system. Although surveys are not necessary to establish rates of confirmed cases in the population, they provide the missing data on the proportion of symptomatic cases in the population, the proportion seeking medical care for their symptoms, ability to obtain diagnoses with confirmatory testing, time to get results, and reasons for not seeking care or testing.

Experience Using the Qualtrics Web Panel for a Cross-Sectional Survey

Jennifer Gerber, *Johns Hopkins Bloomberg School of Public Health, RTI-SRD*

Janesse Brewer, *Johns Hopkins Bloomberg School of Public Health*

Rupali J. Limaye, *Johns Hopkins Bloomberg School of Public Health*

Andrea Sutherland, *Johns Hopkins Bloomberg School of Public Health*

Madeleine Blunt, *Johns Hopkins Bloomberg School of Public Health*

Taylor A. Holroyd, *Johns Hopkins Bloomberg School of Public Health*

Gail Geller, *Johns Hopkins Bloomberg School of Public Health*

Bruce Carleton, *British Columbia Children's Hospital, University of British Columbia*

Daniel A. Salmon, *Johns Hopkins Bloomberg School of Public Health*

Web panel respondents are used to receiving compensation for taking surveys, not for follow-up phone interviews. We aimed to conduct phone interviews with 20 survey respondents as part of a pre-test. After revising the survey based on their feedback, we aimed to collect an additional 1,925 responses between November 2019 and February 2020. During the pretest, respondents were shown a consent statement, asked for their phone number and email address. We noted they would be emailed a \$20 Amazon gift card after completing a 20-30-minute-long phone interview about their survey taking experience. We aimed to recruit individuals of varied sociodemographic backgrounds, but did use quotas due to the small sample size. Several of the 131 pretest respondents provided fake contact information, having no apparent intention of completing the phone interview. Some pretest respondents volunteered how many surveys they took per day and characteristics about their lifestyle.

In the follow-on survey, we used quotas based on the Census, American Community Survey, and Current Population Survey to enroll a sociodemographically representative sample of the U.S. Survey items measured vaccine attitudes and beliefs, prioritization of vulnerable populations during a vaccine shortage, and trust in public health authorities, all of which are relevant to measuring willingness to accept COVID-19 vaccines and prioritization of limited vaccine and therapeutic supplies. Due to difficulty filling quotas for young parents, racial/ethnic minorities, and people from Western states, we ignored quotas when enrolling the last 400 (approximate) of 1,925 respondents. Qualtrics underrepresents those without home Internet access and does not collect information about nonrespondents, limiting our ability to assess potential selection bias. This presentation will interest anyone considering using a web panel or measuring attitudes and beliefs related to COVID-19 vaccines and therapeutics.

Measuring Immigration Attitudes and Experiences

Immigrationalization of Redistributive Preferences

Michelle Io-Low, *Stony Brook University*

Research looking at race and welfare has traditionally focused on black-white relations. The prominent studies in this literature show that White Americans' racial biases towards Black Americans are consistently associated with their attitudes towards welfare. While there is an undeniable link between White Americans' racial attitudes towards Black Americans and their welfare preferences, recent studies have emerged to study welfare attitudes outside this Black-White dichotomy. In this project, I examine how White Americans' attitudes towards immigrants affect their redistributive preferences. Utilizing panel data from The American Panel Survey (TAPS), I construct longitudinal structural equation models to demonstrate how individuals' 2012 immigration attitudes is associated with their 2016 redistributive preferences. Using panel data, I navigate around issues related to omitted variable bias and reverse causality. My results reveal that the more positive a given individual feels about immigration in 2012, the more supportive they are of redistribution in 2016. I discuss the implications of these findings, particularly how this relates to contemporary anti-immigrant discourse.

Humanitarian Frames and Americans' Attitudes Toward Syrian Refugee Resettlement

Rita Nassar, *Indiana University*

A rich body of literature has shown the ease with which prejudicial attitudes can be triggered. Less work has been done to examine how inclusive attitudes toward outgroups can be promoted. This article looks at whether humanitarian frames can reduce opposition to outgroups by leveraging the case of Syrian refugee resettlement. Using a survey experiment embedded in the 2019 Cooperative Congressional Election Study, I examine whether different types of humanitarian frames are effective at reducing Americans' opposition to Syrian refugee resettlement. I find that frames focusing on the struggle of refugees but also concluding in a positive way, as is common in NGO stories, are more effective at reducing opposition to refugee resettlement than those that put the struggles of war at the center of their focus. In line with previous work, I find a general preference for Christian over Muslim refugees. Humanitarian frames are however found to reduce opposition to both Christian and Muslim Syrian refugee resettlement.

Anti-Immigrant Attitudes and White Support for Anti-Democratic Norms in the United States

Beyza Buyuker, *University of Illinois Chicago*

We argue that in a racially and ethnically diverse society such as the U.S., public support for anti-democratic norms among the white majority may be driven, in part by threat related to immigrants. Following racial priming theory, we suggest that for members of the white majority who harbor strong negative priors about immigrants, immigrant-origin groups' demands for democratic inclusion may be perceived in zero-sum terms. We further argue that threat to the white cultural primacy by outgroups may drive anti-immigrant whites to support anti-democratic norms that may better protect the elevated social status of whites over democratic norms that extend equality and cultural representation to racialized immigrant groups. We test this theory via two original threat priming experiments and an LDV analysis using 2017-2018 panel data from the VSG. Our findings suggest that when threat from immigrants is salient, it causes declining support for democratic norms among anti-immigrant whites. The LDV models also show that anti-immigrant attitudes are a strong and positive predictor of support for anti-democratic norms even accounting for the lagged DV.

Spatial Variation in Public Support for the US–Mexico Border Wall

Timothy B. Gravelle, *SurveyMonkey*

A signature policy of former US President Donald Trump was his plan to build a wall the length of the US–Mexico border. While the existing research has identified several political and demographic correlates of support for (or opposition to) the wall, there are contradictory findings relating to the effect of proximity to the US–Mexico border. Existing research also does not provide estimates of local support. Using multilevel regression and synthetic poststratification (MrsP) and data from public opinion surveys conducted between 2016 and 2020 by the Pew Research Center and SurveyMonkey, this letter presents county-level estimates for support for the US–Mexico border wall. The results demonstrate that while a majority of the American public opposes the construction of the wall, there is substantial variation in county-level support. Support for the wall is highest in areas where Trump received strong support in the 2016 election; support also increases with closer proximity to the US–Mexico border.

It's Complicated: Within Household Proxy Reporting Across Languages and Household Types

Angie O'Brien, *U.S. Census Bureau*

Surveys often rely on respondents to report on the behaviors or characteristics of other household members, otherwise known as proxy reporting. It seems reasonable that the accuracy of these proxy reports may be greater within nuclear families than within larger and more complex households, due to more intimate knowledge about close family members (Schwede, 2005). However, it remains unclear how this accuracy may vary by topic or by type of relationship. We define complex households as those with people other than nuclear family members (parents and their children) living together. In 2019, we conducted cognitive interviewing of a household survey in a large U.S. city in English and Spanish. A number of questions required in-depth knowledge of personal issues about other household residents. Some examples are: income, receipt of public benefits, assets, debt, and questions on sexual orientation and gender identity. In this talk, we examine responses individuals gave about other people who lived with them, comparing those who answered for nuclear family members (who are presumably more likely to know details about each other) and those who answered for people with whom they had other types of relationships. We note challenges respondents experienced with proxy reporting for others in their households, along with differences across languages. Additionally, we interviewed three sets of respondents who lived together (one set from a nuclear family household and two sets from complex households) and we examine whether they answered the questions accurately about each other and whether they answered similarly about their other co-residents. Our results could lead to insights on how accurate and reliable proxy reports are across different topics and different types of households. Furthermore, we aim to provide methodological recommendations on how to best maximize the utility of proxy reporting in household survey designs.

Understanding Mode Effects

Comparing Propensity and Extended Mixed Mode Methods to Estimate Mode Effects

David Cantor, *Westat*

Andrew Caporaso, *Westat*

To estimate mode effects it is necessary to separate channel of communication from the propensity to use particular modes (selection effects). One method relies on predicting and controlling for the propensity to respond by a particular mode. This approach relies on accurate predictors of these propensities. A different approach, proposed by Vannieuwenhuyze and colleagues (2010, 2012), use parallel surveys that separately estimate the selection and mode effects by assuming measurement and representative equivalence between the parallel surveys (Extended MM). They show estimates using propensity models differ from those using the Extended MM. The purpose of this paper is to compare these methodologies to estimate selection and mode effects using three parallel surveys. The analysis will use data collected as part of the Health Information National Trends Survey (HINTS). The parallel surveys administered the HINTS as a single mode paper survey and two mixed mode surveys using web and paper modes. One survey gave respondents a choice between web and paper (web choice). The second offered the same choice but also an incentive to respond by web (web bonus). The three surveys had the same response rates (approximately 30%). Twice as many respondents to the web bonus survey used the web when compared to the web choice method. Prior analysis of mode effects for this experiment found several instances of possible mode effects (Cantor, et al., 2020) when examining 40 selected measures using methods that control for propensity to respond. The analysis will estimate mode effects for an expanded set of outcomes using both the propensity and the Extended MM approaches. Two different comparisons will be discussed. One will compare the estimates using propensity and the Extended MM. The second compares the estimates from the two mixed mode surveys that differ by the proportion who used the web.

Mode Differences Across Four Countries and 15 Elections/Referendums: 2020 and Beyond

Claire Durand, *Université De Montreal*

Timothy Johnson, *University of Illinois Chicago*

From the 1970's to 2000, RDD telephone polls among landline phones gradually replaced face-to-face surveys and became the most common mode of administration for electoral polls in most Western democracies. The turn of the 21st century, however, witnessed increasing barriers and costs to reaching people by telephone and the emergence of new technologies which led to the development of new data collection modes. Primary among these were online polls which mostly relied on opt-in panels of volunteers, and Interactive Voice Response (IVR) polls of randomly sampled telephone numbers.

Over the past 20 years, both online and IVR polls have become common alternatives for the collection of pre-election polling data. The use of these new modes, however, remains controversial with concerns often expressed regarding their accuracy and bias relative to traditional interviewer-assisted polls. In this paper we report pooled analyses from 15 elections/referendums conducted across four countries (Canada, France, United Kingdom and United States) between 2005 and 2020 to evaluate the accuracy and bias of around 1,300 interviewer-assisted telephone and face to face polls, online and IVR surveys in representing nationwide election outcomes. Comparisons between modes are performed for the leading party, the left-most and the right-most party and control for national, election-specific and poll-specific variables. We present empirical findings on the differences between modes, in estimates and in trends, according to the contexts of different elections. We conclude with a discussion of the implications for the future of election polling across countries.

Comparing Data Quality Among Telephone, Web, and Mixed Modes for Shape Your Future Cross-Sectional Survey

Jonathan Baldwin, *University of Oklahoma Health Sciences Center*

Sixia Chen, *University of Oklahoma Health Sciences Center*

Ashley White, *University of Oklahoma Health Sciences Center*

Laura Beebe, *University of Oklahoma Health Sciences Center*

Recently response rates for telephone surveys have declined causing a shift to other modes including web-based surveys and mixed mode surveys. In this paper, we compare mode effects on the Shape Your Future Cross-sectional Survey (SYF) data quality among telephone (SYF 2018), web (SYF 2019), and mixed modes (SYF 2020). The goal of this health communications campaign is to improve the health of Oklahomans by addressing physical inactivity, poor nutrition, obesity and tobacco use. Survey respondents were Oklahoma adults with children living in the home. Data collection was funded by the Oklahoma Tobacco Settlement Endowment Trust. Methods: Outcome and demographic data on each version of the SYF were assessed for differences in missing percentage using a Kruskal Wallis test. Unweighted SYF demographic and outcome data were compared across survey modes, and compared to BRFSS 2018 weighted estimates by Euclidian distance.

All analysis was completed using SAS 9.4 procedures, and statistical tests were adjusted for multiple comparisons to maintain an overall 5% alpha rate. Results: Missing percentages were different across survey modes. ($p=0.0072$) The mixed methods mode had the largest missing percentage (4.6%) compared to the telephone (2.7%) and web based mode (1.9%). Demographic estimates differed, considering gender, the telephone survey estimated 38.4% male, the web based 15.8% male, the mixed methods 39.3% male, and the 2018 BRFSS estimated 46.6% male. Additionally, respondents on the online survey indicated a much lower physical activity time. BMI estimates on the telephone and mixed methods modes were close to BRFSS estimates, but web based estimates were approximately 3 points higher. Lastly, the cost of the web-based survey was approximately 47% less compared to the cost of the telephone survey. Conclusion: There is evidence the survey mode has an effect data quality, which could cause misallocation of health improvement resources.

A Simulation Evaluating Three Techniques for Accounting for Selection in a Mixed-Mode Survey

David Biagas, *U.S. Bureau of Labor Statistics*

When collecting household survey data, non-random subsets of respondents often complete mixed-mode surveys in their preferred mode. Estimates of mode effects, or differences in survey outcomes observed across various modes of completion, are thereby conflated with selection effects—the fact that a non-random and often very different set of respondents complete the survey in their preferred mode. This paper compares different approaches for accounting for selection effects when assessing mode effects, using data from the 2018 Consumer Expenditure Interview Survey (CEQ). The CEQ is designed to be completed in-person, but is also administered by telephone at the request of respondents or if deemed necessary by field representatives to elicit survey cooperation from respondents. Three different approaches that can be used to adjust for selection effects are compared: 1) controlling for potential differences in respondent characteristics; 2) creating selection effect propensity adjustment scores from a logistic regression model; and 3) creating selection effect propensity adjustment scores from a random forest model. While previous work compared these approaches, there was no gold standard for which to evaluate the accuracy of these different approaches. The current study improves upon previous work by randomly imposing mode effects for which to better assess the accuracy of these approaches. The results of this paper highlight the importance of adjusting for selection effects when assessing mode effects and compare the effectiveness of three different adjustment techniques.

Impact of Mode Design on Estimates of Change – A Quasi-Experimental Approach

Alexandru Cernat, *The University Of Manchester*

Joseph W. Sakshaug, *Institute for Employment Research (IAB), Ludwig Maximilian University of Munich, University of Mannheim, Germany*

Longitudinal surveys are the main data collection tool used to estimate change in the social sciences. Recent developments have accelerated the move from traditional single mode longitudinal studies to mixed mode designs. Nevertheless, there is a concern regarding the possible impact of implementing a mixed mode data collection on estimates of change. Such a design may inflate coefficients of change due to the change in the mode of answering at the individual level. In this study we use the Understanding Society Innovation Panel, a representative yearly panel from the UK to investigate the impact of mode design on estimates of change. We carry out two types of comparisons. Firstly, we compare a single face to face design to a sequential web-face to face design based on a random allocation of individuals. Secondly, we compare face to face answers to web ones using weights to correct for selection into mode. By analysing 41 variables we find that the mean and variance of estimates of change is the same across groups. Nevertheless, the correlation between intercept and slope, an estimate of convergence in the data, is significantly different across groups for most of the variables. We discuss implications of these findings for panel data in general.

Considerations in Data Harmonization for Multimode Surveys

Hanna Popick, *Westat*

Mina Muller, *Ipsos Public Affairs*

Eric Jodts, *Westat*

This presentation will address data harmonization challenges in multimode studies and data approaches to consider when a multimode data collection obtains inconsistent data across modes. Paper and web surveys each have different strengths, so there can be a benefit to offering each of these options to respondents during a data collection effort. The first part of the presentation will address the process of harmonizing the data from multiple modes, including receiving more than one response from the same respondent, and assigning final result codes. The second part will focus on different data strategies that can be applied when respondents using a paper survey provide responses that are inconsistent with the data from the web survey. One benefit of a web survey is that it enforces logic and response constraints, such as allowing only one response to a question, using skips to present only applicable questions, and applying edit checks to address any inconsistencies. Though a paper survey can be designed to maximize clear instructions, one cannot prevent a respondent from selecting more than one response on a single-response question, for example. The presentation will cover different question types, provide examples and address considerations that should be taken while editing the data based on the intended goal of the question and how the data will be used.

Integration of Administrative Records and Survey Data

Being Agile During a Pandemic: The American Community Survey Approach

Sandra Clark, *U.S. Census Bureau*

The American Community Survey (ACS) is the largest household survey in the United States. It is the only source of comparable, quality information about the people in our nation's communities. The ACS provides current demographic, social, economic, and housing information about America's communities and is used to aid in the distribution of more than \$675 billion a year in federal funds. Increasing nonresponse rates, natural disasters, and pandemics present challenges to the ACS and make it more difficult to obtain survey response. The COVID-19 pandemic exemplified these challenges, forcing the program to research alternative strategies.

Building closures, temporary staff reductions, and suspended or limited operations made it difficult to send mail materials, receive and process incoming paper questionnaires, and conduct in-person interviews. All of which had a significant impact on response rates. This panel will address how the ACS plans to use alternative data sources to help address some of the challenges related to obtaining data from federal prisons, predicting housing unit status, and adjusting for nonresponse bias, all of which were exasperated by COVID-19. Additionally, we will demonstrate the Census Bureau's COVID-19 Dashboard, which provides current demographic and economic information related to the pandemic all in one spot. This effort ensures that data users can quickly find and access the data they need to make informed, data-driven decisions.

Predicting Vacant Housing Units in the American Community Survey

Andrew Keller, *U.S. Census Bureau*

The American Community Survey (ACS) spends resources to determine if a nonresponding address is occupied or vacant. We develop a model for predicting ACS vacancy using survey paradata, information about the neighborhood, and administrative data. The goal is to use this vacancy information to make decisions regarding ACS operations. For example, field visits can be reduced for units modeled as vacant. The quality of model predictions are analyzed retrospectively against outcomes from the field. We examine differences between predictions and field results. The purpose is to understand the characteristics of the occupied units misclassified as vacant. Another part of the analysis is to see how ACS estimates might change if occupied units are incorrectly classified as vacant.

Prisons Shutdown Too: Using Administrative Records to Replace In-Person Interviews of the Federally Incarcerated

R. Chase Sawyer, *U.S. Census Bureau*

Sandra Clark, *U.S. Census Bureau*

The devastating effects of the global COVID-19 pandemic have altered lives across the world. Survey data collections have not been immune to these effects, including those conducted by the U.S. Census Bureau. One major concern for the American Community Survey (ACS) is the collection of group quarters data, particularly the data of incarcerated individuals. These data collection efforts have been limited and in some cases halted to hamper the spread of COVID-19.

These difficulties have not stopped the desire of the ACS to innovate and continue publishing high quality estimates about America's people and economy. Given the current pandemic, the ACS cannot collect data from the usual interviews of individuals in federal prisons. To remedy this loss of data, the ACS is working to utilize administrative records from the Bureau of Prisons (BOP) as a replacement to survey responses. This presentation will detail the steps the ACS is taking to harmonize the administrative record data with the survey questions used during in-person interviews. It will also look at some preliminary results for item missing data rates and other survey metrics. Overall, viewers will be able to learn about how the nation's largest household survey is innovating to adapt to a global pandemic and how administrative records may be a suitable replacement for survey responses.

Survey Isolation During COVID-19: The Effects of Suddenly Relying on Address-Matched Phone Numbers for Interviewing Households.

Jonathan Eggleston, *U.S. Census Bureau*

John Voorheis, *U.S. Census Bureau*

Tim Trudell, *U.S. Census Bureau*

Yarissa Gonzalez, *U.S. Census Bureau*

The COVID-19 pandemic greatly impacted data collection for household surveys, curtailing in-person interviewing across the world. For example, the U.S. Census Bureau conducted numerous in-person surveys before the pandemic, but temporarily switched to phone data collection for these surveys in March of 2020. To contact newly-sampled households, the Census Bureau relied on vendor-provided phone numbers matched to addresses. This sudden change in interviewing protocol and generally the effects of the pandemic on respondent behavior may have profound effects on who is able to be contacted for Census Bureau surveys (noncontact bias) and who is willing to respond conditional on being contacted (refusal bias). To investigate these biases, we use the Current Population Survey and the American Community Survey. We first use administrative data matched to sampled addresses to see how the characteristics of respondents and nonrespondent in 2020 compare to prior years. Next, we focus specifically on the extent to which phone number match quality can explain changes in noncontact bias, as these phone numbers were the primary method of contact. Preliminary results suggest that the lower a household's income is, the less likely the household is to have their phone numbers correctly matched to their current address. This correlation potentially introduces a nonresponse bias with respect to income that was not present before in these surveys. We discuss the reasons for some households having lower match quality as well as a possible correction for this bias through weighting.

Census Data for COVID-19

Andrew Hait, *U.S. Census Bureau*

The Census Bureau provides a wealth of data that policy makers can use to better understand the potential impacts of the coronavirus pandemic on the nation's people and businesses, but only the Economic Indicators provide the most timely data to measure the ACTUAL impact. To provide more comprehensive data on both American businesses AND household, another approach was needed.

This session will provide an overview of 4 weekly surveys that the Census Bureau created to measure these impacts; the Household and Small Business Pulse Surveys, the Weekly Business Formation Statistics program, and Community Resilience Estimates. From their inception, the process to create them, get approval from OMB, to collect and analyze the data, and to disseminate the data was unlike any other program conducted by Census.

This session will also provide a live demo of the COVID-19 Data Hub, a portal that Census Bureau staff created with Esri to disseminate key data from selected annual Census programs as well as these new weekly datasets. This platform allows a wide range of users to access and use these key statistics in their work. The demo will also include views of data visualizations created by these programs. Finally, the session will include a glimpse into the future plans for these programs and Hub, including the future addition of more datasets to guide the recovery from the pandemic.

Measuring Health Insurance in the American Community Survey: Assessing the New Premium and Subsidy Items

Joanne Pascale, *U.S. Census Bureau*

Angela R. Fertig, *University of Minnesota*

Kathleen Thiede Call, *University of Minnesota*

Since 2008, the American Community Survey (ACS) has included a module to measure health insurance coverage. In 2019, items about premiums and subsidies were added to the ACS questionnaire with the aim of using the survey to produce a measure of subsidized marketplace coverage. We use an experimental study to examine the accuracy of survey items in the ACS health insurance module including the new premium and subsidy questions. Data come from the CHIME survey, a reverse record check study where households with individuals enrolled in public and private health insurance plans (non-group, subsidized marketplace, unsubsidized marketplace, Medicaid, and a state-specific public program called MinnesotaCare) were administered a telephone survey that included the ACS health insurance module and the new premium and subsidy items. Survey data were matched back to enrollment records, which indicated coverage status at the time of the survey. We examine concordance between the survey and the records on coverage type (e.g., direct-purchase, Medicaid) and presence of a premium and subsidy. Preliminary results indicate that enrollees in non-group coverage outside the marketplace, and subsidized marketplace enrollees, usually report their coverage type as “direct purchase” (84.7% and 85.7%, respectively). Similarly, Medicaid enrollees reported coverage type as “public” 84.5% of the time. However, unsubsidized marketplace enrollees reported coverage type as “direct purchase” 60.9% of the time, and employer-sponsored insurance 20.8% of the time. Private coverage enrollees reported a premium 95.3-100% of the time, and Medicaid enrollees reported no premium 74.6% of the time. Non-group and unsubsidized marketplace enrollees usually reported no subsidy (93.0% and 80.9%, respectively), and subsidized marketplace enrollees reported a subsidy 72.4% of the time. We also explored various methods of categorizing coverage type, including subsidized marketplace, using the full range of survey items available.

Linked Survey and Transactional Data for Financial Health Measurement: Early Findings

Andrew Dunn, *Financial Health Network*

Thea Garon, *Financial Health Network*

Necati Celik, *Financial Health Network*

Helen Robb, *Financial Health Network*

The Financial Health Pulse is an ongoing annual survey of U.S. financial health that relies on a custom survey administered to the Understanding America Study panel. In collaboration with research partners, we invited survey panelists to share financial transaction data, yielding a sample of approximately 800 individuals for whom we have both survey and transactional data. As financial health measurement becomes even more important during the COVID-19 pandemic, these data provide us with alternative measures of financial health. With the ability to tie participants' transactional data to their responses in Pulse surveys, we compare survey responses to observed linked accounts. We explore where transactional data can complement surveys and where the two data sources align. For example, of those who linked a transactional account, a majority of participants (85%) linked a checking account and reported owning one in the survey. This percentage drops to 68% and 77% for savings accounts and credit cards respectively.

Furthermore, we begin to examine the extent to which transactional data can complement survey data during critical times. Transactional data and responses from three surveys fielded during the COVID-19 pandemic show directional consistency as individuals' financial lives were disrupted during the COVID-19 pandemic. For instance, a 6% increase in the proportion of respondents with 3 months of savings between 2019 and 2020 correlates with an increase in median liquid savings of \$1,553 between February and July 2020 in our transactional dataset. These early findings demonstrate utility in using transactional data to measure the financial health of individuals; transactional accounts can provide valuable insights into the financial health of households that closely align with their self-reported survey responses, allowing for near real-time assessment of financial health. However, that alignment can be sensitive to how transactional measures are defined and sample composition differences.

Using Administrative Data and Machine Learning to Address Nonresponse Bias in Establishment Surveys

Benjamin Kufner, *Institute for Employment Research (IAB)*

Joseph W. Sakshaug, *Institute for Employment Research (IAB)*

Stefan Zins, *Institute for Employment Research (IAB)*

In recent years, participation rates have been declining for the IAB Job Vacancy Survey (JVS), one of the largest establishment surveys in Germany (N=110,000), which mainly aims to quantify the size of the unfilled labor demand. The declining response rates pose a risk for increasing non-response bias. In establishment surveys, non-response analyses and non-response correction are usually limited to use only a few auxiliary variables and apply simple models. We overcome these limitations by using the Establishment History Panel (BHP), a rich administrative data set on the population of all employing establishments in Germany covering a variety of employer and employee profile characteristics. This enables us not only to test more theory-driven hypotheses, but also to assess nonresponse bias over time. In the final step of our analysis, we use estimated response propensities from regression and machine-learning models to construct establishment-level weights to evaluate model performance with respect to bias reduction. For the evaluation, we present aggregate bias measures for the long-run trend in addition to individual estimates of bias. The results from these analyses give insights to what extent we can measure and potentially reduce non-response bias using a rich administrative data set as well as data-driven machine learning methods. Summarizing, our paper will shed light on whether these additional tools and data are worthwhile for reducing nonresponse bias and furthermore provide a blueprint for other establishment surveys on how they might use big data methods to improve their nonresponse adjustment procedures.

Question Wording Experiments

How Much Do They Agree? Examining Variation in Agree-Disagree and Item-Specific Questions Based on Data Quality Indicators and Question Characteristics

Jennifer Dykema, *University of Wisconsin-Madison*

Dana Garbarski, *Loyola University Chicago*

Nora Cate Schaeffer, *University of Wisconsin-Madison*

Nadia Assad, *University of Wisconsin-Madison*

Cameron Jones, *University of Wisconsin-Madison*

Tiffany Neman, *University of Wisconsin-Madison*

Agree-disagree (AD) questions (e.g., “I am extremely satisfied: strongly agree...strongly disagree”) are among the most frequently used response formats to assess subjective phenomena because they are easy to write and efficient to administer. These positive features, however, may be offset by increased cognitive burden for respondents, which may reduce data quality and increase response effects. In recent research, questionnaire designers advocate for item-specific (IS) questions that directly ask about an item’s underlying response dimension and use relevant response categories (e.g., “How satisfied are you: not at all...extremely”). An IS item’s direct method of questioning is argued to be less cognitively burdensome and less likely to produce response effects. However, despite strong recommendations to use IS questions, only a limited number of studies experimentally compare AD and IS questions and their results present a somewhat mixed picture regarding which format is associated with higher data quality. One reason for these mixed results is that AD and IS questions vary not just based on their basic questioning manner, but also based on other question characteristics that may affect respondents’ cognitive processing and subsequent data quality.

To further explore data quality issues for AD and IS questions, we identified questions included in previous experimental studies (approximately 300 AD-IS question pairs) and coded the questions using two question evaluation tools, the Survey Quality Predictor (SQP) and Question Understanding Aid (QUAID). We then conducted a structural analysis of the features of the AD and IS questions to identify and code major dimensions along which the questions varied (e.g., their response dimensions, polarity, number and labeling of response categories, etc.). We use data quality scores from SQP and QUAID to predict which response format is associated with higher data quality and explore whether select question characteristics moderate effects for this unique sample of questions.

Measuring Attitudes Towards Artificial Intelligence in Medicine. An Experimental Study in the LISS Panel Comparing Agree-Disagree Scales, Construct-Specific Scales, and Label-Points.

Marieke Haan, *University of Groningen*

Yfke Ongena, *University of Groningen*

Jeffrey Frikken, *University of Groningen*

Although psychometric properties in agree-disagree scales (ADS) have been advocated (Willits et al 2016), such scales may be more subject to acquiescence bias (Dykema et al 2019) and extreme responding (Krosnick and Presser 2010; Liu et al 2015). Through acquiescence and the repeated use of response categories, ADS may increase common method variance and are not necessarily better to measure the underlying construct. Questions that use a construct-specific scale, such as happy-sad, exciting-boring (CSS, also called item-specific, Hhne and Lenzner 2017), alert the respondent to the underlying response dimension. However, the intent of CSS may be less clear than ADS, since ADS remind respondents that an evaluation is being asked for, while CSS may give the impression that knowledge or facts are desired. Also, CSS are often designed as end-label only items (e.g., happy-sad) because they often do not allow for true midpoint labels. Therefore, not much is known about the effects of (fully)-labeled CSS versus ADS on data quality.

We conducted a 2*2 experiment in a questionnaire measuring attitudes towards artificial intelligence in medicine (CS versus AD and end-labeled versus fully-labeled scales) to study how the different scale versions affected data quality. Respondents were randomly assigned to one of the conditions. Data were collected in April 2020 using the LISS Panel which is representative for the Dutch population (n=2411).

Based on the literature we expected that: 1. The CSS will yield responses with higher reliability and lower acquiescence- and ERS-bias than the ADS, and this difference will be larger for the fully-labeled scale than the end-labeled scale, and 2. The CSS will be associated with longer processing time than the ADS, and this difference will be larger for the fully-labeled scale than the end-labeled scale. Our initial analyses show a higher internal consistency for ADS than for CSS.

Does Adding a Third Gender Response Option Really Improve Measurement?

Ashley Amaya, *Pew Research Center*

Anna Brown, *Pew Research Center*

Emily Vogels, *Pew Research Center*

Accurately measuring gender (a social construct, different from the biological construct of sex) is critical to a variety of analyses in the social sciences (e.g., voting behavior, the wage gap). Historically, researchers have limited gender response options to 'male' and 'female'. This method excludes or inaccurately measures the estimated 1.4 million U.S. adults who do not identify as either a man or a woman or identify with a gender that is different from their sex. To promote inclusivity and accurate measurement, researchers have begun to include response options other than 'male' and 'female'. While the previous research has made large strides in methods to ask more inclusive gender questions, there is limited research on whether this change actually improves overall data quality. Many surveys in the U.S. are conducted in both English and Spanish, but little of the existing research has addressed how to phrase a revised gender question in Spanish. Moreover, 56% of U.S. adults oppose the inclusion of an option other than 'male' and 'female' on online forms. These individuals may be alienated and break off from the survey when they encounter a revised gender question. Finally, several studies have found the words (e.g., 'transgender') used in these new questions to be confusing.

This presentation addresses whether the inclusion of a third response option to gender actually improves data quality. Specifically, we used an experimental design with a Hispanic oversample to field various gender questions on an online non-probability panel to determine whether the inclusion of a third option 1) improved measurement, 2) was understood in English and Spanish, and 3) introduced nonresponse bias. We found that measurement was improved with minimal adverse effects and adopted a revised question to measure gender: Do you describe yourself as a man, a woman, or in some other way?

Obamacare and Trump's Trade War: How Political Cues Influence Survey Responses

Gabriel Miao Li, *University of Michigan*

Josh Pasek, *University of Michigan*

Survey methodologists have long known that different ways of phrasing the same question will solicit different answers. The present study aims to disentangle whether changing question wordings through attaching political cues will influence respondents' choices, and to assess which question(s) is better in eliciting the presumably "correct" responses. We hypothesize that the most knowledgeable respondents should also be the most capable ones to recognize the identity concealed by question wording variations, and we investigate what question will lead the less knowledgeable individuals to approximate the answers they would give if they were more knowledgeable. With two broad national samples of Americans, we conducted two online survey experiments examining the influence of question wording variation on individual's attitude toward three different public policies, namely healthcare, immigration, and foreign trade. In general, we find that attaching political cues to a survey question changes participants' reported attitudes. More importantly, with a novel method we proposed, political knowledge and partisanship are found to be important factors determining whether people could recognize the survey questions worded differently were measuring their attitudes toward the same issue. For two of the issues under investigation, immigration and foreign trade, as we expected, the more knowledgeable the individuals are, the more similar answers they give to the questions worded differently, suggesting that they are more capable to recognize the identity of different versions of the questions. The impact of political knowledge differs for Democrats and Republicans, depending on the issue's partisan ownership. However, the pattern is not substantiated for healthcare, which propels us to reconsider the potential difference between misperception and ignorance; and we find that for an issue like healthcare reform

which is highly divisive party-wise and surrounded by misinformation, the impact of knowledge is contingent on both partisanship and the misperception people hold about the given issue.

Forced Choice for All?

Rachel Horwitz, *U.S. Census Bureau*

Elizabeth Nichols, *U.S. Census Bureau*

There has been much debate in the survey community between using a select all that apply option or requiring respondents to select yes or no for each possible response (i.e. forced choice design) in web surveys. The select all that apply format is easier to display, particularly on mobile devices, and less burdensome for respondents in terms of completion times. However, research shows there are more selections made in the forced choice format than in select all, suggesting deeper cognitive processing as respondents attend better to each option. Much of the research on this topic involves similar types of questions asking attitudes, opinions, or events. Researchers have taken this “more is better” finding and have extended it to all types of multi-item questions, even those where data show many respondents only select one affirmative response.

The present study provides results from a web survey experiment testing the acquiescence rates between forced choice and select all formats on three types of questions. Included in the survey are a series of questions about the respondent’s current status where there is typically just one affirmative response (e.g., type of school a child attends), a series of questions that asks about events (e.g., activities done with children), which are likely to evoke multiple responses, and an opinion question. We have a twofold aim: (1) replicate the previous research for opinion and event questions where the forced choice design should elicit more affirmative responses and (2) expand the previous research to determine whether the same finding holds for factual questions with typically one affirmative answer. Additionally, we evaluate whether instructions on how to answer each question format (“select all that apply” and “select Yes or No for each option”) have an impact on the amount of data provided.

How to Ask About Caste Attitudes in India: Results From a Split Form Experiment

Kelsey Starr, *Pew Research Center*

Neha Sahgal, *Pew Research Center*

Jonathan Evans, *Pew Research Center*

Ariana Monique Salazar, *Pew Research Center*

Manolo Corichi Gomez, *Pew Research Center*

One of the hardest challenges that face survey researchers is bridging the gap between research concepts and respondent understanding. For example, surveys in India commonly ask attitudes about lower caste groups – these groups have legal definitions that may not totally overlap with how the public commonly speaks about them. For example, Scheduled Castes and Scheduled Tribes have separate legal definitions in India, but are often spoken about as one group by the general public. When asking public opinion about these groups, should researchers ask about them together or separately? Asking about these groups together saves questionnaire space and is in line with the public discourse on lower castes, but it is also possible that both terms may not be equally familiar to the respondents, and people may have different opinions about each group. To answer this question and others, Pew Research Center ran a pilot study (n= 2,045) in four Indian states in preparation for a larger national survey. Alongside testing for question sensitivity and fieldwork practices, the pilot study included split-form experiments to test if there was an advantage to asking how people felt about discrimination against Scheduled Castes, Scheduled Tribes and Other Backward Classes and about each group separately. In this paper, we discuss the results of the experiment, how they influenced decisions for the national survey and lessons learned that can be applied to other surveys in the country.

Measuring Gender: Testing Alternatives to Traditional measures

Dan Cassino, *Fairleigh Dickinson University*

Survey research is finally moving away from conflating gender identity and sex, aligning our practice and measures with the accepted theoretical understanding that sex and gender are separate (though correlated measures). This measure is especially important for researchers looking at political and social attitudes, as people with traditional gender identities have very different attitudes and reported behaviors than those with less traditional gender identities, even controlling for factors like sex, age and partisanship. However, the best-validated measures of gender identity, like the Bem Sex Role Inventory, are unwieldy, and too long to be incorporated into most surveys. This paper makes use of a novel survey experiment to test several measures of gender identity against gold standard measures like the BSRI in order to determine which best function as a potential replacement for the BSRI in surveys, as well as which best predict political and social attitudes.

What is in a Word: Accidents, Leaking and Losing

Melissa Constantine, *University of Minnesota*

Charles Cain, *University of Minnesota*

Sara Putnam, *University of Minnesota*

Leslie Rickey, *Yale University*

Todd Rockwood, *University of Minnesota*

Emily Lucacz, *University of California San Diego*

Diane Newman, *University of Pennsylvania*

Linda Brubaker, *University of California San Diego*

Kyle Rudser, *University of Minnesota*

Incontinence research is a subject for which question threat is high. Fecal incontinence research has routinely used the terminology of 'accidental leakage' since the late 1990s (Rockwood) to reduce question threat. Urinary incontinence (UI) is presumed to be less threatening than fecal, but it is still embarrassing and potentially stigmatizing and thus, likely to be affected by terminology used in the question. Two experiments were evaluated in an e-panel sample of women (n=654). First is a wording experiment, in which two commonly used questions in UI research are compared against an item specifically developed to reduce question threat in fecal incontinence (accident).

- In the past year, have you ever leaked urine, even just a drop or two?
- In the past year, have you ever lost urine, even just a drop or two?
- In the past year, have you ever accidentally leaked urine, even just a drop or two?

The second experiment was limited solely to the accidentally leaked urine item in which endorsement is compared between a dichotomous response options and vague quantifiers, which included a declarative "Yes" prior to each category. Dichotomous: No, not even once or Yes; or Vague Quantifiers: "No, not even once, Yes, but only once or twice", "Yes, and it has happened several times" or "Yes, and it has happened many times".

The use of the term accident significantly increased endorsement (accidental 68%, leaked 61% and lost 60%, $p > .1$ one tailed). Furthermore, the use of vague quantifiers in which there are three levels of yes, increase endorsement an additional 10% for the accidental item (78%, $p > .1$ one tailed). This study shows that contrary to assumptions currently made in UI research, the threat of UI is still significant enough to warrant the use of terminology, such as accidental that reduces question threat.

Web Survey Participation and Recruitment

Long Term Effect of Methods to Improve Subsequent Survey Response After Recruitment for a Multi-Purpose Probability Sample Panel

Kanru Xia, *NORC at the University of Chicago*

Lindsay Liebert, *NORC at the University of Chicago*

Ipek Bilgen, *NORC at the University of Chicago*

Vicki J. Pineau, *NORC at the University of Chicago*

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

Obtaining response to surveys from panel samples has unique challenges since a panelist's experience and different survey interventions they receive over the course of panel tenure can influence their cooperation behavior. Probability based sample panels devote significant resources both to recruiting panelists and ensuring high levels of cooperation in completing assigned surveys after panel recruitment. Accordingly in early 2020, a field test examined different methods to improve panelist survey cooperation using AmeriSpeak®, a representative probability based sample panel of U.S. households. The tested methods included the use of advance letter mailings; use of pre-paid incentives in advance letter mailings; tailored design of advance letter mailings; and the inclusion/exclusion in advance letter mailings of instructions on accessing surveys directly using a custom URL or accessing the panel portal log-in. Results showed that the use of advance letter mailings increased cooperation rates by four percentage points, but the pre-paid incentives did not have significant impacts. Furthermore, generic design had the highest cooperation rate among the three tested mailing designs, which were generic, detailed, and with URL instructions, when no pre-paid incentives were given. These results showed the effectiveness of pre-notification for a particular survey. However, the unique challenge, and often advantage, of sample panels is that a panelist's experience can have long term effects on their cooperation behavior, because panelists are repeatedly recontacted to complete surveys. Thus, this study will also analyze if there are any long term impacts of the tested methods on panel retention and panelists' willingness to complete subsequent surveys assigned over time, as well as the duration of such potential effects. This study will also assess if there are diminishing returns when a panelist receives multiple treatments over time.

Nonresponse Follow-up Study of Online Probability-Based Panel Recruitment

Frances M. Barlas, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

Mansour Fahimi, *MSG*

Past research has found that a low response rate does not necessarily serve as an indicator of nonresponse bias in survey estimates. However, there is still cause for concern that differential nonresponse or systematic exclusion of segments of the population are greater threats to representativeness. Probability-based online panels typically recruit new panelists using an initial recruitment survey. Once recruited, nonresponse investigations have focused on differences between panel members who respond to a given online survey and those panel members who did not respond, often comparing demographics and attitudinal aspects that might differ. Such examination of differential nonresponse at the survey stage is more common than examining the potential for differential nonresponse at the recruitment stage when building the panel. We conducted a mailed nonresponse follow-up study to investigate the possibility of nonresponse bias during recruitment to Ipsos' KnowledgePanel, the largest probability-based online panel in the US that is built using address-based sampling. We created a new communication protocol, with a different incentive structure, and with different branding unrelated to the KnowledgePanel recruitment. Shortly after completing a wave of recruitment, we initiated the non-response survey, sending a survey to households we had successfully recruited in that wave of recruitment as well as sending a survey to nonresponding households. We had responses from 866 nonrespondents (a 26.2% completion rate) and 673 responses from recruited households (a 67.4% completion rate). We examined primary demographics, secondary demographics, and several external benchmarks to compare the differences between responders and non-responders. We found that across most of the variables we evaluated, households recruited to KP exhibit compositions almost identical to those of households not recruited. Differences between recruited and

nonrecruited households were rarely statistically significant. We review some of differences in terms of how recruitment can be improved to improve the robust representativeness of the panel.

A Tale of Seven Web-Only Government-Sponsored Surveys

Kris Pate, *Internal Revenue Service*

Scott Leary, *Internal Revenue Service*

Brenda Schafer, *Internal Revenue Service*

Rizwan Javaid, *Internal Revenue Service*

Patrick Langetieg, *Internal Revenue Service*

Martha Stapleton, *Westat*

Kerry Levin, *Westat*

It's generally assumed that web-only surveys will perform better with businesses than with households, because businesses are generally tech-savvy and accustomed to conducting business online. However, Kashihara (2019)¹ found that certain businesses prefer to respond by paper. Reaching individuals can be difficult, because even internet savvy individuals may be uncomfortable completing web surveys or find even the web-friendliest mobile surveys daunting. They may simply be wary of providing information online. Also, although a majority of individuals have smartphones and laptops, research shows that individuals living in rural areas or age 65 or older are much less likely to have internet access.²

The IRS surveys taxpayers to gather information about the time and money they spend to comply with their tax reporting responsibilities. These data collections have in the past been conducted as multi-mode (paper and web) surveys. Given that increasingly more surveys are conducted via web,³ the IRS is exploring moving some or all of its burden surveys to be web-only. However, while web-only surveys generally cost less to administer than paper surveys,³ data quality and potential nonresponse bias must be considered.

This paper will look at seven web-only surveys recently fielded by the IRS to individual taxpayers, small tax-exempt organizations, businesses affected by the Tax Cuts and Jobs Act, pension plan administrators, and filers of estate tax, gift tax, and excise tax returns. We will compare response rates across surveys and between within sample strata, as well as compare nonresponse bias between the web surveys. The paper will also compare the web-only response rate to other recent individual and business multi-mode surveys conducted by the IRS and discuss whether differences such as the availability of data or topic sensitivity may influence response rate. We will also analyze and compare nonresponse bias between the web surveys.

Web Invitation Pre-Incentives: Preliminary Findings From a Pre-Test With College Students

Jodi Walton, *Abt Associates*

David Jernigan, *Boston University*

Finding ways to maximize response rates has long been a focus of survey research efforts in order to reduce nonresponse bias. Incentives can help establish participant interest and goodwill, create a sense of reciprocity and reinforce the legitimacy of a research effort. Using a combination of both prepaid and contingent incentives is often recommended, based on research using mail, in-person and phone recruitment methods. The effectiveness of using prepaid incentives for a web survey where email invitations are sent to a selected respondent group is an area where additional research is merited.

The upcoming Boston University Harms to Others from Drinking Among College Students study provides an opportunity to explore whether prepaid incentives included in email survey invitations may impact response rates. A pre-test was conducted with Boston University sophomore and junior students in spring 2020 (data collection concluded prior to BU's spring break before the onset of the pandemic). An experiment was designed offering \$2.50 prepaid incentives to half of those invited (51 students among 100 selected). The link to the incentive was embedded in the invitation email, but was not mentioned in the subject line. While the target response rate of approximately 50% was easily achieved, pre-incentives didn't seem to affect survey completion rates. Half of the completed surveys were from those receiving the incentive, while half came from those without an incentive. The initial response to the invitation was

much better than expected, since 52 surveys were completed without requiring any reminders or follow-up invitations. This presentation will discuss characteristics related to redemption of offered incentives and the response rate, as well as opportunities to explore the impact of prepaid incentives during data collection scheduled for Fall 2021. Funding for the BU Harms to Others from Drinking Among College Students study is provided by NIH grant 5R01AA025980-03.

Mobile Optimization Strategies: Effects on Web Survey Participation and Fieldwork Progression

Marc Asensio Manjon, *University of Lausanne*

Caroline Roberts, *University of Lausanne*

Jessica M. E. Herzing, *University of Lausanne*

In recent years, smartphone penetration has continuously increased around the world and a growing part of the population relies exclusively on smartphones to access the internet. Coverage rates for web surveys are better than ever, but to maximize response rates, efforts must be made to adapt survey designs to accommodate 'smartphone dependent' participants. Furthermore, to capitalize on new data collection opportunities offered by mobile devices, there is an interest in actively encouraging and normalizing responding to surveys on mobiles in the general population. A range of mobile optimization strategies are available for this purpose, but not much is known about their relative effectiveness and impact on survey costs and errors. We address this question in the present study, through a comparison of three strategies to optimize mobile device experience in web surveys, which were implemented in a probability-based, three-wave election study conducted in Switzerland in 2019: (1) the standard approach of providing a URL to a browser-based survey and optimizing the display of the questionnaire on smartphones (N= 8000); (2) adapting the invitation to promote mobile response and providing a QR code to access the survey (N= 1088); and (3) providing a QR code to download and participate via a smartphone application (N=1087). We compare the three mobile optimization strategies to draw conclusions about their relative impact on a) response rates and sample composition – overall and on mobile devices; b) estimates for target survey variables; and c) the progression of fieldwork, to draw conclusions about which strategy provides the best balance in terms of cost efficiency and representation of the target population.

Do They Exist? Experiment to Assess Panel Conditioning and Opinionation Effects in AmeriSpeak® Panel Surveys

Vicki Pineau, *NORC at the University of Chicago*

David Dutwin, *NORC at the University of Chicago*

Ipek Bilgen, *NORC at the University of Chicago*

In this study, we define panel conditioning as the effect observed in a person's survey responses that is influenced by their panel tenure and panel experiences. The principal aim of the study is to investigate potential conditioning effects on survey responses in an AmeriSpeak Panel survey fielded one year apart, seeking to understand whether survey data quality improves or degrades as a result. Secondly, the same data will be used to study the prevalence of systematic nonresponse in opinion questions, assessing changes in the existence and/or strength of opinions. We will focus on investigating three assumptions related to panel conditioning and/or opinionation: (1) Cross-sectional differences in survey responses between more and less tenured panelists between the first and second fielding of the survey may be due to panel conditioning and/or differential attrition, (2) Repeated and/or longitudinal differences between responses to the same survey to the same respondents fielded 12 months apart may be due to panel conditioning, and (3) Differences in responses to opinion questions may be due to panel conditioning and/or hard-to-recruit status. This paper will present findings from a randomized experiment and sample matching to make high/low tenure as equivalent as possible (as is also true for easy/hard-to-reach respondents). The statistical response matching approach is meant to disentangle attitude and behavior change from panel composition change due to panel attrition and will control for several variables such as age, education, NRFU status, etc. Some assessments will be true experiments and others, quasi-experiments. For both experimental and quasi-experimental tests, we utilize sample matching between groups.

Controlling Errors Associated With Probability and Nonprobability Samples

General Purpose Multiply Robust Data Integration Procedures for Handling Non-Probability Samples

Sixia Chen, *University of Oklahoma Health Sciences Center*

David Haziza, *University of Ottawa*

Recently, there has been an increased interest in combining probability and nonprobability samples. Non-probability samples are cheaper and quicker to conduct. However, the resulting estimators are vulnerable to bias as the selection probabilities are unknown. To adjust for the potential bias, estimation procedures based on parametric or nonparametric models have been discussed in the literature. The validity of the resulting estimators depends on the validity of the underlying models. Also, nonparametric approaches suffer from the curse of dimensionality and poor efficiency. We propose a data integration approach by combining multiple outcome regression models and propensity score models. The proposed approach can be used for estimating general parameters including totals, means, distribution functions and percentiles. The resulting estimators are multiply robust in the sense that they remain consistent if all but one model are misspecified. The asymptotic properties of point and variance estimators will be established. The results from a simulation study and a real application show the benefits of the proposed method in terms of bias and efficiency.

Doubly Robust Inference for Count Data With Varying Exposures in Non-Probability Samples: A Model-Based Approach Based on Gaussian Processes of Propensity Prediction

Ali Rafei, *University of Michigan*

Carol A. C. Flannagan, *University of Michigan Transportation Research Institute*

Michael R. Elliott, *University of Michigan*

The ubiquitous availability of large-scale unstructured data has led to a growing interest in the use of such data for producing official statistics. However, the non-probabilistic nature of the sampling mechanism raises serious concern over the potential selection bias when making finite population inference. In the presence of a relevant probability sample, one can use augmented inverse propensity weighting to make doubly robust inference by combining the idea of propensity modeling with that of prediction modeling. However, this method may no longer be applicable when the rate of a rare event is of interest per unit exposure and the average exposure is substantially different across the two datasets. In addition, as a general drawback of design-based approaches, this adjustment method may perform poorly when there is evidence of influential pseudo-weights. We propose an alternative model-based approach using a partially linear Gaussian process regression, which remains doubly robust under these circumstances. It utilizes a negative binomial prediction model with a flexible function of the estimated pseudo-selection probabilities as a predictor to impute the rare outcome for non-sampled units of the population, and the varying exposure is treated as offset in the model. We show that the Gaussian process regression behaves as a non-parametric matching technique based on the estimated propensity scores. As a second advantage, our proposed method can be implemented under a Bayesian framework. Using the 2017 National Household Travel Survey as a benchmark, we apply our method to the naturalistic driving data from the second phase of the Strategic Highway Research Program (SHRP2) to estimate the traffic accident rates per driven mile and per calendar year in the United States.

Getting it Right: The Accuracy of Sampling Frame Data and Predicting Survey Contact

Julie Pacer, *Abt Associates*

Marci Schalk, *Abt Associates*

Stas Kolenikov, *Abt Associates*

Establishment surveys are susceptible to a unique set of challenges compared to household surveys. Unlike a household survey, an establishment survey collects information from a representative of a business who speaks for that business, usually during business hours. Nonresponse is a considerable factor in establishment surveys. The sampling frame may lack a contact name at the sampled business or it may provide outdated contact information. In attempting to reach a business by phone, the rise of automated menus makes it challenging to reach a person to identify a respondent. Furthermore,

depending on the role of the intended respondent and size of the business, the intended respondent may not respond to a survey invitation due to competing work duties.

Abt Associates conducted the Family and Medical Leave Act (FMLA) Wave 4 Surveys for the U.S. Department of Labor, which includes a survey of employers and their Human Resources benefits administrators regarding their establishments' use and understanding of the FMLA. The screening portion of the survey verified the existence of the business and identified a respondent. The survey utilized the Dun & Bradstreet Market Identifiers file as a sampling frame, and the auxiliary frame data allowed for an in-depth analysis of survey eligibility and response propensity based on sampling frame information. In this paper, we determine the accuracy of this sampling frame information when we had contact with the establishment, particularly contact information, job title, state, business size, and industry. We also use sampling frame variables to predict survey contact using logistic regression and random forest models. The results will inform recommendations for future use of the sampling frame in establishment surveys.

New Approach for Handling Drop Point Addresses in Mail/Web Surveys

Rachel Harter, *RTI International*

Joseph McMichael, *RTI International*

S. Grace Deng, *U.S. Energy Information Administration*

A drop point is a single USPS delivery point or receptacle that services multiple businesses, families, or households (USPS CDS User Guide). Residential drop units are the individual housing units served by the drop point address. For the most part, address-based sampling frames list the number of units at a drop point address but will not contain information identifying specific units. Drop units comprise less than 2% of all residential addresses in the United States but they tend to be concentrated in certain large cities. In Queens NY, for example, drop units constitute almost 28% of address units. The problem with drop units for address-based surveys with mail contacts is that, without names or unit identifiers, there is no way to control which unit receives the various mailings. This leads to distorted selection probabilities, renders the use of cash incentives by mail impractical, and precludes traditional methods for mail nonresponse follow-up. Alternatively, excluding drop units results in coverage error, which can be considerable for some subnational estimates. The authors propose a substitution approach when a drop unit is sampled—replacing the unit with a nearby unit in a non-drop point building containing the same number of units. This paper discusses the selection of the substitute, the theoretical underpinnings of this approach, and some of the issues encountered when implementing the approach for major national surveys including the 2020 Residential Energy Consumption Survey.

Equity and Inclusion: Do Our Weighting Algorithms Really Bring the Voice of Marginalized Populations to the Table?

Erin Caldwell, *Polco*

Michelle Kobayashi, *Polco*

With all of the national emphasis on equity and inclusion, it is important that the methods we devote to creating representative samples of opinion perform as promised. When we weight, are we really doing a better job of representing those with less of a voice or are we simply inflating the response of people who are similar to the folks already completing surveys? Are we making false assumptions about what it means to be like? To answer this question, we mined our database of 100+ matched probability and non-probability pairs of administrations of The National Community Survey™ (The NCS). The NCS is a templated, benchmark used by local governments across the nation to measure quality of life, community livability, local government performance and civic participation. Probability-based surveys have been aggregated in this database since 1997 and non-probability surveys (a second survey completed alongside the probability surveys) since 2013. The database includes the data for more than 850 communities (probability samples) and 100+ matched pairs of probability and non-probability samples.

Through this analysis, we will: 1) Better understand how non-response varies by survey mode and which modes do a better job at reaching those traditionally hard to reach. 2) Investigate the variability in response of populations subgroups often combined in weighting algorithm, and 3) Quantify the overall

impacts of weighting on survey results based on population subgroup calibrations. Results analyzed at the both the national and local community level provide insight into the areas of weighting where we need to ensure we are, in fact, helping all voices be heard.

Outcomes of Salted and Suppressed Addresses in the 2020 Census Address Canvassing Operation

Nancy Johnson, *U.S. Census Bureau*

Laura Norena, *U.S. Census Bureau*

Jennifer Hutnick, *U.S. Census Bureau*

Ahmad Harvey, *U.S. Census Bureau*

As part of the original design of the 2020 Census, the Census Bureau identified four major innovation areas from the 2010 Census. This included the 2020 Census Address Canvassing, which was reengineered to include a suite of operations, conducted both in the field and in the office, which updated the address list and map data for the 2020 enumeration. This presentation focuses on the field component of the address canvassing operation and, specifically, how often listers correctly added and erroneously missed valid addresses from the Census Bureau's preliminary address list that were suppressed from their workloads as well as how often listers correctly deleted or failed to delete fabricated addresses that were "salted" into their workloads. Analysis is stratified by urban/rural and address categorizations. Overall, we find that listers were less successful at taking correct action for suppressed addresses compared to the sample of salted addresses, although correct action for both types of addresses was taken relatively less often in rural areas.

Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)

Andrew Seidenberg, *National Cancer Institute*

Rick Moser, *National Cancer Institute*

Brady T. West, *University of Michigan*

Methodological issues pertaining to transparency and analytic error have been widely documented for publications featuring analysis of complex sample survey data and contribute to problems with rigor and reproducibility. The availability of numerous public use datasets to researchers without adequate training in using these data likely contributes to these problems. In an effort to introduce standards for reporting analyses of survey data and promote replication, we propose PRICSSA (Preferred Reporting Items for Complex Sample Survey Analysis), an itemized checklist to guide researchers publishing analyses using complex sample survey data. PRICSSA is modeled after other checklists (e.g., PRISMA, CONSORT) that have been widely adopted in other research domains. Our initial draft of PRICSSA items includes a variety of survey characteristics, such as mode, data collection dates, response rate, and sample selection process. Additionally, essential analytic information is also included, such as sample sizes for all estimates, missing data rates and imputation methods (if applicable), disclosing if any data were deleted, specifying what survey weight and sample design variables were used along with methods of variance estimation (including how strata with a single PSU were handled, if applicable), and reporting design-adjusted standard errors/confidence intervals for all estimates. PRICSSA also recommends that authors make all corresponding software code available as an online supplement to any publication. To improve upon our initial draft list of items, we plan to seek input from survey methodologists, survey researchers, and survey data providers, and make PRICSSA available for public comment. Once finalized, we plan to inform journal editors about the utility of PRICSSA, publish PRICSSA in multiple peer-reviewed journals, and distribute PRICSSA via professional societies (including AAPOR) and survey research email lists. We suggest that the widespread adoption of PRICSSA will help greatly improve secondary analyses of complex sample survey data and promote scientific rigor and replication.

5/12/2021, 11:00 AM - 12:30 PM
Concurrent Session C: Live Sessions

Mini-Conference: AAPOR Task Force Report on the Performance of 2020 Pre-Election Polls

AAPOR Task Force Report on the Performance of 2020 Pre-Election Polls

Joshua Clinton, *Vanderbilt University*

Jennifer Agiesta, *CNN*

Camille Burge, *Villanova University*

Megan Brennan, *Gallup*

Majorie Connelly, *NORC at the University of Chicago*

Ariel Edwards-Levy, *HuffPost*

Bernard Fraga, *Emory University*

Emily Guskin, *Washington Post*

D. Sunshine Hillygus, *Duke University*

Chris Jackson, *Ipsos*

Jeffery Jones, *Gallup*

Scott Keeter, *Pew Research Center*

Kabir Khanna, *CBS*

John Lapinski, *University of Pennsylvania*

Lydia Saad, *Gallup*

Daron Shaw, *University of Texas-Austin*

Andrew Smith, *University of New Hampshire*

David Wilson, *University of Delaware*

Christopher Wlezien, *University of Texas at Austin*

This is a discussion of preliminary findings of the AAPOR Ad Hoc Task Force on the Performance of Pre-Election Polls. It will discuss how well pre-election polls did at predicting the outcome of presidential, gubernatorial and senatorial contests in the 2020 general election. It will also examine some potential explanations for what happened.

Mini-Conference: Blood From a Stone? Using Qualitative Techniques to Understand Why Households Don't Participate in Surveys (QUALPOR Panel)

Blood From a Stone? Using Qualitative Techniques to Understand Why Households Don't Participate in Surveys (QUALPOR Panel)

Margaret Roller, *Roller Research*

The purpose of the AAPOR Qualitative Public Opinion Research (QUALPOR) Affinity Group is to advance the discipline of qualitative research within AAPOR by offering a forum in which AAPOR members can share best practices, exchange resources, and discuss new approaches to qualitative research design as well as ways to improve on the integrity of these designs along with the quality and usefulness of the outcomes. In keeping with its mission, QUALPOR is pleased to submit a five-paper panel for the 2021 AAPOR conference.

This panel consists of five presentations that focus on the methods and results of a qualitative study undertaken by the National Center for Education Statistics to understand the drivers of nonresponse in a self-administered household survey. The study included both observations of nonresponding addresses and in-depth, semi-structured interviews with over 80 nonrespondents to the National Household Education Survey (NHES).

In the first presentation, Andrew Zukerberg introduces the study and provides details on the methodology. The second presentation, by Ashley Kaiser, discusses the observation component of the study which focused on determining the characteristics of addresses that are prone to nonresponse, including NHES mailings that are undeliverable as addressed, and to assess the accuracy of the information on the sampling frame. The next two presentations discuss the interview findings. In the third presentation, Mahi Megra talks about participants' mail receiving and processing behaviors and the results of an activity in which participants sorted and reviewed 12 pieces of example mail. The fourth presentation by Rebecca Medway discusses participants' attitudes towards survey research and their self-reported engagement with the NHES contact materials. The final presentation by Melissa Scardaville combines the observation and interview findings to describe the seven typologies that offer insight into the primary reasons why participants in their study chose not to respond to the NHES.

A Methodology for Understanding Non Response in a Household Survey

Andrew Zukerberg, *National Center for Education Statistics*

Rebecca Medway, *American Institutes for Research*

Melissa Scardaville, *American Institutes for Research*

Ashley Kaiser, *American Institutes for Research*

Mahi Megra, *American Institutes for Research*

Response rates continue to fall on household surveys. Falling response rates and rising cost have led some researchers to utilize address based samples and mail based contacts with the household. While the frames for address based samples provide some information on the characteristics of address, it is often not accurate. Thus, researchers are left with little information on non responding households. 2019 NHES conducted a two part qualitative follow up study with non responding addresses to learn more about the characteristics of households that did not respond and the factors that went into the household's decision not to respond. The first part of the study was an observation of the non responding addresses in seven geographic areas. The second phase involved sending trained ethnographic researchers to a subsample of non responding addresses in four areas to conduct in depth interviews with the person in the household that handles the mail. This component looked at the visual cues that respondents focus on in processing mail and the larger context of how non responding households view the world and how this impacts their decision (not) to participate in surveys. The studies were novel in the fact that they focused on actual mail non responders and sought to understand the context around the decision not to participate in the study. This presentation provides details on the methodology used in the two phases, including how they were selected and the protocol development and provides context for

other presentations proposed as part of a panel. It will be useful on its own to researchers that seek to better understand non respondents in their studies

You've Got Mail: Using a Mock Mail Sorting Activity With Nonrespondents to a National Household Survey to Better Understand Mail Receipt and Processing Behavior

Mahlet Megra, *American Institutes for Research*

Melissa Scardaville, *American Institutes for Research*

Danielle Battle, *American Institutes for Research*

Rachel Hanson, *American Institutes for Research*

The reliance on mail-based data collection using address-based sampling to survey U.S. households has increased in recent years (Harter et al., 2016). Following this trend, the National Household Education Surveys program (NHES) has used mailed invitations to contact households since the 2012 administration. Even before the household member evaluates the mailed survey invitation, gaining response from a household to a mail request requires that a) the survey reaches the household successfully, and b) the household member opens the envelope.

To evaluate the current NHES mailing procedures and to optimize the design of future administrations, a qualitative interview study was conducted with over 80 households that did not respond to the first three NHES:2019 mailings. Interview participants were asked about receiving mail (e.g., how often they check mail, mail delivery challenges they face) and about their general mail-handling process (e.g., when and how they sort mail, how they decide whether or not to open specific pieces of mail). As part of a mock mail sorting activity, interview participants reviewed 12 pieces of example mail, including an NHES:2019 mailing, that varied on several aspects, such as the type of mail (e.g., bill, solicitation), envelope design (e.g., color, size), and postage type. Interviewers observed household member(s) sorting these mail pieces and prompted household members to capture their thoughts about what they were doing and why.

This presentation provides a summary of participants' mail receiving and processing behaviors and explores themes in their reactions to the mail pieces. For example, participants considered factors such as the degree of personalization of the addressee line, familiarity with/perception of the sender, and the size of the envelope when deciding whether to engage with the mail pieces. Findings from the study will inform mail survey protocols for future NHES administrations and other mailed surveys.

What You See is What You Get: Address Observations in a Nonresponse Follow-up Study

Ashley Kaiser, *American Institutes for Research*

Andrew Zukerberg, *National Center for Education Statistics*

Paul J. Lavrakas, *Independent Consultant*

Mahi Megra, *American Institutes for Research*

Sidney Wilkinson-Flicker, *American Institutes for Research*

To better understand characteristics of nonresponding households, 760 address and neighborhood observations were conducted across the United States as part of a nonresponse follow-up study to the 2019 National Household Education Survey (NHES). The NHES is a nationally representative household survey that uses an address-based sampling (ABS) frame to collect information about educational experiences of children in the U.S. In 2019, an in-person follow-up was conducted to understand more about nonresponding addresses. The purpose of the observation component was to determine the characteristics of addresses that are prone to nonresponse or having their NHES mailings be undeliverable as addressed (UAA) and to assess the accuracy of the information available on the NHES ABS frame.

The observation instrument included items designed to measure the accuracy of the frame (e.g., occupancy status, presence of children, and household income) and challenges in receiving mail (e.g., structure type, mail access type). It also included a series of items that captured characteristics of the address that were not available on the frame (e.g., presence of items related to privacy or pride in education). For each item, observers recorded rich contextual information about the house and neighborhood. This presentation shares the observable characteristics of nonrespondent and UAA

addresses. Indicators of privacy concerns were observed for just over a quarter of observed nonrespondent addresses. Addresses with UAA outcomes were “problematic” (e.g., no mailbox in sight) more often than non-UAA addresses. Finally, there was a considerable range in the agreement rates between the observed data and the ABS frame. Drawing on these findings, the presentation will conclude with implications for the contact materials and procedures used in future NHES data collections and other household surveys that use an ABS frame.

Exploring Experiences With and Attitudes Towards Survey Research: Results of Qualitative Interviews With Nonrespondents to a National Household Survey

Christopher Paek, *American Institutes for Research*

Danielle Battle, *American Institutes for Research*

Cameron McPhee, *SSRS*

Over the past few decades, surveys have faced persistent declines in response rates (Brick and Williams 2013). Given the global nature of this phenomenon, researchers have devoted considerable attention to understanding the factors that contribute to it. However, the drivers of nonresponse are complex and much remains unknown about sample members’ reasons for nonresponse and the best way to counteract those concerns. In response to these trends, the National Center for Education Statistics conducted an in-depth qualitative study focused on better understanding the drivers of nonresponse to the National Household Education Survey (NHES) and other household surveys. The study included over 80 semi-structured interviews with households that did not respond to the first three NHES:2019 survey mailings. These 90-minute interviews covered a variety of topics (e.g., survey experiences, government attitudes, mail sorting behavior). They took place in sample members’ homes at four study sites across the country.

This presentation focuses on three topics. First, we discuss participants’ prior experiences with and general attitudes toward surveys; for example, commonly reported reasons for negative attitudes towards surveys included being too busy, having survey fatigue, and feeling like survey participation doesn’t make a difference. Next, we discuss the survey-specific factors that participants said influence their survey participation decisions, such as the survey topic, offered response mode(s), length and incentives. Third, we report the extent to which interview participants said they had engaged with the NHES survey mailings; while most remembered receiving at least one mailing, more than half of those who opened at least one mailing said they had rejected the survey request. We conclude by discussing the implications of this qualitative interview study for understanding nonresponse to the NHES and similar household surveys.

What Drives Survey Nonresponse? Using Qualitative Research With Households That Do Not Respond to a National Government Survey to Develop Typologies of Nonresponse

Melissa Scardaville, *American Institutes for Research*

Extensive research has been done in the survey methodology field to try to understand why people have become less likely to respond to surveys. Theories of nonresponse range from individual-level drivers such as busyness and privacy concerns to societal-level factors, such as declining confidence in public institutions and growing concerns about security and identity theft (ASA Task Force on Improving the Climate for Surveys 2017). However, precisely why people decline, and the degree to which these drivers vary across households from diverse sociodemographic backgrounds, remains unclear.

Using data from over 80 unstructured interviews conducted with households that did not respond as of the third mailing to the 2019 National Household Education Survey (NHES:2019), the presentation outlines the combination of factors that influenced interview participants to respond to the survey request. This study, rooted in the ethnographic tradition and the emic approach, combines 90-minute in-person interviews with household observations to explore the cultural frameworks participants use to interpret their world and guide behavior. Interviews were conducted with households in four geographically diverse locations and in English and Spanish.

The presentation describes seven typologies that offer insight into why these individuals chose not to respond. We examined participants' behaviors and attitudes across several key factors identified through analyzing interview and observational data. Each participant was placed in a single typology that explained the primary reason for nonresponse. For some typologies, excessive time constraints, having multiple adults in the household, and/or mail delivery challenges stopped participants from responding. For others, their beliefs about the federal government, concerns over data security, and/or assuming the NHES survey request was not relevant to them drove their decision. The findings can shed light on the various drivers that contribute to survey nonresponse to the NHES and to household surveys more generally.

Mini-Conference: Novel Data Collection Solutions During a Novel Coronavirus Pandemic

The Collection of Remote Biomeasures for the 2020-2021 Pretest of the National Social Life, Health, and Aging Project: Design and Mode Considerations

Ned English, *NORC at the University of Chicago*

Colm O'Muircheartaigh, *NORC at the University of Chicago*

Katie O'Doherty, *NORC at the University of Chicago*

Anna Wiencrot, *NORC at the University of Chicago*

Becki Curtis, *NORC at the University of Chicago*

Sara Walsh, *NORC at the University of Chicago*

Stephen Smith, *NORC at the University of Chicago*

Martha McClintock, *The University of Chicago*

The National Social Life, Health, and Aging Project (NSHAP) is a longitudinal, population-based study that examines the interaction between aging, social relationships, and health outcomes of older adults as detailed in O'Muircheartaigh et al. (2009, 2014). NSHAP has followed a nationally-representative probability sample of older American adults as well as their co-resident romantic partners, with Cohort 1 born from 1920-1947 and Cohort 2 born from 1948-1965. In each wave we collected biomeasures to complement self-reported estimates of health, including anthropometrics, sensory measures, frailty measures, as well as blood and other biological samples. While biomeasure collection has been conducted face-to-face by trained field interviewers during the first three waves of data collection (2005, 2010, and 2015), we designed and implemented an approach for remote self-administered biomeasure data collection during the pretest of the 2020-2021 wave. As part of our remote biomeasure pretest we designed experiments based on biomeasure administration time and specific biomeasure combinations. We present the importance of survey mode for biomeasure recruitment, administration time, biomeasure combinations, incentive, and respondent type on the feasibility of their remote collection for social science and health research surveys. Our results show the promise of collecting detailed information in non-face-to-face survey designs, which will see expanding prominence during and after COVID-19.

Venue-Based Sampling for a Hard-to-Reach Population During the COVID-19 Pandemic

Petry Ubri, *NORC at the University of Chicago*

Vicki Pineau, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

Sabrina Avripas, *NORC at the University of Chicago*

Roy Ahn, *NORC at the University of Chicago*

Ghenet Besera, *CDC, Division of Reproductive Health*

Mary Goodwin, *CDC, Division of Reproductive Health*

Venue-based sampling (VBS) involves recruiting from venues where members of the population of interest can be found. VBS is often used for hard-to-reach populations for which no sampling frame exists or for whom traditional sampling methods may be ineffective or increase risk for participants due to confidentiality concerns. In 2019 we successfully conducted a pilot study in one U.S. metropolitan community using VBS for a hard-to-reach population of women who were born or whose mothers were born in countries with a high prevalence of female genital mutilation/cutting (FGM/C). The purpose of our pilot was to test the methodology (VBS) and procedures (in-person interviewing and questionnaire instruments) to better understand experiences and attitudes about FGM/C as well as reproductive health and access to health care services. We worked with community organizations to identify physical venues and female interviewers with the language skills, community knowledge, and cultural sensitivity for recruiting at venues, enrolling participants, and conducting in-person interviews. Once the pilot was completed, the main study was planned for administration across 4-5 U.S. metropolitan communities in the fall of 2020.

The COVID-19 pandemic required modification of the main study VBS approach and interview mode. Because VBS venues' physical locations were closed, or operated at limited capacity, and presented

safety concerns for our interviewers and participants, the VBS recruitment strategy was re-designed to include contact with respondents through virtual venues such as listservs, virtual events or workshops, and phone or web-based communication with venue staff. Because in-person interviewing was also not possible, a telephone interview was used. This presentation summarizes modifications to the main study due to the pandemic. Lessons learned may be useful for studies using similar recruiting approaches for hard-to-reach populations. Procedure modifications may have implications for studies that rely on in-person recruitment and interviewing in the context of a pandemic.

Negotiating Survey Data Collection Challenges Posed by the COVID-19 Pandemic: Developing an Electronic Questionnaire Device for the European Social Survey

Rory Fitzgerald, *City University London*

Debbie Collins, *NatCen Social Research*

Joanna D'Ardenne, *NatCen Social Research*

May Doušak, *University of Ljubljana*

Maurice Martens, *CentERdata*

Arnaud Wijnant, *CentERdata*

The European Social Survey (ESS) is a biennial, large scale, cross-national general social survey that has been conducted using in-home face-to-face interviews with a representative sample of the general population since 2002. Since the COVID-19 pandemic struck, face-to-face interviewing has been paused in almost all ESS countries. Whilst it is hoped that fieldwork might restart in 2021 there remains considerable uncertainty if face-to-face interviewing will be possible in all countries in the short to medium term.

In response to this the ESS has commissioned the development of an ESS Electronic Questionnaire Device to allow respondents to complete the survey in a self-completion mode, without an interviewer present and without an internet connection. The device will be a tablet tailored so that the hour-long ESS questionnaire can be completed in digital format. It is being designed for simple touch screen, off-line completion in order that those without any experience of using the internet or even a computer can complete the survey. Survey agency staff would leave the device with households after completing screening at the doorstep reducing contact between interviewers and households to a minimum. The device might be used in parallel with on-line data collection where target respondents are able and willing to use their own device. Although the pandemic was the inspiration for this tool it is expected to have a longer term utility as a mode to help complement a move towards web surveys in comparative research.

This paper will outline the process of developing the device and provide results from user testing in 4 test countries in Europe. The authors will reflect on the possibility of using the device on a larger scale and the likely utility of the device as part of a longer term mixed mode future for the ESS.

The Mannheim Corona Study - Insights Gained Through Daily Data Collection and Data Augmentation

Carina Cornesse, *University of Mannheim*

Ulrich Krieger, *University of Mannheim*

The outbreak of COVID-19 has sparked a sudden demand for fast, frequent, and accurate data on the societal impact of the pandemic. At the onset of the pandemic, our interdisciplinary research group at the University of Mannheim therefore set up the Mannheim Corona Study (MCS), a rotating panel survey with daily data collection, on the basis of the long-standing probability-based online panel infrastructure of the German Internet Panel (GIP). In a team effort, our research group was able to inform political decision makers and the general public with key information to understand the social and economic developments from as early as March 2020 as well as advance social scientific knowledge through in-depth interdisciplinary research.

This presentation gives insights into the most compelling findings from the MCS. In particular, we will show how our daily data collection allowed us to explore societal developments that were so fast that they might have been missed entirely with less frequent data collection. Moreover, we will demonstrate the

manifold possibilities for augmenting the MCS data with data from other sources. For example, we will show how the data gathered in the GIP prior to the pandemic provided us with important information on how the MCS respondents' lives changed as a result of the pandemic. Furthermore, we will show how we gained insights into the association between the changes in people's lives and the epidemiological development of the pandemic by linking the MCS survey estimates to official COVID-19 statistics.

Opportunity in a Time of Crisis: Adapting an Annual Health Survey to Produce Monthly COVID-19 Data

Todd Hughes, *UCLA Center for Health Policy Research*

YuChing Yang, *UCLA Center for Health Policy Research*

Jacob Rosalez, *UCLA Center for Health Policy Research*

Ninez Ponce, *UCLA Center for Health Policy Research*

Arina Goyle, *SSRS*

Recurring annual surveys faced numerous challenges with the onset of the COVID-19 pandemic, but they also had to consider the opportunities it created. Could a recurring annual health survey with data collection occurring throughout 2020 adapt its survey content, weighting, and dissemination plans in order to produce close to real-time estimates that would be useful in informing the impacts of the pandemic?

The California Health Interview Survey (CHIS) was implemented in 2001, and has been conducted annually since 2011. The survey is designed to collect data across most of the calendar year, and then produce annual estimates at the county level for most areas in California, releasing those estimates in the Fall of the next year. With 2020 data collection underway prior to the start of the pandemic, and using methods that were able to continue despite the pandemic, the CHIS explored how to add COVID-19 questions to the survey mid-cycle and produce and release interim monthly estimates in a timely manner. This paper summarizes the steps CHIS took to adapt data collection, data processing, and data dissemination tools, including new dashboards to produce and release monthly 2020 estimates at the state level, as well as pooled estimates across months for individual counties and county groups, crossed by key demographic and health indicators.

5/12/2021, 2:00 PM - 4:00 PM
Concurrent Session D: Prerecorded Sessions

Language and Translation Issues in Questionnaire Design

Cognitive Interviewing for Test Questions: Special Considerations in Design and Translation

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

Daniela Glusberg, *Research Support Services Inc.*

Chandana Vuyyuru, *Feinberg School of Medicine, Northwestern University*

Evelyn Velazquez, *Feinberg School of Medicine, Northwestern University*

Stacy Bailey, *Feinberg School of Medicine, Northwestern University*

To minimize survey error, in survey design we aim to develop questions that are as simple and easy to understand as possible, while measuring what we intend to. The goal in item development for tests and assessments is related but, unlike survey questions, items have right and wrong answers, and difficulty is part of the design, not something to be avoided. In the survey industry we cognitively test survey questions and respondent materials to ensure they are interpreted by respondents as the question designers intended. For test items, cognitive interviewing is also a useful tool: it can help uncover if an item is interpreted differently not only by different individuals, but also across specific subgroups.

Cognitive testing is also a key tool for assessing the translation of survey questions, as well as test items, into new languages. In analyzing cognitive interviewing findings of test items, analysts must determine if items are not being interpreted as intended because of the formulation of the item, the language used, or because of cultural differences in the target population, rather than the intended difficulty the question was designed to uncover.

In our presentation, we will discuss issues around the cognitive testing of test items, and how it can diverge from cognitive testing of translations of survey questions. We will use examples from recent cognitive testing of the Spanish translation of items for a new digital health literacy tool. We will present interpretation issues we found in testing including translation issues, differing cultural interpretations, and participant errors in responses which do not represent item problems at all, as they are precisely what the items want to measure: participant's level of health literacy. We will discuss how we separated comprehension issues due to translation and cultural context from comprehension issues related to health literacy.

2020 Census Language Support Program: Overview and Findings From the Most Robust Program Built by the U.S. Census Bureau

Lily Kapaku, *U.S. Census Bureau*

The goal of the U.S. Census is to count everyone "once, only once, and in the right place". This includes the counting of limited-English-speaking (LES) households. In order to reach this goal and to provide LES households with an opportunity to respond immediately and in more languages than ever before, the U.S. Census Bureau developed the language support program, which is a significant expansion of the 2010 Census efforts, and is the most robust language program the agency has ever built. For the 2020 Census, respondents could respond online or by phone in English and 12 non-English languages (Spanish, Chinese, Vietnamese, Korean, Russian, Arabic, Tagalog, Polish, French, Haitian Creole, Portuguese, and Japanese). The Census Bureau also provided additional language support materials in 59 non-English languages. This includes both video and print language guides. Video language guides, narrated in 59 non-English languages, assist respondents in responding online. Print language guides, written in 59 non-English languages, assist respondents in filling out the paper questionnaire. By providing internet and telephone response options in English and 12 non-English languages and language guides in 59 non-English languages, the Census Bureau provided support to over 99% of all U.S. households. This presentation will focus on how the Census Bureau developed each of the aforementioned components and will highlight preliminary findings from the 2020 Census language support program.

Lost in Translation: Investigating Spanish-Speakers' Response Error on a Household Education Survey

Talia Kaatz, *American Institutes for Research*

Ashley Kaiser, *American Institutes for Research*

Maria Payri, *American Institutes for Research*

Multilingual and multicultural research is becoming increasingly relevant for survey researchers in the U.S. As the Hispanic population in the U.S. continues to grow both in size and as a percentage of the total population, the stakes to properly reach and represent this population increase. Hispanic Americans also are increasingly diverse in country of origin, where they reside, and their education experience. The languages they speak and their experiences with U.S. institutions and cultural concepts may vary by numerous factors (e.g., generational status, length of time since immigration, and country of origin). This diversity presents new challenges in questionnaire design and translation. The National Household Education Survey (NHES), used to produce national estimates on educational experiences, offers an example of these design and translation challenges. Using NHES data, this presentation aims to answer two main questions. First: Are the patterns observed in responses to the Spanish questionnaire also seen among Spanish-speaking respondents who complete the English questionnaire? Second: Among Spanish-speaking respondents, do response patterns vary across other indicators of English proficiency and immigration experience? This presentation will combine response data from NHES:2019 with cognitive interview findings for NHES:2023 to compare response patterns and indicators of response error by Spanish and English proficiency, immigration status/experience, and questionnaire language. Preliminary findings suggest that translation attempts for some items may be highly prone to measurement error, potentially because certain educational terms and concepts (e.g. homeschooling) are not familiar to many monolingual Spanish respondents. Additional preliminary analysis suggests disproportionately high rates of missing and inconsistent responses across these items among Spanish-speaking respondents. These findings will illustrate a mixed-method approach to identifying conceptually problematic items for translation and determining some underlying factors associated with Spanish-speaking respondents' understanding of key concepts. The presentation will attempt to identify areas of improvement for survey design.

How Does Back Translation Fare Against Team Translation? An Experimental Case Study in the Language Combination English-German

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

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

Comparability of questionnaire translations is a requirement for any sound use of cross-national or cross-cultural research data. The TRAPD translation approach that foresees several translations and interdisciplinary review discussions between translators and survey experts is regarded as best practice to ensure comparability (Harkness, 2003; Mohler et al., 2016). The TRAPD approach, conceived for the European Social Survey in 2002, was and is a clear departure from a simple back translation model that had dominated the cross-cultural landscape before. Despite the success of TRAPD (and related models, such as Pan; de la Puente, 2005), back translation is still in widespread use across data-collecting disciplines, and so the calls have been louder in recent years to provide empirical data on the usefulness (or lack thereof) of back translation and on the respective benefits of different translation approaches (e.g., McKenna; Doward, 2005). This paper will address the research gap by a small-scale case study (English; German) that pits ISSP translations produced according to the team approach (parallel translations, team discussions) against a simple back translation model, the latter of which takes initial translations from the team approach as its starting point. Based on problematic aspects in the initial translations, we trace how these issues are (not) solved or addressed in the two translation approaches. Besides these qualitative assessments, we will report results from a split-ballot web survey where we compare the two resulting German translations to each other. Data was gathered in November 2020 among 960 German respondents (to take account of the two German versions) and 480 British respondents. The web data analysis will include both quantitative statistics and analyses of open-ended probes.

Español al Otro Lado: How Spanish and English Speakers Respond to a Dual-Language Survey Instrument

Kathryn Bell, *University of Massachusetts Boston Center for Survey Research*

Floyd J. Fowler Jr, *University of Massachusetts Boston Center for Survey Research*

Lee Hargraves, *University of Massachusetts Boston Center for Survey Research*

Carol Cosenza, *University of Massachusetts Boston Center for Survey Research*

Researchers are always looking for ways to encourage participation and increase response rates. Sometimes this includes offering surveys in multiple languages to make surveys more accessible to those whose primary language is not English. This project evaluates how providing a dual-language (Canadian style) English and Spanish survey and materials affects the response rates of Spanish-speaking participants compared with English speakers across different modes of data collection.

In a study of modes of data collection, we assigned a sample of 3200 patients who were discharged from a New England hospital in Fall 2019 to four different groups. Those who provided an email address at the time of their hospital stay were randomly assigned to one of three groups: 1. contacted via mail 2. contacted via email 3. contacted via email then via mail if they did not respond to the email request. Those who did not provide an email address at the time of their hospital stay were placed in a fourth group and were contacted via mail. The entire sample received a dual language (English and Spanish) letter and version of the Hospital Consumer Assessment of Healthcare Provider and Systems (HCAHPS) survey. Those contacted via email were linked to a URL where the survey was available in both languages.

Once data collection was completed, we reviewed response rates based on language in which the survey was completed in comparison to their primary language indicated in their hospital records. Of the 3039 who were eligible, 145 were identified as Spanish speakers in their hospital records. We will report on how many identified as Spanish speaking answered in Spanish versus English, how their response rates compared with those who were not labeled Spanish speakers, and how response rates of Spanish surveys differ across each mode of data collection.

Strategies to Increase Self-Response Among Non-English Speakers in an ABS Mail Push-to-Web Study

Todd Hughes, *UCLA Center for Health Policy Research*

Royce Park, *UCLA Center for Health Policy Research*

Brian Wells, *UCLA Center for Health Policy Research*

Ninez Ponce, *UCLA Center for Health Policy Research*

Margie Engle-Bauer, *SSRS*

Obtaining response from speakers of languages other than English in a self-administered survey is a challenge with few solutions that work effectively in all areas and in all languages. Although multi-mode surveys often rely on interviewer-administered methods to complete surveys with non-English speakers, it is desirable to enhance self-response in other languages to improve cost efficiencies and coverage. A 2018 field experiment using address-based sampling (ABS) and mail push-to-web recruitment for the California Health Interview Survey (CHIS) showed some promise for the use of targeted bilingual mailings in English and Spanish to encourage self-response by web or inbound CATI. This paper continues this research by presenting results from the 2019-2020 statewide CHIS which expanded this approach beyond Spanish to include targeted mailings with messaging in Chinese, Korean, Vietnamese, and Tagalog both on the outside of the envelope and in the materials inside. Addresses selected for the survey were allocated to one of three strata: English-dominant materials, Spanish-dominant materials, or Asian language-dominant materials. Sampling rates by strata were adapted to increase the yield of web and CATI interviews in the non-English languages to improve coverage and reduce nonresponse bias. This paper will analyze these methods, compare the results by language, and consider implications for future work.

Race in America: Discrimination, Equity, and Views of Police

The Burden of Discrimination: Effects of Discrimination Experiences on Quality of Life Indicators

Beatrice Abiero, *Ipsos Public Affairs LLC*

Randall K. Thomas, *Ipsos Public Affairs LLC*

Jasmine Stephens, *Ipsos Public Affairs LLC*

Frances M. Barlas, *Ipsos Public Affairs*

Past research has shown that people of color, women, and people with disabilities are more likely to experience discrimination. This discrimination can take place in a variety of contexts, including applying for a job, buying a house, being admitted to schools, or receiving health care. The connection between the cumulative impact of discrimination on quality of life has been under-explored. This study sought to remedy this deficiency by investigating how various forms of discrimination are related to quality of life. As part of a larger web-based study on discrimination, we used two different sample sources. We had over 12,000 respondents in total with approximately 6,000 from Ipsos' probability-based KnowledgePanel and 6,000 from an opt-in non-probability sample. We randomly assigned participants to consider discrimination experiences across different time periods ("Ever", "Past 5 years", and "Past year") for a fixed set of nine contexts (job-related, education, etc.). We created a discrimination measure formed from the total number of situations in which a specific form of discrimination (due to race-ethnicity, gender, disability, etc.) was experienced. Overall, Black participants experienced more discrimination than Hispanic respondents, and Women experienced more discrimination than Men, consistent with past research. In a subsequent survey section, we asked participants to rate five indicators related to quality of life, including healthcare access, education, and jobs available to them. We compared differences by the form of discrimination respondents experienced as well as demographic characteristics (race, immigrant status, gender etc.). We found that respondents who reported higher incidents of discrimination reported a lower self-rated quality of life, particularly for Black participants and women. Our study goes beyond measuring discrimination by providing a comprehensive overview of the impact of discrimination on key factors, which can inform opportunities for promoting equity.

Crisis, Plunge, and Recovery of Public Confidence in the Police

Jonathan Kelley, *International Survey Center*

Mariah Evans, *International Survey Center*

Charlotte Corday, *International Survey Center*

This paper depicts the trajectory of Americans' confidence in their police across five months before, during, and after the nation was riveted by Minnesota police murdering a Black civilian and the implications thereof for our civic community and our national commitment to equal protection. Data from the International Social Science Survey and the international WVS/EVS surveys show that Americans' confidence in their police is, on average, fairly typical of other advanced societies, coming 30 th out of 110 worldwide. Within the US there was sharp political divergence. Democrats' confidence in the police plunged dramatically around the time of the killing, and then rebounded even more strongly later. But this was only among Democrats. Republicans' views were stable across the period. Regression analysis reveals that that this pattern holds controlling for social and demographic factors, including race. The patterns of change in confidence in the police were replicated for trust in the police. In addition, Blacks and Whites were equally confident in the police across most of the period, except that shortly after the murder Black' confidence plunged briefly and temporarily lower than their White peers' the subsequent rebound in confidence was especially large among Blacks. Further multivariate analysis reveals that the degree of confidence people feel in the government as a whole is by far the largest determinant of how much confidence they feel in the police. Sociodemographic factors other than race play no clear role. Racial prejudice is irrelevant. In short, all this is mainly party politics.

Pre- and Post-Floyd Perceptions of Segregation

Peter Woolley, *Fairleigh Dickinson University*

Daniel Cassino, *Fairleigh Dickinson University*

David Zatz, *Fairleigh Dickinson University*

While many Americans believe systematic segregation is a thing of the past, recent studies show that racial segregation in public school systems and municipalities persists in many states even those thought to be Blue and liberal are among the most segregated. We use original public opinion survey data in one national study and two statewide studies, coupled with census data, to show crucial distinctions between actual municipal and school populations and public perceptions of integration and segregation; and we investigate how response patterns to this sensitive social and political subject changed when protests of the death of George Floyd made race more salient.

Police and Community Relations in Los Angeles, Fall 2020

Alejandra Alarcon, *Center for the Study of Los Angeles*

Brianne Gilbert, *Center for the Study of Los Angeles*

In the wake of nationwide demonstrations for racial justice prompted by the murders of countless Black individuals, Los Angeles, with one of the largest police departments in the nation, has become a focal point for critical discussions around police and community relations. This research focuses on the attitudes and opinions of Los Angeles residents toward the Los Angeles Police Department following the Summer 2020 demonstrations. In a Fall 2020 study of 1,750 adult city of Los Angeles residents, the Center for the Study of Los Angeles at Loyola Marymount University set out to gauge how residents feel that the LAPD is doing with respect to its mission (to safeguard the lives and property of the people the LAPD serves, to reduce the incidence and fear of crime, and to enhance public safety while working with the diverse communities to improve quality of life). Residents serviced by the LAPD were asked several questions on policing in Los Angeles in a mixed-mode (phone, online, and face-to-face) survey administered in four languages (English, Spanish, Mandarin, and Korean), with a minimum of 400 residents surveyed within each of the four LAPD service bureaus. The data are analyzed by key demographics (race/ethnicity, age, gender, and household income), with a focus on the attitudes and opinions of Black residents. This survey is the first of a multi-year study on police and community relations in Los Angeles. As the LAPD works to better community relations, this research provides a point of reference for progress in that work.

Crime, Violence, and Political Participation

Shawna Mullenax, *PSB Research*

Carew Boulding, *University of Colorado Boulder*

Kathryn Schauer, *University of Colorado Boulder*

This paper explores the relationship between crime victimization and political participation, making the argument that the observed (and oft-cited) relationship found in surveys is the result of response bias. We show that people who are more extroverted, efficacious, and opinionated are more likely to respond positively to a question about victimization and they are more likely to participate in politics. The implications of this finding are important and widespread – both for survey researchers interested in sensitive questions and for research on the relationship between crime and political engagement.

The Development of a Hybrid Survey to Assess Police-Community Partnerships. Can this Tool Help Mend the Chasm?

Michelle Kobayashi, *National Research Center, Inc/Polco*
Parker Quinn, *Polco*

Media related to police-community relationships demonstrates an all time low since April of 2020. Survey data can help us to understand this relationship more closely, identify key drivers of perceptions and help us monitor if the efforts made to rebuild trust are working. However, survey topics which generate strong emotions such as police equity may make it difficult to get representative samples of perceptions and priorities from our communities and accurately track these trends over time. The National Research Center at Polco developed the National Police Services Survey™ in 2018 as a tool for local governments across the nation to assess and monitor resident perceptions of safety, interaction with the police, and ratings of service quality and trust. The survey also is used to identify community safety priorities. The survey has been administered to a national panel in 2018 and 2020 and now is being used by individual cities to track their own performance and relationships.

Comparisons of the national 2018 and 2020 surveys will demonstrate the changes in resident perceptions over the time period (pre and post the George Floyd event) and provide insights into non-response bias in police-related surveys. Analysis of the data from 5 communities conducting both probability and non-probability police surveys in Fall 2020/Spring 2021 will augment our knowledge on what residents come to the table to talk about the police and the best methods to calibrate the responses to ensure representation.

More Than a Feeling: Impact of Discrimination Events on Affective Experience

Jasmine Stephens, *Ipsos Public Affairs*
Randall K. Thomas, *Ipsos Public Affairs*
Beatrice Abiero, *Ipsos Public Affairs*
Frances M. Barlas, *Ipsos Public Affairs*

A person's identity is multilayered, encompassing such factors as gender identity, age, race, ethnicity, ability status, and sexual orientation. The intersection of these factors underscores whether one faces discrimination and, if so, the types of discrimination experienced. Discrimination pervades critical aspects of life, including well-being across the life course and one's ability to build social capital. While many studies have investigated the consequences of single forms of discrimination (e.g., race-ethnicity), few have examined the wide range of discrimination that can take place from an intersectional lens. In this study, we investigate the long-term impacts that prior experiences of discrimination across a wide variety of forms have on the general emotional well-being of a person. Using an online survey with two different sample sources—6,000 participants from Ipsos' probability-based KnowledgePanel and 6,000 participants from a non-probability opt-in sample—we asked respondents about their experiences with discrimination due to various factors, such as gender, race-ethnicity, and ability status. In a separate section of the survey, we asked about the frequency with which respondents currently experience a series of diverse emotions (e.g., sadness, stress, anger, fear, happiness, excitement). We created a discrimination measure based on the forms of discrimination reported (e.g., racism, sexism) across the number of contexts in which they occurred (such as in employment or housing access). We examined how the experiences with discrimination related to overall emotional experience. We detail these effects specifically for race-ethnicity and gender. Our study adds to the understanding of the complexity of how discrimination is experienced by utilizing an intersectional framework that accounts for overlapping identities and experiences.

Push-to-Web and Multimode Data Collection

An Experimental Evaluation of Three Approaches for Improving Response to Household Screening Efforts in National Mail/Web Surveys

Brady T. West, *University of Michigan*
James Wagner, *University of Michigan*
Rebecca Gatward, *University of Michigan*
Htay-Wah Saw, *University of Michigan*
Mick Couper, *University of Michigan*

Survey researchers have carefully modified their data collection operations in response to the ongoing COVID-19 pandemic, which has made in-person interviewing difficult. For large national surveys that require household screening to determine survey eligibility, cost-efficient screening methods that do not require in-person visits require additional evaluation and testing. A new national study known as the American Family Health Study (AFHS) recently initiated data collection with a national probability sample, using a sequential mixed-mode mail/web protocol for push-to-web household screening (targeting persons aged 18-49). To better understand optimal approaches for this type of national screening effort, we embedded three randomized experiments in the AFHS data collection. The first tested a form of nonresponse follow-up, where a subsample of active sampled households that had not yet responded to the screening invitations were sent a priority mailing with a \$5 incentive, adding to the \$2 incentive provided for all sampled households in the initial screening invitation. The second tested the use of bilingual respondent materials, where mailed invitations to the screener were provided in both English and Spanish to 50% of households predicted to have a high likelihood of speaking Spanish and 10% of all other households. Finally, after we learned of substantial delays with the delivery of a postcard reminder about the initial mailed screening invitations, we experimented with placing the postcard reminders in an envelope for a subsample of newly released households. Analyzing these approaches, we find that: 1) the nonresponse follow-up was especially effective; 2) consistent with prior work, the bilingual approach increased the proportion of eligible Hispanic respondents identified among completed screeners, especially among high-likelihood Spanish-speaking households; and 3) the envelopes were not effective for the postcard reminder. We will conclude with a discussion of next steps for the second AFHS sample replicate.

Improving Participation in a Mail Push-to-Web Design

Jenny Marlar, *Gallup*

Self-administered mail and mail push-to-web surveys depend on a compelling set of materials to motivate respondents to participate. Participants must see the survey invitation in the mail, open the envelope, read the accompanying materials, and complete the survey request. The invitation components must be carefully designed, and each choice can have a significant effect on response rates. While there is research and guidance on the design of envelopes, letters, and incentives (for example, Dillman, Smyth and Christian, 2014), there are few studies that experimentally test combinations of features and possible interaction effects between design elements.

In this paper, we will share the results of a large mail push-to-web experiment, which utilized an address-based sample sent to more than 100,000 randomly selected U.S. households in April and May 2020. We tested a combination of envelope designs, letter wording, and incentives, using a fully crossed 3x2x2 design. Respondents were randomly assigned to receive one of three envelopes, which varied by size and branding. Two versions of language were used in the survey letters and reminders postcards – one which was generic and provided little detail about the nature of the survey and a second which shared more information about the objectives of the study. Finally, respondents were randomly assigned to receive a \$1 incentive in the first communication only or the first and third communication. We examine response rates for each design feature and the interaction of design features. We also test differences in substantive responses. We briefly discuss the cost implications of each design. We will conclude by sharing our recommendations for practitioners and future experiments.

Please Just Open the Envelope! Reporting on Three Experiments Aimed at Maximizing Response in ABS Surveys

Sarah Glancey, *SSRS*
Robyn Rapoport, *SSRS*
Rob Manley, *SSRS*
Christian Kline, *SSRS*

As general population research shifts from RDD telephone to Address-Based Sampling (ABS) and push-to-web designs, researchers seek the most effective ways of ensuring respondent cooperation. This puts a focus on the initial recruitment materials sent via mail. Specifically, envelope design is critical to engaging respondents since opening the letter is the necessary first step to response.

We will report on three experiments conducted in 2019 and 2020 as part of two surveys. Both surveys aimed to better understand health care experiences of adults in the US. All three experiments tested aspects of envelope design to find ways of maximizing response without additional costs. The first experiment tested three ways to address the envelopes:

- Current Resident
- State Resident
- Household Name or Current Resident (if name was available with the sample record provided)

The second and third experiments tested envelope size and the inclusion of a custom, study-specific logo. Sampled households were randomly assigned to receive both invitation and reminder letters in one of the following four formats:

- 6X9 Envelope with logo
- 6X9 Envelope without logo
- #11 (4 1/2" X 10 3/8") Envelope with logo
- #11 (4 1/2" X 10 3/8") Envelope without logo

Our research found that envelopes addressed to the State Resident and mailings sent in a #11 envelope without a logo yielded the highest completion rates. Additionally, both surveys' samples were stratified to increase the portion of interviews among harder-to-reach populations (e.g., low income, Hispanic, or African American). We will present the experimental results overall as well as by stratum, identifying differential effects of these treatments.

Impact of Initial Mode of Data Collection on Attrition in a Longitudinal Study of Young Adults

Ting Yan, *Westat*
Jonathan Wivagg, *Westat*
Charles Carusi, *Westat*
William Young, *Rutgers University*
Cris Delnevo, *Rutgers University*
Daniel A. Gundersen, *Dana-Farber Cancer Institute*

Young adults are generally hard to survey, presenting multiple forms of difficulty. They are hard to locate and contact due to high mobility. They are hard to persuade and exhibit high levels of resistance to survey participation. As a result, they pose a greater challenge for longitudinal surveys. This paper attempts to understand the role of mode of data collection in young adults' decisions to stay in a longitudinal panel. We draw on data from the National Young Adult Health Study (NYAHS). NYAHS is a longitudinal study (three annual waves and 2 brief between-wave follow-up surveys) of adults aged 18-34 initially recruited in 2019 through RDD sampling of cell phone numbers nationwide. Once screened in, two-thirds of young adults were randomly assigned to continue with Wave 1 main interview over the phone and one-third was asked to switch to completing it online. The web mode yielded a 8% response rate (n=131) whereas the phone condition in the low 20% (n=700). We propose to examine whether or not this initial assignment of mode leads to differential propensities to respond to later survey requests.

We will first compare response rates to Wave 1 follow-up interview and Wave 2 main interview by the initial interview mode. As a random half of the young adults in the phone condition were asked to complete Wave 1 follow-up interview online, we will examine young adults' propensities to participate in Wave 2 main interview as a function of the initial mode and the number of times invited to the web mode. Lastly, we will investigate how predictors of attrition (e.g., item nonresponse in the initial interview) interact with the initial mode of data collection in affecting young adults' likelihood to participate in later interviews. The findings will have important practical implications on how to survey young adults.

A Population-Based Health Survey Built for West Virginia

Summer Hartley, *West Virginia University, Office of Health Affairs*

Emily Robinson, *West Virginia University, Office of Health Affairs*

Natalie Ezzie, *West Virginia University, Office of Health Affairs*

Daniel Christy, *West Virginia University, Office of Health Affairs*

Summer Hartley, *West Virginia University, Office of Health Affairs*

Christina Mullins, *West Virginia Department of Health and Human Resources, Bureau for Behavioral Health*

Cynthia Beane, *West Virginia Department of Health and Human Resources, Bureau for Medical Services*

West Virginia (WV) is experiencing more than 20 years of a sustained epidemic of fatal and nonfatal drug addiction and misuse, and leads the U.S. in several negative mental health indicators. Accurate, population-level, mental health estimates have been challenging to obtain in WV at the local level, and these local-level estimates are critical to informing the state response to this epidemic. Data-driven strategic planning is especially important now, given the increased prevalence of substance use and mental health issues that are contributing to rises in adverse health and economic population outcomes throughout the State. Therefore, the WV Department of Health and Human Resources is working with West Virginia University to establish a population-based, statewide health survey that will collect representative data at the county level. This is the first known attempt at this type of data collection in the State, and nationally with an almost completely rural population. This survey will utilize a multi-modal approach, taking into consideration mode sequencing and frequency of follow-up to produce the highest response rates possible, while maintaining cost effectiveness. The sequencing of the survey modes will be informed by the distribution of WV characteristics related to internet access, income levels and literacy. The sampling approach will be based on the survey modes and distribution of Medicaid recipients throughout the State, as informed by Medicaid Claims data. The survey will also perform a multi-variate test on incentive structures to determine which incentive is more effective at increasing survey response rates among West Virginians.

The survey promises to provide many benefits to WV, such as representative, county-level data that will inform resource allocation and strategic planning at the state and local levels via a robust data dissemination plan. The survey will also address data gaps vital to program/policy decisions.

"Translating the Sender" in Spanish-Language Web Surveys: A Contact Experiment

Katy A. Lindquist, *University of Maryland*

Marcus A. Boyd, *University of Maryland*

Amira Giadala, *University of Maryland*

Barnett S. Koven, *University of Maryland*

The issue of low web survey response rates has led to an increased interest in a variety of aspects of survey and invitation design and numerous experiments to ascertain how factors like subject lines, the sender, URL placement, invitation length, and invitation content impact survey completion rates (Fan and Yan, 2009). However, few existing studies have examined survey invitation design outside of English-language contexts. This study is a survey contact experiment among businesses in Panama and Costa Rica that explores survey invitation design in a particular Central American, Spanish-speaking context. Drawing on literature about the importance of (perceived) affinity in web-communications and surveys (Guéguen, Jacob, and Morineau, 2010), we examine whether a more complete linguistic and iconic translation of invitation into Spanish impacts whether recipients are willing to participate in an online

survey. Recipients are randomly assigned to receive one of two different email survey invitations: an incomplete translation version or complete translation version. Both will be in Spanish and contain substantively the same information. However, for the complete translation version of the email invitation, University of is translated as Universidad de anywhere it appears, including in subject, headers, signatures, footers, and logos as well as the email body text. The incomplete translation version leaves the institution untranslated in all places – body of the email, subject, logos, etc. We also vary the sender name, with the incomplete translation email coming from an Anglophone name and the complete one from a Hispanophone name. By more and less fully translating these features of the invitation, we will ascertain whether creating a greater linguistic affinity between the (institutional and personal) sender and Spanish-speaking recipients has an impact on survey participation and completion rates. This will help researchers better understand how to consider survey invitation design elements in non-English contexts.

Lessons Learned From a Multi-Wave Sequential Mixed-Mode Survey of Apprenticeship Program Participants

Jessica Pleasants, *Abt Associates*

Julie Pacer, *Abt Associates*

When designing a multi-stage protocol that includes a follow-up survey, studies should take a number of steps to ensure high response rates at follow-up. For example, introducing the follow-up survey to participants at program enrollment and collecting participant contact information both prepare the participant and position the researcher for effective future engagement. Attempting to reach program participants who were not aware of a future survey or did not provide adequate contact information can lead to low response rates.

The American Apprenticeship Initiative (AAI) Evaluation sponsored by the U.S. Department of Labor and led by Abt Associates surveyed U.S. apprentices about their apprenticeship program experiences. The study collected data via a web survey with a telephone nonresponse follow-up component. This presentation describes strategies to increase response in a follow-up survey that was not part of the original AAI study design. The evaluation team implemented an incentive experiment to determine whether a higher incentive amount would result in a higher response rate. In addition, the incentive experiment included a two-tiered option in which the incentive amount decreased after an initial early response period, to determine whether a higher early response incentive would result in faster survey completion. The AAI survey was also administered in two waves, which allowed for changes in the data collection protocol such as including the apprenticeship sponsor name and using targeted outgoing phone numbers by displaying a local area code in phone data collection. We compare response rates across the incentive experiment groups and data collection waves. We also summarize our recommended strategies for conducting an unanticipated follow-up survey with a known population.

Evaluating Data Quality and Identifying Falsification

No More Waiting by the Phone...Results From the Integration of Multiple Technologies for Large-Scale In-Person Survey Case Validation

Ryan Hubbard, *Westat*

Large-scale in-person surveys rely on multiple forms of validation to identify falsification and ensure data quality. Traditionally, national surveys where large-scale direct observation was not possible, telephone validations where respondents verify the date, time, and content of an interview became standard. Additional forms of verification such as CARI and more recently GPS systems now offer the ability to validate a larger number of cases, more rapidly, at a lower cost. However, multiple validation tools present challenges to staffing, training, and managing a validation team.

The Medical Expenditure Panel Survey (MEPS) replaced a per-interviewer-percentage based telephone validation operation with a system that integrates GPS, CARI, phone and mail into a set of cascading stages. If a case cannot be validated at a given stage, either because data is not available or review did not produce positive validation, it moves to the next stage. A management system determines the proper stage for a case and assigns it to a validator. The process begins with the most economical approach, working toward more costly validation efforts. The first stage, GPS, uses automated analysis followed by manual GPS review if necessary. Cases not validated through this approach are reviewed in CARI, then telephone and finally mail.

After several months of use, results indicate a much larger volume of cases may be validated, from less than 25% through phone validations to over 90% through the new integrated system. Thousands of cases were evaluated using automated GPS analysis; however, CARI proved to encompass the bulk of manual validator review. This is a detailed review of the validation methods employed on MEPS, the proportion and outcome of cases validated using each method, the overall contribution to data quality for the integrated system compared to telephone validation, and areas for improvement in the validation technologies and their integrated management.

Improving Data Quality: Does Data Cleaning Reduce or Increase Bias in Sub-groups?

Randall K. Thomas, *Ipsos Public Affairs*

Frances M. Barlas, *Ipsos Public Affairs*

Megan Hendrich, *Ipsos Public Affairs*

To improve accuracy of results, many researchers exclude cases from analyses when participants have demonstrated sub-optimal behaviors, such as speeding or straightlining. Increasingly, researchers have employed 'red herring' or 'trap' questions in their surveys to catch inattentive participants. We have seen some go so far as to use aggressive cleaning criteria that deletes 20 percent of participants. This raises questions about the validity of the survey results, but also has cost implications as replacement sample is often required, and those being cleaned are often less likely to respond in the first place (e.g., younger, people of color, etc.). How much data cleaning is too much? For this research, we analyzed data from two different web-based surveys with items for which external benchmarks existed from high quality sample surveys that would allow for bias estimates. Study 1 included a sample of 1,297 participants from Ipsos' probability-based KnowledgePanel and 2,563 participants from non-probability online samples. This survey included 24 questions with available external benchmarks to investigate bias. Study 2, had over 3,300 participants from KnowledgePanel and 5,800 from non-probability online samples. We explored the effects of data cleaning at different rates (from 2.5% to 50%) for effects on average error from benchmarks, both overall and for specific subgroups (e.g., male or female; Black, Hispanic, or White). While we found that average error for probability-based sample was lower than opt-in sample, we also found that that cleaning does not reduce error at an overall level, nor does it increase error. At the subgroup level, there is some evidence that error is somewhat higher for some groups. Looking at covariance measures, we found that excessive cleaning protocols may actually increase bias, especially for subgroups. We generally recommend caution for any data cleaning and instead recommend increasing attention to designing shorter and easier-to-complete surveys.

Identifying Interviewer Falsification Using Speech Recognition: A Field Study

Hanyu Sun, *Westat*

Ting Yan, *Westat*

Gonzalo Rivero, *Westat*

Survey management staff have long used Computer Audio-Recorded Interviewing (CARI) to assess interviewer performance, validate interviews, and to evaluate the performance of survey questions. The success of such an evaluation, however, often depends on labor-intensive coding in a timely manner. The coders first need to listen to the interactions between the interviewer and the respondent, and then provide their assessment of how the interviewer performs based on some criteria such as whether or not the interviewer falsified the case and whether or not the interviewer followed the standardized interviewing techniques. Due to resource constraints, typically only a small number of items in the questionnaire or a small portion of the interview will be listened to and coded based on pre-specified criteria. At Westat, we developed an assessment tool that automates the coding and the evaluation process. The tool first transcribes CARI audio recordings into text and then measures the distance between the transcript and the questionnaire. The tool also detects the number of conversational partners in the recording. Both this auto-detection of the number of speakers and the distance measures can be used to detect interviewer falsification. We conducted a proof of concept study with recorded lab interviews in 2019 and found that the speech recognition approach could detect the number of speakers correctly for most of the lab interviews. We also found higher agreement between the transcript and the speech when an interviewer reads the question verbatim. This paper extends the 2019 study in two important ways. First, we will replicate the 2019 analyses on recorded field interviews. Second, we will evaluate the effectiveness of the auto-detection of the number of speakers and the distance measures in detecting interviewer falsification in a field in-person study. We will also evaluate the performance of the tool at identifying poor interviewer behaviors.

Respondent Burden Over Time: A Three-Stage Model

Andrew Hupp, *University of Michigan*

Kallan M. Larsen, *University of Michigan*

Frederick G. Conrad, *University of Michigan*

Michael F. Schober, *The New School*

While survey researchers are in agreement that respondent burden should be reduced wherever possible without reducing data quality, conceptualizations of respondent burden itself remain relatively imprecise. Burden is measured using anything from objective indicators like response rates and questionnaire length to the rare self-reported subjective indicator. Interview duration remains the most common measure of respondent burden, but asynchronous survey modes, such as texting, where interview duration is difficult to define, are increasingly common. In this paper, we synthesize existing burden literature to form an actionable model of respondent burden that can be applied across existing modes, regardless of synchrony or technological demands. This model separates respondent burden into three categories: anticipated burden, in-the-moment burden, and remembered burden. Each stage of burden has corresponding indicators, including but not limited to response rates (anticipated burden), data quality (in-the-moment burden), and attrition in longitudinal studies (remembered burden). Using findings from recent studies that implement newer survey modes (web, texting, and video), we present examples of how to use this new model to more precisely identify sources of burden and thus strategies to potentially reduce burden.

Evaluating Machine Learning Algorithms to Detect Interviewer Falsification

Silvia Schwanhäuser, *Institute for Employment Research (IAB)*

Joseph W. Sakshaug, *University of Mannheim; University of Munich (LMU); Institute for Employment Research (IAB)*

Yuliya Kosyakova, *University of Bamberg; University of Mannheim; Institute for Employment Research (IAB)*

Natalja Menold, *University of Dresden*

Peter Winker, *University of Giessen*

Interviewers play a vital role for the quality of survey data, as they directly influence response rates and are responsible for appropriately administering the questionnaire. At the same time, interviewers may be enticed to intentionally deviate from the prescribed interviewing guidelines or even fabricate entire interviews. Different studies have discussed various possibilities to prevent and detect such fraudulent interviewer behavior. However, the proposed controlling procedures are often time consuming and their implementation is cumbersome and costly.

One understudied possibility to simplify and automate the controlling process is to use supervised machine learning algorithms. Even though some studies propose the use of unsupervised algorithms like cluster analysis or principal component analysis, there is hardly any literature on otherwise widespread methods like neural networks, support vector machines, decision trees, or naïve Bayes. This is mainly driven by the lack of appropriate test and training data, including sufficient numbers of falsifiers and falsified interviews to evaluate the respective algorithms.

Using data from a German experimental study, including an equal share of falsified and real interviews as well as real-world data from a German panel survey with fraudulent interviews in different waves, we address the question: How well do supervised machine learning algorithms discriminate between real and falsified data? To do this, we evaluate the performance of different algorithms under various scenarios. By utilizing different data sources and working with different subsets for training and testing the algorithms within and across datasets, we provide additional evidence regarding the external validity of the results. In addition, the setting allows us to draw conclusions on the different strategies and behaviors of falsifying interviewers.

Finding Fraud in Public Polls: Employing Semantic Network-Based Methods for Identifying Fraud in Online Sampling

Leib Litman, *Touro College*

Peter Kowalski, *Versta Research*

Cheskie Rosenzweig, *CloudResearch*

Joe Hopper, *Versta Research*

Jonathan Robinson, *CloudResearch*

Adam Dietrich, *CloudResearch*

Online opt-in panels are increasingly being used in political polls and public opinion surveys, with 80% of current public opinion polls using at least some online respondents. However, online data collection suffers from numerous threats to data quality, including acquiescence bias, satisficing, and random responding. This increases the error of point estimates and creates illusory correlations. Researchers use a variety of approaches to mitigate against online data quality threats, including attention checks and other in-survey measures. However, these approaches suffer from multiple methodological limitations, most notably the lack of validation. As a result, typical solutions are either too stringent and prevent good respondents from participating (false positives), or too weak and allow bad participants into the survey (false negatives). Here, we present several solutions to mitigate against both random responses and satisficing. We use a semantic network model approach to create sets of words with a quantifiable associative similarity. Weights are assigned to word-pairs based on the analysis of English language corpora, and specific objective difficulty thresholds are assigned by varying the weights between targets and response options. These stimuli are then used as measures of attention, engagement, comprehension, and inauthentic behavior in pre-survey screens. This approach was further improved upon by adding additional quality measures that check for acquiescence bias. A library of validated stimuli was developed and employed to examine the quality of respondents across several large online surveys. Data across these studies suggest that pre-survey screens prevented over 70% of problematic respondents from entering surveys, had a low false positive rate, and performed nearly as well as postsurvey quality review. Overall, we propose that this approach can be used at scale to systematically prevent problematic respondents from entering online surveys across the vast network of opt-in panels that supplies respondents for online polls.

Evaluation of Data Quality Indicators in Online Panel Providers Used to Identify and Remove Inauthentic Survey Response

Emily Geisen, *Qualtrics*

Brad Smith, *Qualtrics*

William Belden, *Qualtrics*

Research suggests that nonprobability samples produce survey estimates that are more biased than probability samples. Yet, market researchers and academics are increasingly using online nonprobability panel providers due to their low cost, speed, and convenience. As a result, researchers use weighting and modeling approaches to reduce the bias and increase representativeness of online panels. While efforts such as sample-matching and raking have been shown to improve the quality of nonprobability panels, they still do not perform as well as probability samples on average (Dutwin & Buskirk, 2017). One possible reason online panels perform poorly may be because they are more likely to include “professional” survey takers who are less motivated to provide accurate responses and primarily complete surveys to receive incentives. If online panel respondents who falsify responses, satisfice, or otherwise answer in an inauthentic way can be identified and removed (and ideally replaced with motivated respondents), it could – along with weighting techniques – help to reduce bias in surveys using online panel providers.

To explore this, we surveyed over 2,900 respondents from four online panel providers and analyzed a variety of data quality indicators: responses to honeypot or invisible questions (to identify bots), contradictory or inconsistent responses across questions, straightlining, trap questions, nonsense or illogical open-ended responses, item nonresponse, speeding, failing attention checks, falsifying responses, or otherwise providing an invalid response (e.g., answer of 5,000 for age), and comparisons of data to benchmarks.

We will present an evaluation of these data quality indicators by panel provider, determine which quality indicators most consistently and accurately predict respondents who provide inauthentic responses, and determine whether removing these respondents can improve key survey estimates relative to benchmarks. The implications of the findings on conducting survey research with online panel providers will also be discussed.

Measurement, Formatting, and Design

Making Cognitive Reflection Test Robust in Online Surveys

Wenhuan (Edwin) Wang, *University of Michigan*

Josh Pasek, *University of Michigan*

Jan Van den Bulck, *University of Michigan*

The Cognitive Reflection Test (CRT; Frederick, 2005) is a popular measure of cognitive effort that relates to outcomes ranging from information seeking behaviors to the acceptance of inaccurate claims. The measure employs a few questions that ask respondents to think through problems where the intuitive answer differs from an easily derived correct answer. Despite the utility of measuring CRT, scholars have raised concerns that the logic problems used as part of the measure suffer from challenges including familiarity issues (Haigh, 2016; Stieger; Reips, 2016), floor effects (Thomson; Oppenheimer, 2016), and an over-reliance on arithmetic modes of thinking (Sinayev; Peters, 2015).

This paper presents results from three survey experiments comparing traditional CRT items with alternative items in both broad national and student samples. The first study (using respondents from a national online non probability panel vendor) examines the influence of potential familiarity by randomizing respondents between conditions using traditional measures and items that use the same arithmetic concepts, but different examples. The second study (in a student participant pool) examined the extent to which non-arithmetic CRT-style questions yielded similar relations with criterion variables. And the third study (drawn from another nonprobability online vendor) examines the susceptibility of both original and alternative measures to survey context and survey placement, which is particularly important given the cognitive complexity of the items. In a series of assessments of criterion validity for various items, we find mixed support for the contention that CRT measures suffer from familiarity issues, but greater evidence that respondent attention levels and survey fatigue can influence these responses (cf. Chandler, Mueller, Paolacci, 2014; Meyer, Zhou, Shane, 2018). We discuss how these results should be understood and what they imply for future use of CRT measures.

Designing Dual Question Grids for Mail Surveys: An Experiment

Jolene D. Smyth, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

Surveyors sometimes place two questions side-by-side in a dual-question grid format in mail surveys to avoid redundancy, save space, and accommodate skip patterns. For example, one might ask who in the sampled household did certain household tasks before and during the coronavirus pandemic within a single grid, using a vertical line between the two questions to visually differentiate them. This design creates complex visual groupings of information. As a result, item nonresponse rates to each question and to sub-items within each question are often high, especially on the second (or right-hand-side) question (Couper et al. 2013; Dillman et al. 2014). The navigation of dual-question grids is further complicated when they contain skip patterns where whether respondents should answer the second question depends on their answer to the first. Despite their complexity, reduced data quality, and advice to avoid them, dual-question grids continue to be used. Given their continued use, it is desirable to test possible design improvements that can help respondents successfully navigate these grids.

In this paper, we report the results of a dual-question grid visual design experiment in a mail survey of Nebraskans about COVID-19 and race relations fielded September through December 2020 (2400, 24.8% as of November 2020). We compare traditionally designed dual-sided grids that utilize vertical lines to separate questions versus redesigned grids that use combinations of graphical (e.g., vertical lines), verbal (e.g., "If yes"), and symbolic (e.g., arrows) cues designed to help respondents follow complex navigation in these grids. Across five dual-sided grids, we compare item nonresponse rates (including errors of omission), errors of commission (answering items that do not apply), predictors of question and sub-item-level item nonresponse, and response distributions. Results will help inform best practices for designing dual-sided grids in mail surveys to maximize data quality.

Playing Hardball, Softball or Snowball: A Study on the Use of Warnings in a Web Survey

Sebastian Kocar, *Australian National University*

Web surveys have become an increasingly important mode to collect survey data in the era of decreasing response rates and increasing costs. While they often do not have a genuine link to probabilistic sampling, they are a useful tool in push-to-web approaches or when surveying subpopulations for which there is a full sampling frame with available email addresses of potential respondents. In addition to their relatively low cost, web questionnaires can be interactive and can react to what respondents do. One particular approach in computerized web surveys is question warnings, which can help decrease item nonresponse. They can generally be classified into soft warnings, i.e. respondent is asked to respond to a question but can still continue with the survey without providing an answer, and hard warnings, i.e. respondent is asked to respond to an item and cannot continue with the survey without answering to the item first - also known as forced answering.

The literature on the use of warnings, as well as the recommendations in online data collection tools like Qualtrics or OneClick Survey, present limited and often mixed evidence on what warnings are optimal from the survey error perspective. To address this problem, we ran a survey experiment using three distinctive approaches; we programmed three online questionnaires with exactly the same content but with a different level of survey warnings in case of item nonresponse: hard, soft and no warnings. After the introduction page, we randomly assigned our respondents, academics employed at higher education institutions, to different versions of the questionnaire. The results showed that hard warnings decreased item nonresponse while not increasing breakoff rates, which was not consistent with most of the literature on forced answering. In this presentation, I will discuss the possible reasons for that, such as specifics of the studied population.

“Do You Have Anything Else to Add?” Exploring the Value of Final Catch-all Questions in Qualitative In-Depth Interviews

Casey Tesfaye, *Research Support Services, Inc.*

Jessie Engel, *Research Support Services, Inc.*

Researchers make many careful considerations when designing questions for qualitative interviews. While questions composing qualitative interview guides are written with study goals and intents in mind, the interview guides also commonly include a general, “catch-all” question. These are questions such as “Do you have anything else to add before we end our interview today?” These questions provide a way to close the interview and ask respondents to share any additional thoughts that may not have already been fully addressed. Additionally, there are often unanticipated themes that arise during the research conversation, and a final, less structured question affords an opportunity to follow-up should time allow.

For this analysis, we first isolate the set of final questions and associated response dyads from the transcripts of a set of recent qualitative interview studies and analyze both the question itself and the content, insightfulness and tone of the responses gathered, in order to more deeply examine their usefulness and value. We will also examine the level of specificity of the question asked, any observable variation between interviewers introduced in the delivery of the questions, and the word counts from the interviewers and respondents after the question has been asked. Ultimately, we aim to use these findings to develop guidelines for creating follow-up questions and closing processes that are maximally useful. These guidelines will address the specific phrasing of the questions as well as their value within the context of the interview and subsequent analysis.

Capturing the 2020 Census User Experience Using Sentiment and Thematic Analysis of Open Text

Shelley Feuer, *U.S. Census Bureau*

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

Elizabeth Nichols, *U.S. Census Bureau*

In 2020, for the first time in U.S. history, the public had the option to respond to the decennial Census online. The U.S. Census Bureau then re-contacted a sample of census respondents and asked them to complete a short online survey in order to measure satisfaction and user experience. The sample represented a variety of respondent profiles including early and late Census responders and responders who did and who did not use their Census ID. We gained feedback on which areas of the questionnaire were problematic for users and got an overall satisfaction rating. Using an open text question at the end of the survey, we also provided an opportunity for respondents to share more about their experience filling out the 2020 Census.

We conducted manual coding of these text responses, in addition to using Qualtrics Text iQ tool for sentiment analysis, to gain a more nuanced understanding of the respondents' thoughts and concerns. Our thematic analysis combined inductive and deductive approaches to coding, with themes that cover usability issues observed throughout the decade (e.g. separating race and Hispanic origin), feedback from other 2020 Census operations (e.g. duplicate enumeration; vacancies), the 2020 Census (e.g. citizenship question; ending count early), and the 2020 climate (e.g. COVID-19; politics). This talk will present the results from these qualitative analyses, including their relationship to quantitative measures of satisfaction and whether the type and valence of comments are related to other respondent characteristics, such as when they completed their Census. We will also address the benefits and drawbacks of Qualtrics' overall and topic-level sentiment analysis compared to manual coding of open text valence and theme. Together, the findings provide a much more comprehensive picture of the 2020 Census user experience, as well as a clearer path to improving operations for 2030.

Understanding How Respondents Interpret a Demystified NPS Scale

Phyllis Mancino, *IBM*

Annette Tassone, *IBM*

NPS is one of the most widely used customer experience metrics among small businesses and large enterprises. The appeal of NPS is that it is a single number, and it has successfully driven increased focus on customer and employee experience. This focus has become so incessant that companies set strategy, evaluate performance, and determine salaries based on NPS. This focus on the "score" by some customer experience professionals and stakeholders has propagated a movement toward manipulating the scale to improve the ratings.

Creative methods to artificially inflate the score are increasingly being used. Recent consumer surveys show numbers 10 and 9 accompanied with green smiling faces, 8 and 7 with yellow disappointed faces and 0 to 6 with red sad faces. This manipulation turns the rating into a 3-point scale. Even marketing and top companies like NiceReply, British Airways, and Salesforce have used color-coded NPS rating scales. Repeatedly, clients ask for them, and we field comments such as, "But the ratings are wrong; we have to tell respondents what the ratings mean." To understand the impact of manipulated NPS scales, we used a standard NPS and color-treated NPS in a large enterprise survey to measure the sentiment of a workplace tool. Findings revealed that the color-treated scale did indeed lead to significantly higher ratings when compared to the standard NPS question. Further, through in-progress cognitive interviews, we explore how the standard NPS scale is understood, how the manipulated NPS scale is interpreted, and how the scale manipulation affects ratings.

New Ideas for Testing a Psychometric Instrument Using Mixed Methods

Florian Berens, *University of Goettingen*

Kelly Findley, *University of Illinois Urbana-Champaign*

Developing psychometric instruments, two core activities of testing the instrument have so far played a major role: (Think Aloud) Cognitive Interviews and Explorative or Confirmative Factor Analysis. Cognitive Interviews are performed to check the wording of the individual items for comprehensibility and formulation validity. The factor analyses quantitatively test the reliability of the theoretically designed constructs. From the authors' point of view, this course of testing lacks an intermediate stage that uses mixed method approaches and that combines a more detailed examination of the individual items with a review of the validity of the constructs.

The idea of the developed approach is to substitute the formation of factors via correlation measures, which is common in quantitative factor analysis, by human clustering. Interviewees cluster the developed items into factors based on their personal knowledge and their everyday understanding of the questionnaire. The required think aloud and the result of the clustering can be used as different perspectives to evaluate and improve item formulations and intended constructs. The amount of preliminary information the respondents receive is varied in three possible steps.

This work reports on the authors' experiences with different variants of the approach. They were tested on an instrument development for a survey on beliefs about statistics. All variants were tested at two universities in different countries on different groups of respondents. The result is a proposed procedure with which an instrument draft can be substantially improved within two cycles. In the first cycle, items and theoretical constructs are compared individually to check their fit. In the second cycle, the constructs as such are recreated by the respondents.

Using Social Media and Digital Technologies for Sampling and Data Collection

Using Social Media to Recruit Rare Populations for Web Survey Research

Zaza Zindel, *Bielefeld University*

Simon Kühne, *Bielefeld University*

In many countries and contexts, survey researchers are facing decreasing response rates and increasing survey costs. Data collection is even more challenging and expensive when rare or hard-to-reach populations are to be sampled and surveyed. Alternative sampling and recruiting approaches are usually needed in these cases, including non-probability and online convenience sampling. A rather novel approach to recruiting rare populations for online and mobile survey participation uses advertisements on social media. This presentation discusses the advantages and disadvantages when using social media as a recruiting tool for reaching rare populations. By comparing a social media sample survey with a simultaneously collected, probability-based, face-to-face sample, we assess the potentials, appropriateness, but also limitations of social media recruitment compared to traditional sampling approaches. In addition, we will share insights on planning and conducting an ad campaign for recruiting rare populations for web survey research. Finally, using several best-practice examples, we will discuss practical details and potential hurdles in advertisement campaign preparation, during fieldwork, and subsequent data processing.

To sum up, our presentation covers both the conceptual and the practical aspects of social media-based survey recruitment, and thus, may serve as a starting point for survey researchers and practitioners involved in non-probability sampling surveying rare populations.

Exploring the Effects of Advanced Social Media Targeting Methodology on Recruitment of Hard-to-Reach Audiences to Non-Probability Surveys

Amelia Burke-Garcia, *NORC at the University of Chicago*

Ipek Bilgen, *NORC at the University of Chicago*

Justine Bulgar-Medina, *NORC at the University of Chicago*

Robert Henne, *NORC at the University of Chicago*

Dasha Afanaseva, *NORC at the University of Chicago*

One of the major benefits of social media ad-based survey recruitment is the existence of various types of data to target ads to users of these platforms. There are a couple different ways to target social media ads – first, by using the basic data, e.g., demographic and geographic, that social media platforms currently provide and second, by using enhanced data that can be embedded within the social media platforms supplied by third party providers based on external data sources, e.g., historical purchase data. We will examine whether and how much this latter category of data will enhance ad based social media recruitment capabilities to reach niche and hard-to-reach audiences.

To investigate the targeting efficiency, quality, and cost differences among these two approaches that can be used to target audiences within social media platforms, NORC piloted an internal research study in 2020. Web survey was constructed using existing items from national surveys on individual's health and online habits as well as new items related to life changes during the pandemic. Two main audience groups that are generally hard to recruit through probability based studies were targeted – young adults, ages 18-24, and people with low education (defined as anyone who has completed high school as the highest level of education or lower). Five sets of tailored ads with unique URLs that linked to a web-based survey were designed and launched via Facebook and Instagram. Two sets used basic targeting to recruit the sample and the other three used the enhanced targeting to do so. This paper will present findings from our examination of the differences between the samples obtained from basic and advanced targeting on the dimensions of recruited sample composition, survey estimates, and recruitment costs.

Using Paid Advertising on Facebook to Recruit for Fast Turn-Around Surveys: Results From a 5-Topic Survey Participation Experiment

Samantha Chiu, *University of Maryland, Joint Program in Survey Methodology*

Leanna Panasethaned Moron, *University of Maryland, Joint Program in Survey Methodology*

Frauke Kreuter, *University of Maryland, Joint Program in Survey Methodology*

Stanley Presser, *University of Maryland, Joint Program in Survey Methodology*

Xin (Rosalyann) Yang, *University of Maryland, Joint Program in Survey Methodology*

Mobile data collection and online surveys, linked with potential sensor and other meta-data, can provide rapid information sharing to tackle crisis events. Facebook (FB) offers a platform where surveys and big data can be collected via a mobile-devices, through apps, and their online platform. With over 2.7 billion monthly active users, it is the biggest social network worldwide. FB Advertisements for topic-specific surveys are growing in use due to the cost-effectiveness and rapid deployment. However, researchers have little documentation on the recruited population and whether selection bias is present.

Building on leverage-salience theory and tailored invitation letters effects on participation, our research explores the effect of recruitment images using FB Advertisements. Prior experimental designs find context effects due to image associations in survey measurement. Utilizing images in web surveys are found to influence recruitment, participation, and non-response. This can change the user's experience and impact total survey error.

Our research experimented with five images: COVID-19, mental health, finance, data privacy, and a neutral (control) image. We found that COVID-19 images recruited the most participants, but that the neutral image produced the least biased results. The mental health image recruited fewer participants than the COVID-19 or neutral image. Less salient topics such as finance and data privacy recruited the smallest number of participants. Our findings suggest that the image used for recruitment does recruit different populations, yield different samples, and may lead to bias results. Our results are benchmarked against the probability-based University of Southern California's Understanding America Survey in May and June 2020.

Mall-Intercepts for the 21st Century: Geofence Based Survey Engagement

Valerie Lykes, *J.D. Power*

Our traditional survey invitation method for automotive research is a mail-to-online approach. This year, we identified through geofencing that someone visited a dealership, and then invited them to participate in a short survey through digital recruitment. From two studies (N = 500 for first, N=5,000 for second), we will be comparing demographics between this approach and mail-to-online, as well as looking at whether there were any index score differences between these two sources. We hypothesize that attitudes will be similar between the two methods but the geofenced sample will be more representative by age group.

Informed Consent in Social Media Surveys

Adrienne Bradford, *University of Maryland, Joint Program in Survey Methodology*

Samantha Chiu, *University of Maryland, Joint Program in Survey Methodology*

Frauke Kreuter, *University of Maryland, Joint Program in Survey Methodology*

Stanley Presser, *University of Maryland, Joint Program in Survey Methodology*

IRB requirements for informed consent for surveys often result in long and cumbersome statements that may discourage participation by potential respondents. In our study, we tested a more streamlined consent statement by embedding an experiment in the University of Maryland/Facebook International COVID-19 Symptoms Survey that compared a two-page consent process to obtain both informed consent and consent for data sharing with a simpler one-page combined consent statement. We hypothesized that more respondents in the short version would consent and complete the survey. We also examine whether there are differences in composition of the sample or in answers to key survey questions between the two conditions. Our results suggest that researchers should simplify consent statements, especially for social media surveys that will likely be taken via smartphone.

Data Collection of Subpopulations and Institutions

Collecting a Maternal and Child Health Survey in the U.S. Jurisdictions

Brian Kirchhoff, *NORC at the University of Chicago*
Anna Wiencrot, *NORC at the University of Chicago*
Clare Davidson, *NORC at the University of Chicago*
Ben Skalland, *NORC at the University of Chicago*
Erik Amonson, *NORC at the University of Chicago*
Peyton Holleran, *NORC at the University of Chicago*
Michael Kogan, *HRSA / Maternal and Child Health Bureau*

High-quality maternal and child health (MCH) data are required to inform efforts to improve the health and well-being of mothers, children, and families in the 59 U.S. states and jurisdictions. However, the eight jurisdictions (Puerto Rico, the U.S. Virgin Islands, Palau, Marshall Islands, Guam, American Samoa, Northern Mariana Islands, and Federated States of Micronesia) have had limited MCH data compared to the states. The goal of the MCH Jurisdictional Survey was to meet the jurisdictions' need for MCH data.

Following a comprehensive assessment of the jurisdictions' data needs, the National Survey of Children's Health (NSCH) was adapted for use in the jurisdictions, adding maternal health and jurisdiction-specific questions. The questionnaire was pretested in summer 2018. The results of this pretest, plus collaboration with the jurisdictional MCH leads and experts in data collection in these locations informed the development of a field plan for standardized, in-person data collection appropriate for use in each unique location. Data collection protocols, sample selection procedures, and trainings were standardized across all eight jurisdictions. Due to unique challenges, such as a lack of addresses, a number of novel approaches were developed to ensure high quality data were collected. The survey was administered in English and in common local languages. Data collection was conducted once in each jurisdiction between May 2019 and February 2020. Other challenges encountered and overcome included lack of examples from past data collections, lack of current population estimates, and physical challenges inherent to each jurisdictions' environment. Over 200 surveys were completed in each location. These data are being used by the Maternal and Child Health Bureau (MCHB) and the jurisdictions to improve planning and reporting on MCH priorities. Planning is now underway for the next round of data collection, incorporating lessons learned as well as the impact of the pandemic.

Sampling and Post-Hoc Adjustment Strategies for a Multi-Mode Study Targeting Black U.S. Adults

Jazmyne Sutton, *SSRS*
Liz Hamel, *Kaiser Family Foundation*
Eran Ben-Porath, *SSRS*

A hybrid sample design utilizing both traditional random digit dialing (RDD) sample as well as a probability-based web-panel may provide an effective method for reaching a representative sample and provide a means to oversample minority groups such as Black adults. However, using a combination of interviewing modes and sample types raises concerns with representativeness and modes effects. This requires special attention to sample design and data collection as well as tailored weighting adjustments.

This paper discusses the methodology of a survey on race and discrimination that utilized samples stratified to oversample high-density Black areas augmented by a web-based probability panel. To prevent modal effects from affecting only estimates from the Black sample, a sample of non-Black respondents was also reached through this hybrid design. Several post-hoc adjustments were needed to properly account for oversampling of high-density Black strata for the phone component as well as propensity to be included in the panel. Additionally, post-stratification weighting needed to adjust for known population parameters within the sample of Black respondents separately to ensure the sample matched the US Black adult population.

Even with the adjustments, some differences in response were apparent. Results indicate that respondents interviewed via web rather than telephone were more likely to report certain types of discrimination. Moreover, Black web respondents were less likely to express trust in institutions to do

what is right for them and their community than those interviewed via telephone. Overall, mode differences were more apparent among Black respondents. These differences appear to combine mode effects with the underlying differences of respondents reached by phone or online. We consider the implication for future usage of hybrid sample designs.

Weekly Response and Contact Rates From a High-Frequency COVID-19 Phone-Based Surveillance System in Lesotho

Abigail Greenleaf, *ICAP - PHIA*

Haruka Maruyama, *ICAP at Columbia*

Qixuan Chen, *Columbia University Mailman School of Public Health - Biostatistics Department*

In response to the COVID-19 pandemic, with support from the Centers for Disease Control and Prevention (CDC), ICAP at Columbia University established a phone-based surveillance system to collect influenza-like illness (ILI) symptoms weekly from approximately 1,700 respondents who had participated in face-to-face survey in Lesotho, the Population-based HIV Impact Assessment (PHIA) Survey. Lesotho PHIA is a nationally representative survey that took place between December 2019 and March 2020. As part of national surveillance efforts for SARS-CoV-2 transmission, 16 interviewers began enrolling participants on July 1, 2020 by calling 3,020 Lesotho PHIA respondents who consented to be recontacted. Over two weeks, 1,778 participants were enrolled and weekly phone calls enquiring about ILI symptoms of the participant as well as household members listed during the face-to-face survey began the third week of July. The data collected is used to create point-estimates of percent of the population that had ILI each week, fulfilling the first aim of the World Health Organization's COVID-19 surveillance guidance. This analysis will present contact and response rates for the first six months of data collection. First, we will share the disposition codes and final outcomes for the enrollment period. We will then present contact and response rates for six months of data collection, analyzing the impact of interviewers on response rates, text messages for non-respondents, and calling respondents who have been traveling monthly instead of weekly. We will also look at cause of attrition for household members. The results of this analysis will inform future high-frequency, extended period surveillance and phone surveys in low and middle-income countries about the anticipated attrition and workload.

Innovative Methods to Increase Child Interviews in Push-to-Web Surveys

Brian Wells, *UCLA Center for Health Policy Research*

Royce Park, *UCLA Center for Health Policy Research*

Growing difficulties with survey data collection have prompted multiple household surveys to consider methodological redesigns. However, for many household surveys the focus is not just on collecting data from a randomly selected adult, but also to collect data on other household members, in particular children. While large-scale methodological changes will likely benefit the collection of data on children, redesigns are also an opportunity to reevaluate where additional improvements can be made, specifically tailored for the newly adopted sampling frame and mode of data collection. This paper examines the benefits a statewide household survey observed by moving to a new sampling and data collection design, and investigates two methods aimed to increase the number of child survey completes: 1) placing the child survey questions before the adult survey, and 2) oversampling households with children through predictive modeling.

In 2019, the California Health Interview Survey (CHIS) implemented a mixed-mode (web and telephone) design using address-based sampling. CHIS choose to implement two additional methodological changes with the overall design based on positive results regarding the importance of ordering for multiple within-household survey requests (Park et al., 2019) and the potential for success of machine learning methods to predict household attributes (Dutwin, 2018). With all these changes combined, CHIS achieved over 3,000 child completes in 2019 — nearly double the number of child completes compared to 2018 and the highest number of child completes since CHIS 2012. These innovative changes have promise for other household surveys looking to add or improve child data collection.

Exploring Motivators of and Barriers to Racial and Ethnic Minority Survey Participation

Thomas Brassell, *ICF*
Matt Jans, *ICF*
Randy ZuWallack, *ICF*
John Boyle, *ICF*

As survey participation declines across the United States, one particular concern is the implication for research on minority populations that are already difficult to survey. Despite improved methods for surveying these populations, there appears to be little by way of established protocols for engagement. This is partly because research has not adequately addressed motivations to participate. This study explores participation motivations among Black and Hispanic populations. Data come from a national non-probability, census-balanced mobile panel survey of 2,000 adults, which asked questions about potential motivators to participate, attitudes about surveys, survey-taking preferences (e.g., mode, topic), and survey participation history.

The current study examined participation likelihood and motivation by race and ethnicity focusing on differences in attitudes and motivators towards surveys between Black, Hispanic, and White respondents. Black respondents felt they are surveyed more frequently and that surveys are more burdensome (difficult to complete) than White respondents. Black respondents were more likely than White respondents to prefer being surveyed by phone. They were also more likely than White respondents to see non-response to a survey as an expression of disapproval for government actions. Hispanic respondents were less concerned about data privacy than White respondents. Results are discussed in the context of improvements we can make to survey recruitment materials and messaging.

Targeting Spanish-Speaking Households: Results From a Methodological Experiment in the 2019 National Household Education Survey (NHES:2019)

Luke Natzke, *American Institutes for Research*

As the Spanish-speaking population continues to grow in the United States, so too does the need for accurate estimates of this group's attitudes and experiences in national surveys. Historically, Spanish-speaking households have had lower-than-average response rates on the National Household Education Survey (NHES), despite the use of bilingual mailings for households that are flagged as being likely-Spanish-speaking. Previous research suggests that tailored invitation letters can increase response in low-response propensity groups (Lynn, 2016).

This presentation presents the results of an experiment included in NHES:2019 that used targeted materials to increase response among likely-Spanish-speaking households. In this experiment, likely-Spanish-speaking households were identified based on auxiliary data available on the NHES sampling frame and appended from publicly available sources. This group was then sent targeted mailing materials that were developed through a series of focus groups with Spanish speakers to engage or appeal to this group. For example, the letters emphasized key themes that emerged in the focus groups, the paper questionnaire included images of Hispanic families, and a Spanish-first presentation was used for the bilingual materials. However, our findings suggest that, among likely-Spanish-speaking households, the targeted mailings decreased the response rate as compared to the standard NHES materials. We also find that the flag used to identify likely-Spanish-speaking households was possibly imprecise. In this presentation, we summarize results from this experiment and offer lessons learned for future targeted designs for surveying Spanish-speaking populations.

Neurocognitive Assessment Data Collection on the Comparative Health Assessment Interview (CHAI) Research Study

John Sokolowski, *Abt Associates*

Aaron Schneiderman, *Veterans Health Administration, Post-Deployment Health Services*

Christopher Hansen, *Veterans Health Administration, Martinsburg VA Medical Center*

Nicholas Ruther, *Abt Associates*

Ashley Bradbury, *Abt Associates*

The Comparative Health Assessment Interview (CHAI) Research Study, funded by the Veterans Health Administration (VHA) Post-Deployment Health Services at the Department of Veterans Affairs (VA), and conducted by Abt Associates, Inc. is examining the effects of military service, deployment and combat on the health and well-being of post-9/11 Veterans. Data collection for CHAI consisted of two primary components; (1) a Core Survey conducted via web and telephone with a national sample of post-9/11 Veterans and (2) a follow-up Neurocognitive Assessment Study administered in-person to a subsample of Core Survey Veterans in six geographic areas with the highest concentration of Veterans. The CHAI Neurocognitive Assessment Study consisted of 10 tests which seek to (1) describe the cognitive function of post-9/11 Veterans, (2) contrast measures of cognitive function with select measures of health and well-being among post-9/11 Veterans, and (3) examine the differential impact of health and psychological factors on post-9/11 Veteran performance in neuropsychological domains such as executive function, memory, attention, processing speed, and neurobehavioral symptoms. This methodological brief will discuss the methodological design, data collection logistics, and quality control procedures implemented in the administration and completion of the Neurocognitive Assessment Study yielding an 80% response rate (AAPOR RR5). Topics to be discussed include recruitment of Veterans from the Core Survey, recruitment of data collection technician staff with the appropriate level of education and experience required for the highest levels of precision and standardized administration of each test, training and management of data collection technician staff, scoring of assessments and quality control as well as lessons learned.

Concurrent Session D: Live Sessions
5/12/2021, 2:30 PM - 4:00 PM

Mini-Conference: Shifts From In-Person to Phone Collection During the Pandemic

Pilot Testing the Shift From In-Person to Phone Data Collection on the Medicare Current Beneficiary Survey (MCBS)

Samantha Rosner, *NORC at the University of Chicago*
Shena Patel, *NORC at the University of Chicago*
Andrea Mayfield, *NORC at the University of Chicago*
Jennifer Vanicek, *NORC at the University of Chicago*

Medicare Current Beneficiary Survey (MCBS) is a longitudinal survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. The MCBS collects detailed data from Medicare beneficiaries and proxies living in the community and from facility staff on behalf of beneficiaries living in long-term care facilities, including health care use and expenditures, health status, and other factors that affect health care utilization.

Due to the COVID-19 pandemic, public health officials issued guidance limiting in-person interactions, first in long-term care facilities in early March 2020 and then for people living in community settings in mid-March. In response, CMS and NORC paused in-person data collection, the MCBS' primary data collection mode for community and facility interviews, to ensure the health and safety of respondents and interviewers. To continue collecting data while adhering to in-person visit restrictions, NORC conducted fast-track pilot testing to study the shift to telephone-only data collection.

The purpose of this presentation is to summarize the planning and implementation efforts needed to transition to telephone interviewing, which occurred on a compressed timeline. This was particularly challenging given that the MCBS is a continuous, large scale interviewing operation, which interviews 16,000 Medicare beneficiaries during three rounds of data collection annually over a four year period. We will present the pilot testing framework, which assessed the feasibility of conducting Community and Facility interviews by telephone. Additionally, we will discuss the operational results of the pilot testing, including preliminary data collection metrics and interviewer feedback about the more challenging aspects of telephone data collection such as ensuring availability of beneficiary medical records. Finally, we will review how the pilot testing continued to inform telephone interviewing throughout the remainder of 2020.

The Impacts of COVID-19 on Survey Data Quality: Examples From the National Health Interview Survey

James Dahlhamer, *National Center for Health Statistics*

Stephen Blumberg, *National Center for Health Statistics*

Jonaki Bose, *National Center for Health Statistics*

Matthew Bramlett, *National Center for Health Statistics*

Aaron Maitland, *National Center for Health Statistics*

Benjamin Zablotzky, *National Center for Health Statistics*

The National Health Interview Survey (NHIS) is a face-to-face interview survey that allows telephone interviewing when face-to-face interviewing is not feasible. Due to this existing telephone infrastructure, the NHIS was able to quickly transition to telephone-only data collection for quarter 2 of 2020 in response to the pandemic. In quarters 3 and 4, data collection included some in-person interviewing but remained predominantly telephone-based. As expected, the move to telephone interviewing resulted in lower response rates (i.e., a roughly 17 percentage point drop in the household response rate from quarter 1 to quarter 2), the possible loss of coverage (i.e., telephone numbers not available for the full sample), and possible mode effects.

This presentation covers the array of metrics used to assess data quality and inform decisions about 2020 data releases and weighting procedures. Among the quality metrics monitored were response rates, R-indicators of sample representativeness, item nonresponse, response discrepancies, and interview times. In addition, demographic and health measures (e.g., lifetime estimates of chronic health conditions) which should not be impacted by the pandemic were also monitored to gauge coverage and nonresponse error. We summarize comparisons of these quality metrics and demographic/health estimates between quarters 2-4 and quarter 1 of 2020 (and similar time periods from 2019) and highlight select changes to post-collection activities as a result of observed differences (e.g., changes to weighting procedures in quarter 2 of 2020). The presentation concludes with lessons learned to date and implications for future data collection efforts.

Changing Modes on the Fly: Transitioning a Complex Longitudinal Survey From In-Person to Telephone Due to COVID-19

Becky Reimer, *NORC at the University of Chicago*

Kylie Carpenter, *NORC at the University of Chicago*

Liz Kantor, *NORC at the University of Chicago*

Scott Sederstrom, *NORC at the University of Chicago*

Ann Bisognano, *NORC at the University of Chicago*

Typically, when long-standing longitudinal survey projects are transitioned from one data collection mode to another, researchers devote substantial time and resources to planning, field testing, and evaluating the change before implementation. With the arrival of COVID-19 in 2020, many survey projects were forced to change course without any such research plan. Among them was the Medicare Current Beneficiary Survey (MCBS), an in-person survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. The MCBS transitioned to telephone administration in March 2020 after a brief pilot test phase in production. The present study aims to measure and understand the impact of the mode transition on the quality of the data collected across a wide scope of MCBS questionnaire variables. Many techniques included in the body of literature on mode analyses (e.g., De Leeuw and Berzelak, 2016) do not apply to this work due to the lack of experimental design and synchronous data collection across multiple modes. Instead, we assess the stability of response patterns across several years of data collected in-person prior to the pandemic with a model-based approach, and determine the degree to which the data collected via telephone in 2020 continued or broke those trends. The analysis encompasses all three rounds of data collection within the 2020 administration of the MCBS, including both the Community and Facility components. This analysis focuses on changes in reporting of medical events and their associated costs, which were typically collected by interviewers manually reviewing documents (e.g., statements) before the transition. We discuss difficulties separating the impact of the

mode change from the impact of the pandemic itself on MCBS beneficiaries, along with implications of publicly releasing data collected with this unanticipated mode change.

Data Collection During the COVID-19 Pandemic for the Survey of Income and Program Participation

Kevin Tolliver, *U.S. Census Bureau*

The Survey of Income and Program Participation (SIPP) is a face-to-face longitudinal household survey conducted yearly by the U.S. Census Bureau. During a typical data collection period, interviewers are asked to conduct interviews in-person over a four month-span, spanning late-winter and early-spring months. In 2020, the SIPP sampled 53,000 households that were approximately half new sample and half returning sample. As a result of the pandemic, midway through data collection field interviewers were instructed to stop all in-person interviews and conduct only phone interviews. While some returning sample had phone numbers listed from their prior interviews, no new sample had any contact information available. As such, the SIPP had to rely on administrative data to contact new sample and conduct interviews. Additionally, in the wake of the pandemic the SIPP modified their case prioritization strategy, using prioritization functionality as an effort to focus on specific geographies that had few completed interviews.

This work discusses how the administrative phone numbers were provided to FRs the percent of households were reached successfully reached by phone, the percent ultimately ended up with a completed (of sufficiently completed) interview, and if it improved geographic representation.

Using Contact History Instrument Paradata to Explore Changes in Field Survey Data Collection During COVID-19

Matthew Virgile, *U.S. Census Bureau*

Renee Ellis, *U.S. Census Bureau*

The coronavirus pandemic (COVID-19) poses many new challenges to data collection in Federal surveys with a field interview component, including health risks to respondents and interviewers. In response, the U.S. Census Bureau temporarily suspended in-person interviews for all of its ongoing surveys during March 2020, and had Census Bureau interviewers contact survey participants by phone where feasible. Some of these surveys began to permit in-person interviews again later in the year. Besides changes in mode, little is known about other potential differences in survey interaction during this period, such as contact strategies or respondent concerns. One way to measure these potential changes is by using paradata from the Contact History Instrument (CHI). Since 2004, the CHI has gradually been adapted for select Federal surveys with a field interview component, allowing Census Bureau interviewers to record information about each contact attempt with potential respondents.

For our analysis, we use CHI paradata to examine key changes in data collection procedures indicated by contact history records since the start of COVID-19. Some of these changes were by design, such as a shift in mode from personal visits to telephone. Other potential changes may have been initiated by respondents or interviewers; these include changes in contact strategies attempted, specific concerns or forms of reluctance expressed by potential respondents, and final outcomes (e.g. rate of complete interviews). We also consider models predicting the likelihood of obtaining complete interviews based on these factors. Exploring these distributions and their relationships will provide a more thorough understanding of changes in data collection since the start of the pandemic.

Mini-Conference: Partisanship, Policy, and Misinformation in the Context of COVID-19

The Role of Race, Religion, and Partisanship in Misperceptions About COVID-19

Jamie Druckman, *Northwestern University*

Katherine Ognyanova, *Rutgers University*

Concerns about misperceptions among the public abound. Yet, little work explores the correlates of misperceptions in varying contexts – that is, how do factors such as group affiliations, media exposure, and lived experiences correlate with the number of misperceptions people hold? We address these questions by investigating misperceptions about COVID-19, focusing on the role of racial/ethnic, religious, and partisan groups. Using a large survey, we find the number of correct beliefs held by individuals far dwarfs the number of misperceptions. When it comes to misperceptions, we find that minorities, those with high levels of religiosity, and those with strong partisan identities – across parties – hold a substantially greater number of misperceptions than those with contrasting group affiliations. Moreover, we show other variables (e.g., social media usage, number of COVID-19 cases in one's county) do not have such strong relationships with misperceptions, and the group level results do not reflect acquiescence to believing any information regardless of its truth value. Our results accentuate the importance of studying group level misperceptions on other scientific and political issues, and developing targeted interventions for these groups.

Many People Think Governments Created COVID-19 and That Influences Distancing Behavior: A Multivariate Analysis of Nationwide Survey Data From the USA

Sarah Kelley, *Child Trends*

Mariah Evans, *University of Nevada, Reno*

Jonathan Kelley, *International Survey Center*

How did the novel coronavirus get into humans? Public opinion does not match the scientific consensus, and that belief impacts behavior. In fact, we find that less than a majority (44%) of Americans think that the coronavirus came into humans accidentally from a market in China, with substantial minorities thinking instead that it escaped accidentally from a lab (28%) or was released on purpose by the Chinese government (19%). Results of a linear probability model show that accepting or rejecting natural origin is highly partisan (with Democrats more likely to believe the virus had natural origins) and only small differences between demographic groups. Those who doubt the natural origin of the virus do not have different beliefs about the characteristics of the virus: they do not, net of other factors, think the virus is more dangerous nor do they think that social distancing practices will be more or less effective at avoiding it than their peers who believe the virus had natural origins. They do, however, behave differently; they do substantially less social distancing. This suggests that there is a public health interest in understanding the beliefs the public holds about the virus (even those, like beliefs about its origins, which are not immediately about health behaviors). Data come from a large, 4 wave internet survey run by the International Survey Center (n=5126, May-June, 2020).

Does Motivated Reasoning Cross Borders? Ideology, Neighboring-Country Behavior, and Support for COVID-Related Border Restrictions in the United States and Canada

Samara Klar, *University of Arizona*

Isabel Williams, *University of Arizona*

Timothy B. Gravelle, *SurveyMonkey*

In an effort to control the spread of COVID-19, countries around the world have enacted border restrictions. These include a bilateral agreement reached between the United States and Canada in March 2020 to close their shared borders to all non-essential travel, and decisions by both countries to close their borders to refugee claimants. This policy has been politically contentious in both countries. What factors, then, shape Americans' and Canadians' support for COVID-related border restrictions? This paper reports the results of survey-based experiments conducted in parallel in the US (n = 4,875) and Canada (n = 4,426) using SurveyMonkey's endpage data collection methodology. In the US, we find that

self-described conservatives are more likely than self-described liberals to approve of the (conservative) Trump administration's border restrictions. By contrast, in Canada, we find that self-described conservatives are less likely than self-described left-leaning respondents to approve of the (center-left) Trudeau government's border restrictions. These patterns maintain even though both governments enacted the same restrictionist migration policies. When respondents are primed with information about the other country's border restrictions, American liberals become more supportive of the Trump administration's policy, while left-leaning Canadians become less supportive of the Trudeau government's policy. Individuals are therefore not only motivated to endorse (or reject) the policies of ideologically-aligned (or opposed) politicians from their own country, but are also motivated to support (or oppose) the policies of ideologically-aligned (or opposed) politicians from their neighboring country.

Primary News Source, Attitudes and Attitude Change During the Pandemic: Results From a Monthly Monitor

John Boyle, *ICF International*

Thomas Brassell, *ICF*

James Dayton, *ICF*

The fragmentation and politicization of news channels in the United States is well known. The coronavirus pandemic provides an opportunity to assess the impact of this balkanization of news reporting on public opinion about the virus during the pandemic. Nine national surveys with approximately 1,000 adults each were conducted each month from March through December 2020 using Census-balanced samples from a national non-probability panel. Respondents identified the source that they turned to most often for their news, and these were categorized as ABC/CBS/NBC, CNN/MSNBC, Fox, Internet, Social Media or other source. Over the ten months, 23% of respondents identified traditional networks, 15% CNN or MSNBC, and 18% Fox as the source they turn to for their news. Additionally, 15% identify the Internet (e.g., Google news), 11% social media (e.g., Facebook, Twitter), and 18% all others as their usual source of news. The monthly variations in these proportions across waves was small, although statistically significant.

Opinions about personal risk to the virus, economic and public health risk, how to stop its spread, restrictions on businesses and schools, confidence in institutions to control the virus, and trust of information about the virus from those sources was assessed each month. These opinions are charted by news source over the course of the pandemic to determine differences at any given point, but more importantly whether the timing, direction and magnitude of changes were similar across news sources. There were statistically significant differences in a number of these areas by primary news source in March at the beginning of the pandemic. Some opinions changed over the course of the pandemic in response to positive case rates, and other events. However, the timing and slope of these changes differ by news source.

International Polls on COVID-19

Yulia Baskakova, *Langer Research Associates*

Gary Langer, *Langer Research Associates*

Since March 2020 survey shops around the world have produced hundreds of polls on public attitudes towards the coronavirus pandemic, measuring fears, daily practices, assessment of government policies, expectations and hopes. While U.S.-based surveys naturally have held the spotlight in this country, a wealth of international research broadens the scope, shedding light on global attitudes toward this public health crisis.

Our presentation will summarize key topics in COVID-19 studies conducted outside the United States, tracing main trends and comparing results cross-nationally. We will focus on the dynamic of anxiety, adopting safe practices and attitudes toward vaccines in countries with higher and lower prevalence of COVID-19. Our analysis is based on data housed in the Societal Experts Action Network COVID-9 Survey Archive, which we operate in support of the broader SEAN project at the National Academies of Sciences, Engineering and Mathematics. By late fall 2020 the collection included more than 350 probability-based surveys from more than 30 countries outside the United States.

Mini-Conference: Weighting and Estimation in the Wake of COVID

Seroprevalence of SARS-CoV-2 Antibodies: Bootstrap Weighting and Estimation Accounting for Sensitivity and Specificity of Assays in Serology Data From Multiple Commercial Laboratories

Adam Lee, *ICF*

Yangyang Deng, *ICF*

Richard Lee Harding, *ICF*

Ronaldo Iachan, *ICF*

Tonja Kyle, *ICF*

Chris Edens, Centers for Disease Control and Prevention

Ryan Wiegand, Centers for Disease Control and Prevention

Kristina Bajema, Centers for Disease Control and Prevention

The US remains significantly affected by the COVID-19 pandemic, with more than 10 million cases and 237,000 deaths reported through November 10, 2020. While limited testing and undiagnosed mild or asymptomatic infections may contribute to an under-ascertainment of community transmission, seroprevalence surveys are important for improving infection estimates. CDC contracted with ICF, Inc. to support a seroprevalence study using residual blood samples identified by three commercial laboratories serving all 50 states, the District of Columbia, and Puerto Rico. In collaboration with CDC, ICF developed a weighting and estimation methodology to compute state-level seroprevalence estimates of SARS-CoV-2 antibodies while accounting for the bias and variability associated with false positives and false negatives associated with antibody testing. Post-stratification weight adjustments were used to ensure that the distribution of the weight-adjusted non-probability sample of individuals providing lab specimens in each state matches the state's demographic distribution.

To account for both the variability from the antibody testing and variability due to weighting, we used a bootstrap replication method. We built upon the work of Havers et al. 2020, which utilized bootstrapping to account for the diagnostic test uncertainty represented by the sensitivity and specificity. We expanded their approach to include poststratification raking within each replicate to account for the variability due to weighting, and to produce accurate bootstrap estimates and confidence intervals. This presentation describes our approach and its comparability to other estimation methods.

Regression Decomposition as a Tool for Assessing Unknown Mode Effects During the COVID-19 Pandemic

Sarah Staveteig Ford, *U.S. Department of State*

Prior to the COVID-19 pandemic, face-to-face surveys were still considered the gold standard for producing nationally-representative population estimates in many lower-income countries. During the pandemic, ethical considerations necessitated that survey implementers pivot to remote modes such as CATI, despite relatively low phone penetration. Even in the best of times, such mode changes can substantially affect survey estimates. But the pandemic era has ushered social and political changes that may themselves affect public sentiment and behavior. Without opportunity to conduct randomized mode effect studies that would ordinarily precipitate a transition from CAPI to CATI, it can be difficult to reconcile the source of changes in outcomes.

In this paper, I propose multivariate decomposition regression as a tool consistent with the TSE paradigm to help researchers disentangle compositional effects from period changes and other unobserved error. Regression decomposition, first introduced by Blinder (1973) and Oaxaca (1973), has been widely applied by economists to parse wage differences based on characteristics and unexplained factors, such as discrimination. While statistical decomposition remains uncommon in survey research, I posit that this method meets the twin TSE goals outlined by Tourangeau (2017) for mode comparison by maximizing comparability between surveys that minimize within-mode error. Regression decomposition is not a silver bullet, but when used alongside analysis of nonresponse distribution and other tools, it enables researchers to parse out the extent to which changes in particular outcomes are based on compositional sample effects, differential effects of those characteristics on the observed outcome, or unobserved error. I share examples of applied regression decomposition from opinion surveys fielded via different modes in

the same country before and during the COVID-19 pandemic where mode effects were unknown, and assess the method's strengths and limitations.

The Trials and Tribulations of Developing a Weighting Strategy to Address Nonresponse Bias in the Wake of COVID-19

Elizabeth Ayres, *Statistics Canada*

Kristine Villaluna, *University of Lethbridge*

At the onset of the COVID-19 restrictions in March 2020, Statistics Canada suspended all interviewer assisted collection efforts for non-critical programs. This presented several challenges for surveys in collection, particularly those which were interrupted during the earlier stages of collection. In many household surveys at Statistics Canada, collection begins through self-enumeration after invitation letters are mailed out to selected households. After a period of time, non-respondent units are contacted to complete the questionnaire through computer-assisted telephone interviewing (CATI). Non-response follow-up through CATI is particularly important for encouraging the participation of all eligible respondents, and reducing non-response bias.

For some surveys that were in collection at the time the COVID-19 restrictions commenced, it was not ideal to resume collection after the restrictions were lifted. These surveys were therefore faced with lower than expected response rates. For two such surveys, the Survey on Quality of Employment (SQE) and the Survey on Sexual Misconduct at Work (SSMW), the resulting response rates were reduced by approximately 10 and 20 percentage points respectively, as compared to the targeted response rate of 50%. As a result of the suspended collection efforts, responses to both surveys were predominately collected through online self-enumeration.

This paper will provide details about the challenges faced, and the different strategies explored for adjusting for this additional non-response, such as post-stratification using census information and non-response modeling using variables from administrative data sources and paradata. We will also discuss the results of a bias study that was used to inform the final weighting strategy for the SSMW. Lastly, this paper will discuss recommendations for future surveys facing lowered response rates.

Seroprevalence of SARS-COV-2 Antibodies: Lessons Learned About Data Management and Cleaning of Serology Data From Multiple Commercial Laboratories

Davia Moyse, *ICF*

Kelly Martin, *ICF*

Ronaldo Iachan, *ICF*

Tonja Kyle, *ICF*

Chris Edens, *Centers for Disease Control and Prevention*

Ryan Wiegand, *Centers for Disease Control and Prevention*

Kristina Bajema, *Centers for Disease Control and Prevention*

The US remains significantly affected by the COVID-19 pandemic, with more than 10 million cases and 237,000 deaths reported through November 10, 2020. While limited testing and undiagnosed mild or asymptomatic infections may contribute to an under-ascertainment of community transmission, seroprevalence surveys are important for improving infection estimates. CDC contracted with ICF, Inc. to support a seroprevalence study using residual blood samples identified by three commercial laboratories serving all 50 states, the District of Columbia, and Puerto Rico. In collaboration with CDC, ICF developed a data management and weighting methodology to compute state-level seroprevalence estimates of SARS-CoV-2 antibodies. This presentation describes the data cleaning process and data management steps, the development of data comparisons, and lessons learned about utilizing serology data for research.

ICF generated standardized lab data to facilitate weighting to American Community Survey (ACS) population control totals on a biweekly basis. This process included verification that all data values were valid within preset parameters and that specimens were collected within defined time periods. ICF also assessed data for consistency, duplicates, and level of missingness. To further validate biweekly submissions of cumulative serology data, we developed and executed extensive data comparisons to summarize data changes between submissions and classify changes as critical vs. non-critical to analysis. To support this classification, ICF performed data comparisons on both raw cumulative serology data and standardized processed data. Data change summaries highlighted issues such as duplication of data, age discrepancies, and missing data elements and demonstrated the importance of robust and reproducible data quality checks for public health surveillance. Other key lessons included best practices for maintaining an efficient, automated data management process and approaches for a flexible data processing system to respond to evolving data structures.

5/12/2021, 4:00 PM - 5:00 PM
Round Tables: Live Session

AAPOR 75: Looking Back to Look Forward

AAPOR 75: Looking Back to Look Forward

Janice Ballou, *Independent Consultant*

Kathleen Frankovic, *Consultant*

Don A. Dillman, *AmeriSpeak NORC*

David Dutwin, *AmeriSpeak NORC*

Frauke Kreuter, *University of Maryland, Joint Program in Survey Methodology*

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

AAPOR 75: Looking Back to Look Forward --The Meeting Place: The History of The American Association for Public Opinion Research has this description from David Sills: "AAPOR has from its inception been a fruitful meeting place, a crossroads for all manner of professional exchanges." Now, after seventy-five years of professional exchanges, AAPOR is described as "the leading association of public opinion and survey research professionals."

This round table discussion will offer reflections on how AAPOR has gained this reputation by honoring our past and recalling the founders 1946 vision; reflecting on key milestones over the years; and envisaging AAPOR's role in the future. Roundtable participants will include representatives from varied professional affiliations who have contributed to AAPOR's continuance and relevance. The following are examples of questions that will give us the opportunity to look back as we think about AAPOR's future

In 1946 Harry H. Field had a dream to bring together a broad selection of experts in public opinion and allied fields to discuss problems common to all research organizations. His closing message summarized the three day experience: "you have seen for yourselves a team that works" and "this was your conference, and I think it was you who made it a success." In a few words, how would you summarize an AAPOR conference experience?

Over the years, what has been AAPOR's most notable achievement? And, what existing "common problems" should AAPOR address? A 50th AAPOR Conference Panel asked this question: Will the pace and nature of change be so rapid and fundamental that we will witness "the end of surveys and polls as we currently know them?" Twenty-five years later, how do we answer this question?

Understanding Texas's Changing Electorate: Geography, Demographics, & Turnout

Understanding Texas's Changing Electorate: Geography, Demographics, & Turnout

Mark Owens, *University of Texas at Tyler*

Jeronimo Cortina, *Siena College*

Meghann Crawford, *Siena College*

Kirby Goidel, *Texas A&M University*

This roundtable discussion addresses the unique challenges of state level public opinion surveys in Texas. The diversity of the state is immediately apparent, however how should experts best understand asymmetries within the population geographically, demographically, and voter participation? Insights will be shared by Josh Blank (University of Texas), Jeronimo Cortina (University of Houston), Meghann Crawford (Siena College), Mark Owens (UT Tyler), as well as moderator Kirby Goidel (Texas A&M University). Each of these survey researchers have used different methods to sample registered voters in the state at different points in the past two years to explain how changes within the state have affected survey research. The large and diverse Latino population in Texas allows large samples Texas surveys to identify regional and nationality variation within the state. Texas is also simultaneously experiencing rapid urbanization and a lack of connectivity with the lag of rural broadband. The roundtable will discuss if such differences offer meaningful insight for how we can improve the representativeness of surveys, consider who is likely to vote, and how research designs can be adapted to reduce non-response bias. The roundtable will introduce AAPOR attendees to how the state is changing and why the state's election outcomes look differently when infrequent voters become engaged in the election.

Introduction to Bayesian Statistics Using Stata

Introduction to Bayesian Statistics Using Stata

Chuck Huber, *StataCorp*

Bayesian analysis has become a popular tool for many statistical applications. Yet many data analysts have little training in the theory of Bayesian analysis and software used to fit Bayesian models. This talk will provide an intuitive introduction to the concepts of Bayesian analysis and demonstrate how to fit Bayesian models using Stata. No prior knowledge of Bayesian analysis is necessary and specific topics will include the relationship between likelihood functions, prior, and posterior distributions, Markov Chain Monte Carlo (MCMC) using the Metropolis-Hastings algorithm, and how to use Stata's Bayes prefix to fit Bayesian models.

Thursday, May 13, 2021 | Day 3 of Conference

5/13/2021, 9:30 AM - 11:30 AM

Concurrent Session E: Prerecorded Sessions

Mini-Conference: Nonresponse During a Pandemic: Controlling and Understanding

Achieving a High Response Rate on a Panel Survey During the COVID-19 Pandemic

Natalie Fitzpatrick, *University of Michigan*

Debra Horner, *University of Michigan*

In spring 2020, as the Center for Local, State, and Urban Policy (CLOSUP) conducted the 22 wave of its panel survey of local government officials in Michigan, the COVID-19 pandemic presented a number of specific challenges to survey administration: 1) due to a statewide stay-at-home order, many local governments reduced their operations to essential services, meaning fewer respondents “in” the office; 2) although the survey included questions on the COVID-19 pandemic, the bulk of the survey addressed other topics, meaning respondents might not want to spend time answering questions about less pressing issues; and 3) officials were under significant professional and personal stress, meaning investigators were concerned about alienating them during this wave and negatively affecting future participation.

This presentation reports on response rates from the Michigan Public Policy Survey (MPPS), a mixed-mode paper/online census survey of Michigan local governments (1,856 counties, cities, villages, and townships) conducted once or twice per year since 2009. The survey consistently receives responses from over 70% of all Michigan local jurisdictions. The 2020 wave was in the field from March 30-June 1, and received a 72% response rate. To understand how the MPPS was able to maintain its high response rate in the face of extraordinary circumstances, we use daily response tracking data to compare response rates in spring 2020 with 21 prior waves of the MPPS. We saw relatively high response rates during the first two weeks of the survey’s field period, but significant declines over subsequent weeks, particularly among counties and villages. The daily tracking is used to evaluate the effectiveness of variations in weekly recontact email language and other administrative interventions, and to examine the effects of external conditions, including extensions of and changes to Michigan’s stay-at-home order, on survey response rate.

Surveying Nursing Facilities During the COVID-19 Pandemic

Eran Ben-Porath, *SSRS*

Gillian SteelFisher, *Harvard T.H. Chan School of Public Health*

Akilah Evans-Pigford, *SSRS*

Skilled nursing facilities, or nursing homes, have been among the hardest hit by the COVID-19 pandemic. In the span of several months, more than three hundred thousand residents in these facilities were diagnosed with COVID, and more than sixty thousand died as a result. In view of this tragedy, public health researchers are turning to surveys as a means of understanding the ongoing challenges that facility directors have faced and continue to endure. This brief discusses two approaches to reaching nursing facility directors: phone interviews (CATI), and mail-based recruiting to online and hardcopy questionnaires. The CATI approach was used for a quicker turnaround, while the mail recruit involved a more complex set of contact attempts and a sizeable incentive to drive higher response rates. Both approaches proved challenging, but ultimately succeeded in giving voice to nursing facility leadership and shedding light on the experience and needs in this critical part of the American health care infrastructure. As nursing facilities remain highly vulnerable to the impact of COVID-19, we design this brief as guidance for future research. We first discuss the various steps required for reaching respondents: namely, sample sources and contact protocols for each approach designed to identify and reach the appropriate respondent. We then discuss the response rate implications, cost considerations, and possible mode effects associated with these approaches, and then highlight some limitations of approaches to date.

Comparison of Respondents and Non-respondents to a COVID-19 Impacts Survey of Businesses

Martha Stapleton, *Westat*

Hany Sun, *Westat*

Victoria Hoverman, *Westat*

Kerry Levin, *Westat*

Brandon Bedford, *Montgomery County Economic Development Corporation*

Benjamin Kraft, *Montgomery County Planning Commission*

Sarah Miller, *Montgomery County Economic Development Corporation*

With the emergence of the novel coronavirus (COVID-19) in the United States in February 2020, and its unprecedented impact, the Montgomery County Economic Development Corporation (MCEDC) in Maryland initiated the COVID-19 Targeted Industry Adaptation and Recovery Study. The goals of the study are to understand how local business owners are being affected by the pandemic, as well as how they are adjusting to changing circumstances over time. A probability-based sample of a little over 2,400 businesses was drawn from the Dun & Bradstreet database to represent all businesses in Montgomery County. The design involves administering a baseline survey and up to five subsequent follow-up surveys between July 2020 and May 2023. For each iteration, data is collected over approximately two weeks, with up to three reminder email follow-ups. The survey includes 25-30 items and takes about 15 minutes to complete. Survey questions ask about the impact of COVID-19 on business sales and operations, strategies used for adaptation and innovation, and the resources needed for rapid recovery.

This paper will compare baseline respondents with non-respondents on a set of auxiliary variables, including industry, business size, revenue, and ownership. We will also compare respondents who participated in the baseline survey but then dropped off during the first follow-up survey with their counterparts who participated in the baseline and follow-up survey on key survey items, such as, the impact of COVID-19 on sales and operations and resources needed for rapid recovery once the economy reopens.

You Can Only Watch So Much Netflix: How Response Rates to Mail Surveys Increased During the Pandemic

Jordon Peugh, *SSRS*

Susan Sherr, *SSRS*

Royce J. Park, *Center for Health Policy Research, UCLA*

The emergence of the COVID-19 pandemic in the United States brought sudden and dramatic changes to society. Through the Spring of 2020, most people curtailed their activities and movements, and many became cautious about opening mail for fear of virus transmission. For surveys that rely on mailed invitations, there was uncertainty as to whether more time at home would increase response or whether mail remaining untouched for extended periods of time could depress response. This research presents data about the impact of the pandemic on mail-to-web response rates in two studies fielded in the Spring of 2020. These studies collected data both prior to and during the pandemic using the same methodology, allowing for a natural pre/post comparison.

The first study we discuss, the California Health Interview Survey (CHIS), collects approximately 20,000 interviews with California residents each year using an ABS push-to-web methodology. In 2020, fielding began in early March, just a few days prior to the statewide shutdown. Comparing 2020 yields to comparable waves of the 2019 collection demonstrates a 36% increase in mail-to-web response over 2019. Second, we will discuss a tracking study that surveys U.S. adults about their travel behavior. Each month, a fresh ABS sample of households is invited by mail to take a web survey. In March, we saw a 30% increase in web response rates over February results. Overall, the peak COVID months of March/April/May/June averaged a web response rate that was 38% higher than that of the same time period in 2019. This paper will present the findings described in detail, look at the overall impact on the two studies, and track response trends as the pandemic continues.

New York City Community Health Survey (CHS) 2020 and COVID-19: Response to the Pandemic, Obstacles Overcome and Impact on Survey Results

John Sokolowski, *Abt Associates*

Michael L. Sanderson, *New York City Department of Health and Mental Hygiene*

Steven Fernandez, *New York City Department of Health and Mental Hygiene*

Amber Levanon Seligson, *New York City Department of Health and Mental Hygiene*

Michael Witt, *Abt Associates*

Andrew Burkey, *Abt Associates*

Nicholas Ruther, *Abt Associates*

Anders Hansen, *Abt Associates*

The NYC Community Health Survey (CHS), conducted by the NYC Department of Health and Mental Hygiene (DOHMH), is an annual survey of 10,000 randomly selected adult New Yorkers. The CHS helps track the health of New Yorkers, guide health program decisions, understand how different behaviors can affect a person's health, and support health policy initiatives. The CHS also provides valuable information on the extent to which diseases, including COVID-19, and risk factors occur at the neighborhood, borough and city levels. This presentation discusses real time changes made to the CHS survey and protocols due to the COVID-19 pandemic and how the pandemic affected results. The presentation also summarizes our response to the closure of physical call centers and the establishment of a Virtual Call Center (VCC) that allowed telephone interviewers to conduct the CHS from home to ensure their safety and the continuity of the data collection effort. Lastly, we examine the impact of the pandemic and the VCC on results, including response rate, which was initially higher and decreased as pandemic restrictions were lifted, and significant differences in the distributions of responses to key survey questions. We summarize our analysis of changes in survey response and health estimates during the pandemic and present various standardized estimates to explore the potential causes of significant differences in some survey results. Included in our analysis will be an examination of estimates by survey month which will allow us to measure change throughout the data collection period from the height of pandemic restrictions through the easing of restrictions. We examine interviewer effects related to the VCC, nonresponse bias compared to prior survey years due to changing response, and the possibility that differences in health estimates were due to real change in the health status of New Yorkers because of the pandemic.

Impact of the COVID-19 Pandemic on Data Collection Outcomes in a National US Health Survey

Andrew Caporaso, *Westat*

Richard P. Moser, *National Cancer Institute*

Ashley Murray, *National Cancer Institute*

Gordon Willis, *National Cancer Institute*

Terisa Davis, *Westat*

David Cantor, *Westat*

Kelly Blake, *National Cancer Institute*

This presentation will examine response rate and return outcomes for the National Cancer Institute's Health Information National Trends Survey (HINTS) before and after the COVID-19 pandemic declaration, as a preliminary attempt to ascertain the impact of the pandemic on data collection in a national survey. HINTS is a cross-sectional postal survey of US adults that monitors how and where Americans seek and use health information, and measures associated health behaviors and health outcomes. HINTS is currently fielded annually to a sample of about 14,000 households. Cycle 4 of HINTS was underway when the World Health Organization (WHO) declared COVID-19 a pandemic on 3/11/2020.

We will compare response rates, respondent composition, and key HINTS indicators for the Cycle 4 administration to previous HINTS cycles and before and after the WHO declaration, which occurred 16 days after the start of Cycle 4. About two thirds of the 3,865 responses were received after the declaration. Cycle 4 achieved a higher overall response rate than Cycle 3 (2019) and comparable rates to Cycles 1 and 2 (2017 and 2018). The proportion of returns received after the declaration was similar to the proportion received during the same time frame as past HINTS Cycles. The proportion of returns received from the 2 survey mailings was higher than in past cycles. These findings suggest that the

pandemic had a delayed impact on data collection, if any. The presentation will compare the pre- and post-declaration respondents from Cycle 4 with respect to sociodemographic indicators, as well as selected health communication outcomes. We will compare across high and low response areas, as well as early and late responders across prior HINTS cycles to examine whether there was differential response in Cycle 4 before and after the pandemic declaration.

Emergency is the Mother of Invention: Efficacy of Pre-Recorded Follow-Up Voice Messages Versus Live Interviewer Follow-Up Calls on Return Rates

Kimberly Hawkins, *Nielsen*

Robin Gentry, *Nielsen*

Nielsen has found positive results using automated, pre-recorded messages nine days after the diary survey ends, as a final reminder to return the diary. Further, Skalland et al found that automated messages can substitute for live interviewer messages in early recruitment attempts to induce the rate of contact with no significant impacts to response rates (Skalland, 2018).

During the early days of COVID, for the sake of business continuity, Audio diary methodology switched to automated, pre-recorded messages in place of the 1-3 live interviewer follow-up calls that occur throughout the diary-keeping week. Follow-up calls are a way to remind diary-keepers to complete and return their seven-day radio diaries and often serve as the first point of voice contact with a respondent. Once the workforce was reconfigured to work-from-home and large enough to handle the follow-up call volume, the process was reverted back to live follow-up calls

In order to compare the return rates of the two reminder methods in a controlled experiment and determine whether there was a financial benefit to replacing live follow-up calls with automated, pre-recorded messages, a live production test was conducted in Summer 2020. Because the Audio diary service is a dual-frame service, results will be shown separately for both random digit dial (RDD) and addressed-based sample as well as demographics, such as race.

Addressing Nonresponse Error in a Longitudinal Tracking Survey Measuring the Economic Impact of the Pandemic

Michael Karpman, *Urban Institute*

Douglas Wissoker, *Urban Institute*

Timothy Triplett, *Urban Institute*

The pandemic-induced recession caused a sharp increase in unemployment that could affect rates of material hardship and health insurance coverage. To assess the pandemic's economic impact, we fielded a longitudinal, rapid-turnaround tracking survey with a nationally representative sample of adults ages 18 to 64. Tracking survey participants were recruited from a survey conducted in March/April with members of a probability-based online panel. We sampled a subset of these participants for two follow-up survey waves fielded in May and September, each with approximately 4,000 respondents. Estimating changes over time in the pandemic's impact required addressing differential nonresponse between adults whose families lost jobs or work-related income during the pandemic and adults whose family employment was unaffected. Nonresponse rates in the follow-up tracking survey were higher among adults who reported experiencing negative employment effects in March/April, and these differences in nonresponse increased between the first and second tracking survey waves. In addition, employment effects, and therefore nonresponse, varied by demographic characteristics such as race-ethnicity. To mitigate nonresponse bias, we adjusted survey weights to make the September sample representative of the initial March/April sample in terms of employment and hardship outcomes that were reported in March/April. Within each race-ethnicity group, we developed a propensity score for response to the September survey based on these March/April outcomes and other control variables used in weighting. Both the March/April and September samples were split into groups based on these propensity scores and race-ethnicity. We then included the weighted shares of March/April respondents in these groups as control totals in a raking method to reweight the September sample. This reweighting approach aligned weighted estimates for March/April outcomes in the September sample with estimates from the March/April sample, and likely reduced nonresponse error in estimated trends in employment, health insurance, and food insecurity and other hardships.

Collecting Data From or at Establishments

Using a Delphi Panel to Determine Best Practices of Rural Community Engagement for the Tennessee Department of Transportation

Makena Nail, *East Tennessee State University*

Candace Forbes-Bright, *East Tennessee State University*

Kelly Foster, *East Tennessee State University*

Moin Uddin, *East Tennessee State University*

Morgan Kidd, *East Tennessee State University*

Madison Kosky, *East Tennessee State University*

The Applied Social Research Lab (ASRL) at East Tennessee State University (ETSU) conducted a Delphi panel for the Tennessee Department of Transportation (TDOT) to provide recommendations for best practices of rural community engagement in four rural communities: Cookeville, Gordonsville, Brownsville, and Greenville. ASRL then conducted interviews with 24 community leaders about their perspectives on how TDOT can best engage with their community. These interviews, which included questions about community meetings, media and technology use, and barriers to engagement, were analyzed to develop a survey for the Delphi panel. The survey included 32 questions across the domains of community engagement with TDOT, community leadership engagement with TDOT, and community meetings, including questions on virtual, in-person (pre-Covid), locations, times, and media advertising of meetings. The purpose of this survey was to find a consensus with the recommendations of community leaders and to create actionable recommendations for TDOT based on the results. Of these 32 statements, 17 were included in the second round of the survey based on meeting the desired consensus threshold. This presentation will focus on the results of the second round of the Delphi panel and the supporting qualitative data.

Managing Nonresponse in a Required, Weekly Survey

Megan Waggy, *RTI International*

Jessica Williams, *RTI International*

Sara Russell, *RTI International*

Amerine Woodyard, *U.S. Energy Information Administration*

The EIA-878, Motor Gasoline Price Survey and EIA-888, On-Highway Diesel Fuel Survey are two mandatory, weekly surveys for the U.S. Energy Information Administration (EIA). Every Monday morning, RTI International surveys a static sample of approximately 400 stations selling diesel fuel and over 1,000 stations selling motor gasoline to collect their posted fuel prices. This data is aggregated and published by EIA as national and regional level price estimates by close of business Monday, the same day price information is collected. Data are collected through multiple modes: phone, email, website, text message, and fax. To ensure up-to-date and sufficient coverage, stations are rotated on and off sample every few years. Gaining and maintaining cooperation of selected stations is necessary to produce accurate estimates. This presentation will cover an evaluation of our sample to determine if there were any trends in nonresponse based on station characteristics, physical location, current events, and prior reporting history, and discuss possible techniques for decreasing nonresponse among typical nonrespondents.

Just You Wait... and Fill Out This Survey: Methodological Aspects of Waiting Room Surveys

Yfke Ongena, *University of Groningen*

Marieke Haan, *University of Groningen*

Waiting rooms are commonly conducted to study patients' attitudes, behaviors and other characteristics, but guiding literature on waiting room surveys for practitioners is scarce and outdated (Pirodda et al. 2002). In this presentation we synthesize practical guidelines from prior studies and our own experiences. We compare waiting room surveys with similar approaches such as public intercept surveys and drop-off-pick-up surveys, and we discuss our experiences in a waiting room survey that was conducted at the Radiology department of the University Medical Center Groningen. Patients scheduled for a CT or MRI were approached by students working in pairs. One student unobtrusively filled out an observation sheet, noting the waiting room (CT or MRI), the date, time, gender of the approached person, the number of people present in the waiting room, and the number of caregivers accompanying the respondent. After the request to participate, both respondents and non-respondents were asked their year of birth. Out of 249 approached patients 208 (83.5%) were willing to fill out the questionnaire, and for 189 patients (75.9%) a complete questionnaire (with at least 75% of the questions answered) was available.

From our review, we conclude that waiting room surveys, though limited to patients and their care givers, can provide useful information on patients' perspective on health care. Response rates in waiting rooms are usually high, but often not even reported. These surveys also allow for collection of para-data; i.e. relevant information in the circumstances of a request to participate in survey research, and behavior of surveyors can easily be controlled, or investigated in an experimental design.

Four Key Strategies for Achieving a High Response Rate in the Department of Labor's Ready to Work Program Evaluation

Michelle Kahmann, *Abt Associates*

Jessica Flores Pleasants, *Abt Associates*

Sandra Hernandez, *Abt Associates*

Reaching a high response rate in a follow-up survey is particularly important in an experimental research study design. The Ready to Work (RTW) Partnership Grant program, funded and administered by the U.S. Department of Labor (DOL) in 2014, was designed to assist those experiencing long spells of unemployment find jobs and succeed in the labor market. DOL sponsored an evaluation of the RTW grant program using an experimental design in four purposively selected sites. The program targeted those who had been employed for six months or more and aimed to employ them in middle and high-skill positions. Individuals applying to the RTW program at selected sites were randomly assigned to either a treatment or control group through a lottery-like system. The treatment group had access to the range of activities provided by the program including occupational training, work-based training, and job search and employment readiness services while the control group did not but could access employment and training services available in the community. A key data source of the study was a mixed-mode survey conducted 18-months after random assignment to document service receipt and a range of employment-related outcomes. The evaluation team achieved a 78 percent survey response rate.

At the conclusion of data collection, the evaluation team identified specific strategies as contributors in successfully reaching study participants for the follow-up survey. These strategies include collecting critical contact information at enrollment, communicating with the respondent periodically during the interim period between enrollment and the follow-up survey to maintain interest in the study, and equipping interviewers with a customized protocol and tools to engage study participants. This presentation highlights these three strategies and we share our thoughts on how each of these strategies contributed to the overall response rate.

Measuring Abortion Attitudes

An Interdisciplinary, Multi-Dimensional Approach to Developing New Abortion Attitude Measures in the United States: A Project Overview

Kristen Jozkowski, *Indiana University*
Brandon L. Crawford, *Indiana University*
Michael Traugott, *University of Michigan*
Ronna C. Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*

We theorize people's attitudes toward abortion are complex and multi-dimensional, requiring innovative approaches to assess such multi-dimensionality. We are conducting a multi-phase, multi-year project to concurrently develop a comprehensive array of instruments (e.g., individual polling items; item scale sets) in English and Spanish to assess abortion attitudes in the US. The purpose of this presentation is to provide a detailed description of the methodological approaches we are using to develop these bilingual measures.

After conducting comprehensive literature reviews of both prior content and methodology, we began formative data collections in English and Spanish. After cognitive pre-testing, we fielded an open-ended survey to a sample drawn from NORC's AmeriSpeak panel (N = 608). These data provided insights into people's salient behavioral, normative, and contextual beliefs regarding abortion. We conducted two pilot surveys (N = 512; N = 1586) comprising existing and experimental items aimed at distinguishing people who may be consistently polarized in their abortion perspective (anti- or pro-abortion) from those with more nuanced perspectives, with follow-up pilot interviews (N = 24). After refinement, we will administer a finalized version of this survey to NORC's AmeriSpeak panel (N = 600) and conduct follow-up in-depth interviews (N = 200). We also conducted four additional formative studies designed to test item content and instrument design hypotheses related to abortion attitude measurement. We will integrate these findings with other results to identify common themes and underlying constructs that will drive item development for our new abortion attitude scales and polling items. Multi-stage pilot assessments and cognitive interviews will be used to assess readability, item-level content validity, psychometric functioning, and consistency across survey modes, with revisions made as needed. Finally, we will administer a comprehensive set of new items to a large, nationally representative sample of US English- and Spanish-speaking adults (N = ~26,000).

Legal or Illegal: Examining Differential Response Patterns in Abortion Attitude Items When Asked About Legality Versus Illegality

Brandon Crawford, *Indiana University*
Kristen N. Jozkowski, *Indiana University*
Ronna C. Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*

Because abortion is a complex social issue, responses to abortion attitude items on surveys may be particularly sensitive to differences in question and answer formats. Specifically, differences in the number of response options, question framing, and dimensionality of answer choices may yield significantly different assessments of abortion attitudes. For example, questions assessing abortion attitudes in specific circumstances, such as the core abortion items in the General Social Survey, focus on the dimension of legality. These questions typically ask if abortion should be legal in a particular context followed by answer choices of "Yes" and "No". However, circumstantial abortion questions are seldomly worded such that the respondent is asked if abortion in that circumstance should be illegal. Although responding with a "No" to these types of circumstantial abortion legality questions is technically the practical equivalent of saying abortion should be "illegal", we hypothesize that participants may be more uncomfortable saying that abortion should be illegal, and therefore less likely to select "Illegal" than "No" on similar questions.

To test this hypothesis, respondents from IPSOS's Knowledge Panel (N=919) were provided both the web equivalent of the core GSS abortion items with response options of "Yes" and "No" and versions of

the questions with response options ranging from “Definitely Illegal” to “Definitely Legal”. Results show there is higher agreement between answering “Yes” to the GSS questions and “Legal” to the modified version (average agreement of 90%) than answering “No” and “Illegal” (average agreement of 72%). In addition to discussing the differences in agreement across the seven items, we will also discuss which demographic sub-groups were most likely to have higher levels of disagreement between the “No” and “Illegal” response options. Our presentation will conclude with a discussion of the relevance of our findings when assessing abortion attitudes and other complex social issues.

All for One or One for All: The Impact of Using Terms With Multiple Interpretations on Item Response Distributions

Ronna Turner, *University of Arkansas*
Kristen N. Jozkowski, *Indiana University*
Brandon L. Crawford, *Indiana University*
Wen-Juo Lo, *University of Arkansas*
Michael Traugott, *University of Michigan*

Determining wording for survey items assessing complex issues like abortion can be difficult. Item response trends can provide important information regarding item interpretation and content validity. We investigated the terms “for any” in reference to circumstances and “at any” in reference to timeframes for abortion items to examine the potential for multiple interpretations of “any” as defined in Black’s Law Dictionary (1990, p. 94) -- ‘Any’...may indicate ‘all’ or ‘every’ as well as ‘some’ or ‘one’ depend[ing] upon the context...”

Using data from three studies, we investigated how participants interpreted “any” within item contexts. First, using pilot data collected via Qualtrics (N = 245), we assessed eleven abortion attitude items modeled after the seven General Social Survey (GSS) abortion legality items. We compared patterns of responses across a range of abortion circumstances (e.g., rape, life endangerment) with the “...if she wants [an abortion] for any reason” GSS item. Significantly fewer respondents support abortion legality under two circumstances than the “for any reason” item. Interview data indicated that some respondents interpret “for any reason” to mean for all reasons or in all circumstances, whereas others interpreted “any reason” to mean in at least one circumstance. We replicated this study using Ipsos’ nationally representative Knowledge Panel (N=919), finding similar trends. Finally, we collected data from a second Qualtrics sample (N=2,442) to investigate responses to the GSS abortion legality items when gestational timeframes were added to item stems. Timeframes included 6, 12, and 20 weeks, ending with “any time” during the pregnancy. More participants indicated abortion should be legal “any time” during a pregnancy than at 20 weeks, suggesting potential diversity in interpretation of “any” in reference to timeframes. Understanding different interpretations of “any” may help researchers contextualize potential inconsistencies in people’s responses to abortion legality items.

A Psychometric Evaluation of 2-Point, 4-Point, and 6-Point Likert Scales to Assess Complexity in Abortion Attitudes

Wen-Juo Lo, *University of Arkansas*
Brandon L. Crawford, *Indiana University*
Ronna C. Turner, *University of Arkansas*
Michael Traugott, *University of Michigan*
Kristen N. Jozkowski, *Indiana University*

There does not appear to be consensus regarding the optimal number of points to use on rating scales when assessing people’s attitudes toward social issues like abortion (Krosnick & Presser, 2010). Although dichotomous scales (e.g., Yes, No) are widely used in many surveys, adding more points to a rating scale provides participants with greater diversity in response options, which could result in more valid and reliable findings when assessing attitudes toward complex issues like abortion. The purpose of this study is to evaluate the Likert scale with three even number response formats in terms of their latent constructs and to compare the intensity of attitudes toward abortion using General Social Survey (GSS) abortion items.

A total of 1,006 participants were randomly assigned to complete one of three sets of questions, which included different modifications to the six GSS items assessing abortion under different circumstances (e.g., rape, life endangerment). The first set of items modified the original dichotomous options from Yes/No to Agree/Disagree. The second set provided a 4-point option by adding “strongly agree” and “strongly disagree” to the 2-point agree/disagree format. The third set provided a 6-point scale, adding “strongly,” “moderately,” and “slightly” to both agree and disagree categories.

We first conducted a chi-square test to evaluate whether our 2-point scale had similar response patterns to the nationally representative GSS data from 2018. Second, we conducted confirmatory factor analysis (CFA) to evaluate whether each scale still held the two-factor model (Muthen, 1981). Third, to investigate the influences between the different number of response options, we aggregated a 4- and 6-point scale to a dichotomous scale in comparison with the standard GSS format. Our findings suggest that adding points to the rating scale demonstrate greater complexity in response patterns. Implications for abortion measurement will be discussed.

Developing New Measures to Assess Abortion Attitudes in the United States

Michael Traugott, *University of Michigan*

Kristen N. Jozkowski, *Indiana University*

Brandon L. Crawford, *Indiana University*

Ronna C. Turner, *University of Arkansas*

Wen-Juo Lo, *University of Arkansas*

Abortion is a nuanced and complex issue, which makes measuring abortion attitudes and opinions conceptually and methodologically challenging. Previous research suggests that people may feel conflicted or ambivalent toward abortion or may hold seemingly competing opinions when responding to abortion polling questions. For example, some people who identify as pro-life and indicate abortion is morally wrong may also indicate abortion should be legal. Given these conflicting perspectives, we theorize multiple dimensions may comprise people’s attitudes toward abortion. Standard polling questions, which can be limited in scope and rely on dichotomous response options, may not be equipped to capture this multi-dimensionality. This may, in part, explain potential dissonance across polling data, public discourse, and legislation. To account for this multi-dimensionality, we assembled an interdisciplinary team to conduct a multi-year study aimed at developing sets of abortion attitudes measures (e.g., individual items and scales) in English and Spanish. This panel will (1) include an overview of the study, (2) describe our methodological approaches, and (3) present findings from formative data collections.

The first presentation will provide an overview of the project goals and specific aims as well as a description of the methodological design for the entire multi-year, multi-method project, including a discussion of our complexity framework. The second presentation will include results from a pilot quantitative survey (N = 1586) and pilot in-depth qualitative interviews (N = 24) conducted with US adults; these data will drive construct development for new survey items. The third presentation will include results from three formative surveys and cognitive interviews demonstrating potential inconsistencies in how people interpret certain wording in abortion attitude items (e.g., “for any reason”, “any time”). The final presentation will include findings from a formative methodological study aimed at assessing item response format issues using replications of the General Social Survey abortion items.

Acculturation and Latinx Abortion Attitudes: An Assessment of Multiple Acculturation Measures

Alejandra Kaplan, *Indiana University-Bloomington*

Megan Solon, *Indiana University-Bloomington*

Brandon L. Crawford, *Indiana University*

Ronna Turner, *University of Arkansas*

Wen-Juo Lo, *University of Arkansas*

Kristen N. Jozkowski, *Indiana University*

Given the growing influence and heterogenous nature of the US Latinx community, measuring Latinx attitudes toward socially contentious topics as well as factors contributing to variance in such attitudes is important for understanding trends in the US social and political climate. This study examines the relationship between acculturation—or cultural adaptation—and US Latinx attitudes toward the legality and morality of abortion. Research suggests that heightened acculturation is associated with more supportive attitudes toward abortion. However, differences in approaches to measuring acculturation across studies make conclusions about this relationship tentative. Thus, in addition to assessing the relationship between acculturation and abortion attitudes for a US Latinx sample, this study compares the predictive properties of several acculturation measures. A web-based survey was administered using Qualtrics' panels to English- and Spanish-speaking Latinx adults (N=109). Acculturation was measured via the Hispanic Acculturation Index (HAI) the Short Acculturation Scale for Hispanics (SASH), and common acculturation proxies of language use and immigrant generation. Abortion attitudes were assessed with two scales related to respondents' opinions on the (1) morality and (2) legality of abortion under various circumstances. Multiple regression models were used to compare the predictive properties of the HAI, SASH, and acculturation proxies alongside common demographics (e.g., gender, education level, bible literalism). Comparison of the regression models showed a significant effect of demographic factors on abortion morality ($F(9,94)=6.74, 0.001, R=0.39$) and legality ($F(9,94)= 8.20, 0.001, R=0.44$), but no significant differences were found between models when the acculturation measures were added. Additionally, no differences were found between different acculturation measures. Given that our sample was relatively varied in acculturation, our results suggest little relationship between acculturation and Latinx abortion attitudes. Nevertheless, we urge further exploration of acculturation with different methods or additional measures to assess and better understand the role of Latinx identity in abortion attitudes.

Topics in Probability Panel Surveys

Examination of Impact of Panelist's Reported Satisfaction Over Time on Panel Retention in a Multi-Purpose Probability Sample Panel

Ipek Bilgen, *NORC at the University of Chicago*

David Dutwin, *NORC at the University of Chicago*

Poulami Maitra, *NORC at the University of Chicago*

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

Susan M. Paddock, *NORC at the University of Chicago*

This paper examines the relationship between patterns of change in panelists' reported survey satisfaction over time and retention in the context of a probability-based web mostly panel. Our main goal is to predict future panelist activity and attrition using: 1) change in survey-level satisfaction items (which are asked at the end of each subsequent survey) over time and 2) survey-level characteristics and interventions such as incentive amount, fielding period, survey length, survey topic, etc. and 3) panelist-level demographics and characteristics. The research question is how do survey intervention and satisfaction trends relate to panelist attributes, data quality, and attrition over time? Panel retention and high survey cooperation rates are imperative to support studies of lower incidence populations and to potentially minimize nonresponse bias. Accordingly, this study will present findings from 31,000 panelists, who responded to three items where they rate their survey experience during each survey after recruitment, using NORC's AmeriSpeak probability-based Panel. These items are asked at the end of each subsequent survey that are administered in various time points during panelists' tenure. The main modes of survey response offered to panelists in these subsequent surveys are web and phone. Panelists included in this study have participated in on average 29 different surveys in different time points during their tenure. This paper will present findings from models which investigate interactions between change in survey-level satisfaction items and survey-level interventions over time. Additionally, we will examine if and how the relationship between survey-specific variables and panelist activity differs by panelist demographics and other characteristics, including recruitment type (mail/phone or in person), mode preference, etc. Based on these results, we plan to implement targeted and responsive survey design strategies based on time series analyses with the aim of forecasting panelist attrition and increasing panel retention and data quality.

Recruitment and Expansion of an Australian Probability-Based Online Panel

Benjamin Phillips, *The Social Research Centre*

Dina Neiger, *The Social Research Centre*

Charles Dove, *The Social Research Centre*

Paul Myers, *The Social Research Centre*

Probability-based online panels are increasingly important sources for polling and opinion research. We present the results of an expansion of an Australian probability-based online panel (Kaczmirek et al 2019) using address-based sampling. The sample design aimed to address existing biases by varying sampling rates across strata defined by area-level age and socio-economic profile. We also systematically varied field methods across recruitment rounds (advance letter vs. no advance letter, pre-paid plus promised incentive vs. promised incentive only, any adult selection vs. next/last birthday). We present our findings with respect to recruitment and profile rates, retention and completion rates through to April 2021, costs, and the demographic profile of panel members. We compare the demographic profile of panel members with respect to age, gender, education, and nativity to Census 2016 benchmarks. This paper contributes to the international body of research on recruitment methods for probability-based online panels (see, e.g., Bertoni 2019; Bilgen, Dennis, and Ganesh 2018; Blom, Gathmann, and Krieger 2015; Bosnjak et al. 2018; Jessop 2018; Knoef and de Vos 2009; Meekins, Fries and Fink 2019; Pedlow and Zhao 2016; Pew Research Center 2015, 2019; Pollard and Baird 2017; Scherpenzeel and Toepoel 2012; Stern 2015; Vaithianathan 2017; Ventura et al. 2017).

Using IVR to Increase Response Rates for Non-Internet Households in the American Trends Panel

Nick Bertoni, *Pew Research Center*

The American Trends Panel (ATP) is a 100% online, probability-based panel that is the primary platform for domestic polling by Pew Research Center. In order to include offline adults in the ATP, these households are provided tablets, cellular data plans, and IT support. The tablets receive survey notifications through SMS messages via a preinstalled app.

Survey-level response rates from tablet-provided adults tend to be in the mid-60s while our other panelists respond at a rate in the mid-to-upper 80s. One reason for this is these households are tech averse, so it is harder to get them to participate. Additionally, the tablets must be turned on in order to receive survey notifications. To help combat this, we also send an oversized postcard to indicate when a survey is ready.

In 2020 Pew Research Center experimented with IVR reminder calls to consenting, tablet-provided panelists to try to boost their response rate. One reason for the interest in IVR reminder calls is that when fielding unexpected polls with short notice, there is not always enough time to print and mail the postcards. That means that if a tablet was not charged or in a place where notifications could be heard, these offline adults would not know a survey was available.

In the experiment, a random subset (50%) of consenting tablet panelists were assigned to a treatment group to receive an IVR reminder call if they had not yet completed as of a certain date. The remaining 50% was assigned to a control group and was excluded from the reminders. This experiment was conducted over three panel surveys. Analysis of the experiments determined that the ATP would use IVR reminder calls to tablet households for quick-turn surveys or surveys with a shorter field period. Findings will be shared along with details of implementation.

Using a (Relatively) High Response Rate ABS Survey to Calibrate Online Panels on Attitudinal Outcomes

Courtney Kennedy, *Pew Research Center*

Nick Bertoni, *Pew Research Center*

Nick Hatley, *Pew Research Center*

Scott Keeter, *Pew Research Center*

Arnold Lau, *Pew Research Center*

Andrew Mercer, *Pew Research Center*

One challenge with using online survey panels is that they may over-represent Democratic leaning adults, unless a corrective weighting adjustment is used. Indeed, several polling organizations surveying online include in their weighting a partisanship adjustment using targets derived from recent RDD surveys. This practice works so long as recent RDD data are available. But as response rates to telephone surveys continue declining, it is unclear how much longer RDD will be viable. This project attempts to provide an answer to the question of how researchers can calibrate partisanship in online surveys without relying on RDD data. Our proposed solution was to conduct a large, rigorous national survey of U.S. adults offering both online and offline (PAPI) response modes. The survey fielded June 1 to August 18, 2020, using a stratified sample of addresses from the U.S. Postal Service Computerized Delivery Sequence File. The first mailing included a \$2 pre-incentive and was an invitation to take the survey online. Sample households were later sent a reminder postcard and then a reminder letter. Nonresponding households were then sent via priority mail a letter asking the adult with the next birthday to complete and return an enclosed paper survey, a postage-paid return envelope, and a \$5 bill. Ultimately, n=4,109 adults completed the survey (45% online and 55% by mail), and the study yielded a 29% AAPOR3 response rate. We refer to this one-off, online + PAPI survey as our attitudinal benchmarking survey. The goal was to produce weighting targets that we could use to calibrate our online survey panel on dimensions such as partisanship (and also religious affiliation), for which no federal benchmark data is available. This presentation reports the weighted estimates from the attitudinal benchmarking survey and explores their use as weighting targets for online surveys.

Use of Email “Flair” to Encourage Study Participation Among Parents of a Nationally Representative Sample of Students

Bethany Vanspronsen, *RTI International*

McCaila Ingold-Smith, *RTI International*

Courtney Waterman, *RTI International*

Our study follows a nationally representative cohort of students to understand factors that lead to student success. Parents of these students are surveyed about family involvement in their child’s education and family characteristics that are key predictors of academic achievement and other student outcomes. Achieving participation from parents in research has become an increasingly difficult task. We contacted parents via phone, letters, postcards, and email to remind them to complete the parent survey. Traditionally, emails have been written with plain black text in a standard greeting-login information-signature format. In an effort to improve engagement with emails and parent response rates, we introduced emails with “flair” such as colorful text, animations, callout boxes, and emojis. Additionally, we utilized personalized subject lines and greetings with the parent’s and student’s names as it has been found that adding the name of the message recipient to the email’s subject line increased the probability of the recipient opening it by 20% (Sahni, Wheeler, Chintagunta, 2018). These emails yielded an uptick in parent participation and, because of the success, we plan to use these tools in future studies. This presentation will review the various elements introduced into the study’s emails and look at comparisons of previous communication materials to those that had been enhanced. Additional and alternative methods of customizing and improving emails will also be discussed.

Forecasting Voters and Votes in the 2020 Election

Revisiting Likely Voter Model 10 Years After

Masahiko Aida, *AppleCart*
Joy Wilke, *Civis Analytics*

Social Scientists understand the importance of replicating research in theory, though researchers often do not practice it. The authors will revisit our earlier work (Rogers, Aida 2013) on the likely voter screen (LV screen) and evaluate if the idea still holds today.

Earlier work found many self-predicted voters do not actually vote (flake-out), and many self-predicted nonvoters do actually vote (flake-in). The authors hypothesize this pattern will still hold. We do not know how a surge in turnout on a historic scale changes the profiles of likely voter screen errors. Did LV screen capture the true electorate in a surge election, or did it fail to identify true voters, possibly contributing to the large errors in the 2020 Presidential and Senatorial polling we have witnessed. We will use large scale voter survey responses (over 1MM interviews) from 2020 presidential election surveys, and individual-level validated votes for analysis

This research will address the following questions: How does the timing of LV response affect accuracy? (pre-COVID, summer, and post-labor day) How does partisanship impact accuracy? Were Trump supporters more likely to be screened out from the LV screen? Overall, how good is the LV screen? What proportion of responses can we categorize as false positives and false negatives? Do we observe meaningful differences between 2018 and 2020?

Strategic Polling Limits the External Validity of Early Election Polling

David Rothschild, *Microsoft*
Ling Dong, *MIT*

Using responses from a series of general election polls conducted between February and March of 2020, we found that a non-negligible share of participants withheld information about their true voter intention by responding strategically to hypothetical questions about potential future vote choice, thus introducing a source of total survey error. Although we propose a combination of ex-ante considerations in survey design and ex-post statistical fixes that can be used to reduce total survey error and infer true voter intention, the lack of repeated outcomes limits our ability to precisely calibrate our models. Our findings call into question the external validity of conclusions based on conditional forecasts using hypothetical polling results and are especially pertinent to early US general election forecasting, where potentially meaningless differences of a few percentage points can be misinterpreted as proof of differences in relative electability of potential party nominees during the Primary election season. We suggest that results from hypothetical election polls should instead be viewed as sources of partial information on future behavior, and interpreted based on contextual variables such as the moment in time relative to the predicted event.

We illustrate how the differences we observed between the estimated vote shares using data from different question iterations and from different points in time are often larger than the differences in estimated support for each of the Democratic candidates. For instance, the 8.5 percentage point difference between the estimated vote shares of the Democratic candidates with the highest (Biden v. Trump) and lowest (Klobuchar v. Trump) estimated support in our first poll iteration was dominated by the difference in estimates between the different polling iterations and across time (as high as 8.8 percentage points after changes to survey design, and up to additional 4.0 percentage points following the consolidation of candidates on Super Tuesday).

Institutions, Parties, and the Timeline of Elections: On the Variation of Pre-Election Poll Performance

Christopher Wlezien, *University of Texas at Austin*

Will Jennings, *University of Southampton*

Political institutions and parties define the set of choices faced by voters and structure the evolution of preferences over the election cycle. We theorize ways in which characteristics of institutions and parties may lead voters' preferences to come into focus earlier or later. We then implement an analysis that simultaneously assesses the influence of different system- and party-level variables on preferences using a dataset of 33,457 vote intention polls in 44 countries since 1942, covering 358 discrete electoral cycles and 329 political parties. The results indicate that both institutions and parties are important to the structure and evolution of preferences and in different ways, some of which are counterbalancing and others of which are reinforcing. These have implications for our understanding of poll performance but also for the effects of election campaigns, the prediction of election outcomes, and party behavior itself.

Forecasting the US 2020 Presidential Election Using Multilevel Regressions and Poststratification: an Evaluation of Modeling Performance at County and State Levels

Nils Osterberg, *Ipsos Public Affairs*

Meng Li, *Ipsos Public Affairs*

Kevin Fenton, *Ipsos Public Affairs*

Predicting US election outcomes using Multilevel Regression and Poststratification models provides important potential benefits when applied to traditional polling methods. Traditional polls face three practical challenges that the application of MRP may help mitigate. Firstly, predicting the national outcome of electoral votes relies on voting behaviors at the state level, or even congressional districts. Secondly, achieving sufficiently large sample sizes to estimate preferences in each state is prohibitively costly. Thirdly, traditional weighting methods face important constraints and may fail to adjust for biases in samples, especially in combination with sample size constraints. The multilevel nature of MRP enables the models to take advantage of important variables that may explain voting behavior but that are not observed directly in surveys. Moreover, the method does not rely on representative survey samples, but rather it implicitly provides a weighted outcome through the poststratification step. This allows MRP models to utilize more data and less costly sampling methods. Where traditional weighting methods fail to account for biases in the sampling, e.g. a proportional underrepresentation of rural voters, MRP may draw on the larger sample size to account for such biases. This paper examines the accuracy of MRP state and county level predictions of the 2020 US presidential election. In doing so, the paper discusses important predictors and evaluates how performance changes with sample size. In short, the study finds that applying MRP to survey data correctly indicated, at the state level, a close presidential race between Joe Biden and the incumbent Donald Trump.

AP VoteCast's Innovative Approach to Election Surveys

David Sterrett, *NORC at the University of Chicago*

Jennifer Benz, *NORC at the University of Chicago*

Trevor Tompson, *NORC at the University of Chicago*

Nadarajasundaram Ganesh, *NORC at the University of Chicago*

Benjamin Skalland, *NORC at the University of Chicago*

In 2020, The Associated Press and NORC at the University Chicago conducted AP VoteCast, a survey designed to deliver a broad and accurate picture of the American electorate. The survey of voters and nonvoters in all 50 states during the week leading up to the 2020 general elections featured 133,000 total interviews, including about 46,000 probability interviews conducted online and via phone and 87,000 non-probability interviews conducted online. Our presentation will highlight the methodological innovations of AP VoteCast for an election in which a record number of Americans voted early or by mail. We will provide an overview of the sampling and weighting methodology, the calibration models used to combine the probability and non-probability samples, and the small area models designed to improve estimates within states. We'll also assess the composition of the VoteCast electorate compared to benchmarks and

examine the influence of key voter groups in the outcome. The presentation will evaluate the success of the approach and review opportunities and challenges for this methodology.

Performance of the Opinion Polls at the 2019 Australian Federal Election

Darren Pennay, *The Social Research Centre*

Paul J. Lavrakas, *Independent Consultant*

Australian national election polls published throughout the 2019 election campaign showed that the Labor Party had the support of the majority of Australian voters. The Liberal-National Party coalition went on to win the election with 51.5% of the vote compared to Labor with 48.5%, almost the mirror opposite of what the polls found. The Association of Market Social Research Organisations and the Statistical Society of Australia quickly launched an independent inquiry to determine what, if anything, went wrong and to provide recommendations on how polling might be improved.

Unlike similar inquiries in the US and Britain, pollsters in Australia provided very limited information to the Inquiry. The lack of access to data sets and to detailed descriptions of the survey methods and statistical techniques used limited the Inquiry Panel's ability to definitively identify the specific factors that contributed to the inaccuracy of the published polls. Nevertheless, by drawing on other data sources and using the Total Survey Error (TSE) framework to structure our analysis, the Inquiry Panel was able to identify key methodological and statistical factors that likely contributed to the polling failure. The Inquiry concluded that pollsters should seek to better understand the biases in their samples and to develop more effective sample balancing and/or weighting strategies to improve representativeness. The Inquiry Panel also noted that the reporting of the polls failed to consistently meet the basic disclosure guidelines for editors and journalists set out by the Australian Press Council.

The Inquiry made 10 key recommendations aimed at improving polling methods and their disclosure. Recommendations include calling for an enforceable Code of Conduct for Election Polling; consistent, enforceable and public disclosure standards; and recommendations to improve polling methods and better educate the media and public on polling.

The presentation will summarise the Inquiry's analysis, conclusions and recommendations.

What Got It Right? An Analysis of Which Methodological Factors Yielded the Least Error in Late-Cycle U.S. 2020 Election Polling

Nathan Nathan, *Kiaer Research*

Both in 2016 and 2020, the polling industry was lambasted for overestimating Democratic support in polls and underestimating Republican support. However, some polls both at the state and national level were considerably more accurate than the typical poll was in the 2020 cycle. This study compares the methodological components of polls released in the days leading up to the election and determines which factors yielded the least error. Readers will walk away with a better understanding of which methodological factors in 2020 polls led to increased accuracy, which had no effect on accuracy, and which had a potentially damaging effect on accuracy. Findings from this study can be used to help professional survey researchers improve their ability to measure public opinion, attitudes, and behavior.

Broad Lessons and Best Practices in Questionnaire Design

Building Towards Best Practices for Questionnaire Design in Text Messaging Surveys

Kevin Collins, *Survey 160*

Oyindamola Bola, *Survey 160*

Text message surveys are an emerging method of survey research. As a mode of communication SMS (short message service) has different features from either web or phone surveys, although with some attributes in common with each. How should researchers design their survey instruments to best fit the specific affordances of this novel mode of survey research?

In this research, we examine that question with observational and experimental data from the 2020 election cycle. We look at several different aspects of questionnaire design, including the effect of survey length and signposting on completion rates, the use of identity confirmation questions to confirm a match to the age and gender of respondents as reported on a voter file, the optimization of introduction language to maximize participation, and other question and response choice wording design decisions.

To answer these questions about questionnaire design and describe new best practices for SMS interviewing, we draw on results from more than 100 of question-wording experiments and meta-data from over 200,000 interviews conducted in 2020.

Developing and Testing New Survey Content for the National Center for Education Statistics: Insight From the School Survey on Crime and Safety Redesign

Riley Burr, *American Institutes for Research*

Jana Kemp, *American Institutes for Research*

Zoe Padgett, *ICF*

Developing and testing new survey content for federal clients is a unique challenge. Multiple stakeholder perspectives must be addressed, respondent burden must be kept low, and the resulting data must be useful for policymakers, researchers, the media, and the public. Occasionally, federal priorities change during the development process, requiring adaptability. This presentation will focus on challenges faced and lessons learned in developing a question bank of school crime and safety-related items for the Department of Education's National Center for Education Statistics (NCES).

NCES administers the School Survey on Crime and Safety (SSOCS) every two years to a sample of approximately 4,800 U.S. public schools. Though small changes have been made to the survey since its first administration in 1999-2000, it has never undergone a full redesign. In 2017, the American Institutes for Research and NCES began the process of redesigning the survey to ensure it reflects the most salient current and emerging topics in school safety. At the onset, the intention was to redesign the standalone SSOCS questionnaire; however, as priorities at NCES changed, we transitioned the redesign efforts into the development of a bank of school crime and safety-related questions that could be used by the Department in multiple ways (e.g., a standalone survey, topic-specific modules, interspersed throughout another NCES survey).

This presentation will discuss the process of developing survey items for federal clients. We will discuss our research and literature review process, consultations with content-area experts, the development of an item outline, the questionnaire design best practices used to design the questions, and the process of conducting remote cognitive interviews with school principals during the COVID-19 pandemic. Additionally, the outbreak of the pandemic inspired a new set of pandemic-related questions. We will discuss how we expedited question development to integrate these newer items into our development process.

Examining the Measurement of Family Caregiving using a National Survey of Older Population: Implications for Survey Design and Analysis

Mengyao Hu, *University of Michigan*

Sarah E. Patterson, *University of Michigan*

Vicki Freedman, *Survey Research Center, University of Michigan*

Nora Lewis, *University of Michigan*

Survey research on family caregiving (FC) has traditionally focused on the measure of primary caregiver (PC), assuming that there is one caregiver provides substantially more care than others among a group of family caregivers. However, it is increasingly recognized that older adults typically rely on a network of caregivers. The type, quality and quantity of care divided by caregivers are complex and often change over time, requiring researchers to extend data collection and analysis beyond a single PC. Moreover, the definition of PC in the literature vary greatly. Some surveys directly ask the care recipient to identify the PC. Others use co-residence status, or duration and intensity of care provided for older adults to identify PC. To date, there is no consensus on the definition of PC and how to best measure FC structure.

This study aims to improve the measurement of FC structure in surveys by examining whether the concept of primary caregiver can correctly capture the dynamics of FC using data from two national surveys - the National Health and Aging Trends Study (NHATS), a longitudinal survey on older adults ages 65 and older, and the National Study of Caregiving, which collects data from up to five caregivers of NHATS respondents. We first examine to what extent the commonly-used identification methods based on co-residence or duration of care fail to identify PC and the level of inconsistency in identification across methods. We then evaluate the change over time of PC identification status (i.e., with a PC identified or not) and family caregiving structures, e.g., size of caregiver network, and composition of caregivers using growth mixture model and latent class growth analysis. Lastly, we compare the predictive validity of PC-centered compared to caregiving network-centered variables. This study has implications on the survey design and analysis to measure FC network.

The Confidentiality Gatekeeper: Strategies for Balancing Data Democratization and Respondent Data Security in a State Health Survey

Andrew Juhnke, *UCLA Center for Health Policy Research*

Parneet Ghuman, *UCLA Center for Health Policy Research*

Hin Wing Tse, *UCLA Center for Health Policy Research*

Data security is an integral part of all processes associated with the storage, dissemination, and overall management of survey data. Having robust review procedures (both on continual and periodic bases) is necessary in order to allow researchers to use and analyze the data in meaningful ways while also ensuring that the confidentiality of the survey's respondents is maintained.

As the largest statewide health survey in the nation, the California Health Interview Survey (CHIS) has a variety of data products, including public use data (public use files and AskCHIS query platforms), as well as confidential data projects, to benefit the public and researchers looking to utilize public health survey data. The Data Access Center (DAC) serves as a central gatekeeper of CHIS data products and is responsible for overseeing avenues for researchers to access CHIS data. The DAC maintains a rigorous process for classifying each CHIS variable according to sensitivity and identifiability, determines the level of availability of variables to the researchers, and conducts data disclosure review to ensure confidentiality and comply with data release protocols.

We examine (1) the role of the DAC, the Data Disclosure Advisory Committee (DDAC), and the Data Disclosure Review Committee (DDRC) when determining CHIS data inclusion in various data products; and (2) the various reviews of data (how does confidential data security differ from public use data), and the use of internal suppression guidelines (i.e. suppression thresholds). We will evaluate CHIS data security and disclosure review procedures, identifying both advantages and drawbacks for each strategy and how they compare to other surveys. Overall, we examine the balance between data security and data democratization to reveal potential ways in which CHIS can provide guidance to those in the survey data industry on best practices, and where improvements can be made within CHIS itself.

Development of “State to Local” Approach to Recruiting People With Intellectual and Developmental Disabilities for Cognitive Interviews

Amanda Titus, *National Center for Health Statistics*

Considerable obstacles while recruiting people with Intellectual and Developmental Disabilities (IDD) for a cognitive interview study led to the development and implementation of a “state to local” recruitment approach to reach individuals in the IDD community. Recruiting hard-to reach populations to participate in research is always a challenge because of the difficulty finding individuals with very specific characteristics and getting past multiple gatekeepers for vulnerable groups. This was the challenge faced by the Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER) during recruitment for an evaluation of survey questions designed to identify community-dwelling adults with IDD.

A “state to local” recruitment approach was developed by the CCQDER operations team. The approach begins with introductions to state-level representatives of organizations and groups which provided services and support to individuals with IDD and their families. These introductions laid the foundation to build relationships at the local level within the IDD community and ultimately our ability to reach respondents. While time intensive, the approach proved to be successful in recruiting individuals with IDD and their caregivers for cognitive interviews and could be modified and modeled for other studies involving hard-to-reach and vulnerable populations. Our presentation will detail our recruitment approach and how it was implemented to gain access to the IDD community.

Nonresponse to Income on the South Carolina (SC) Behavioral Risk Factor Surveillance System (BRFSS) and Associations With Demographic Groups

Jeremiah Bell, *South Carolina Department of Health and Environmental Control*

Harley T. Davis, *South Carolina Department of Health and Environmental Control*

Previous research has shown associations between nonresponse and age, sex, race, and education when utilizing random digit dialing (RDD) telephone surveys. Our objective was to investigate the association between nonresponse to an income question and demographics, and compare those associations between RDD cell phone and landline interviews within South Carolina’s population.

We utilized the 2019 SC BRFSS survey dataset. Income was dichotomized by response (any response vs. don’t know/not sure/refused) and stratified by cell phone and landline mode of completion. Associations were examined using chi square tests and logistic regression among age, sex, race, and education.

1245/7146 (17%) of respondents did not provide a response to the income question; of those interviews with a nonresponse, 527/1245 (42%) were landline and 718/1245 (58%) were cell phone. Among landline interviews there was a significant association between nonresponse and education ($p=0.0066$) but not with sex, age or race. Among cell phone interviews there were significant associations between nonresponse and age, race, and education (all $p=0.002$), but not with sex ($p=0.3452$). Among cell phone interviews, the 18-24 and 65+ age groups had the significantly higher odds of nonresponse compared to the other age groups in the 25-64 age range; minority racial/ethnic groups had significantly higher odds of nonresponse compared to Non-Hispanic whites. Among cell phone and landline interviews, individuals who did not graduate college had significantly higher odds of nonresponse compared to those who did graduate college.

In the 2019 SC BRFSS, nonresponse for income in cell phone interviews were associated with age, race, and education while nonresponse among landlines were only associated with education. Consideration of nonresponse variability among demographic groups in sample design and question development may allow for collection of more complete information from these groups, to include increases in cell phone completion percentages.

Interviewers and Interviewer Effects

Time After Time: Effects of Interview Setting, Interviewers, Respondents, and Questions on Response Times

Daniel Goldstein, *New York City Department of Housing Preservation & Development*

Elyzabeth Gaumer, *New York City Department of Housing Preservation & Development*

Questionnaire length is an important factor in survey cost and response rates. Reducing respondent burden can potentially result in higher data quality and even prevent broken off interviews. Advances in survey technology allow the measurement of overall interview length as well as precise capture of response times to individual questions. An emerging literature has examined the effects on response times of various question features as well as respondent and interviewer characteristics in a multilevel analytic framework in web (Yan and Tourangeau 2008), in-person (Couper and Kreuter 2013; Sturgis et al. 2020), and telephone surveys (Olson and Smyth 2015). This paper analyzes in-person interviews from the New York City Housing and Neighborhood Study (n=1610, AAPOR RR1=71.5%), a natural experiment evaluating the impact of affordable housing on overall wellbeing. Householder interviews (with those that applied to affordable housing with no children, n=933) took on average 60 minutes; caregiver interviews (with those that applied with children, n=677) took on average 90 minutes and included all householder questions plus additional caregiving and child-related questions. Interviews were conducted variously in respondents' homes, at project offices, and in public locations like coffee shops, libraries, or parks. This paper builds on previous work by examining the variability in response time due to interview setting in addition to interviewer and respondent characteristics and question features. We also explore what characteristics of each are associated with response time. Results will provide insight into which interview settings may be more encouraging to respondents as well as which question types and features may pose particular challenges to respondents and which types may be easier to process and answer. This analysis supports improvements in study design that may lower respondent burden and fatigue, produce higher data quality, and reduce break-offs.

Should We Use Conversational Agents and Voice-Based Input in Online Surveys?

Florian Berens, *University of Goettingen*

Sebastian Hobert, *University of Goettingen*

When collecting qualitative data, personal or telephone interviews provide the opportunity to obtain detailed answers from the respondents and to comprehensively assess complex issues. Surveys, on the other hand, whether on paper or online, create a higher degree of standardization and usually leave room for more questions than open interviews. Especially in online surveys, more people can be interviewed cost-effectively and easily. In the design of such surveys, however, the problem often arises that not all survey issues can be represented well in closed questions. Such surveys therefore also use open-ended questions, in which the respondents usually type their answer as free text into a box. However, studies and diverse experience in survey research show that many respondents shy away from this effort and therefore data quality in open questions is often rather poor. What is needed here are innovative approaches that reduce the effort for respondents or provide other incentives to motivate respondents to give better answers.

In a survey experiment with 2x2 design, first-year students were asked about their expectations of university and learning. The question was asked randomized either in a classical, static format or in form of a dialogue by a conversational agent. Depending on the second randomization, the respondents could either type the answer into a free field via keyboard or communicate it verbally via a microphone. The evaluation of the answers shows that verbally given answers are longer and indicate a higher answer quality in other parameters as well. The same applies to answers that were requested in a chat format. At the same time, however, it can be seen that verbal answers in particular increase drop-out considerably.

The Impact of Survey Design Features on Interviewer Effects and Deviant Behavior

Lukas Olbrich, *Institute for Employment Research*

Joseph W. Sakshaug, *Institute for Employment Research, LMU Munich, University of Mannheim*

Yuliya Kosyakova, *Institute for Employment Research, University of Mannheim, University of Bamberg*

Interviewers represent a well-known source of error in face-to-face surveys. Previous literature has found considerable interviewer effects on survey estimates and multiples studies have provided evidence on single interviewers who intentionally deviate from fieldwork instructions. However, it is still unclear to which extent specific survey design features facilitate interviewer effects and deviant interviewer behavior. In this study, we use data from eight rounds of the European Social Survey where countries differ in survey design features to explore this question. Specifically, we focus on differences with regard to sample selection methods (individual-level register, other register, random route) and the instrument mode (CAPI, PAPI). Based on principal-agent theory, we develop a theoretical framework for interviewer behavior where the field agency's ability to monitor interviewers and the interviewers' variety of tasks play a crucial role in enhancing interviewer effects and deviant behavior. To test the hypotheses on the design features derived from this framework, we use a two-step approach. First, we employ multiple indicators of data quality as dependent variables of multilevel location-scale models estimated for each country in each round separately. These models allow for constructing four distinct measures of interviewer behavior for each estimation and for approximately 200 country-round observations. In the second step, we use this sample to estimate to which extent sample selection methods and the instrument mode are associated with the interviewer measures. The results will show under which circumstances both survey practitioners and researchers must pay particular attention to interviewer behavior.

The Effect of Interviewer Speaking Pace on Respondent Behaviors in Telephone Interviews

Angelica Phillips, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

Jolene Smyth, *University of Nebraska-Lincoln*

Telephone interviewers are typically trained to speak at a pace of two words-per-second to enhance respondent cognitive processing. Although interviewer speaking pace varies across different question characteristics such as question length and complexity (Phillips 2020; Olson, Smyth, and Ganshert 2019), whether the pace at which interviewers ask questions is associated with subsequent respondent speaking behaviors has received scant attention. One might hypothesize that questions that are asked more quickly will be less likely to be understood and thus receive more requests for clarification and uncodable answers, but be less likely to be interrupted due to the fast pace. On the other hand, interviewers who ask questions more quickly may be communicating the need for a fast adequate response and may reduce the amount of time that the question needs to be held in working memory, decreasing uncodable answers.

This paper examines the association between question-level interviewer speaking pace and respondent behaviors among the first two conversational turns in telephone interviews using behavior coded transcripts from the Work and Leisure Today 2 Survey (AAPOR RR3=7.1%). On average, interviewers speak at a pace of 3.15 (n=42272, SD=0.88) words-per-second, and 46.0% of interviewers speak at a pace of 3 words-per-second or slower. We dichotomize interviewer speaking pace into less than or equal to three words-per-second or greater than three words-per-second, and examine whether this indicator is associated with respondent behaviors such as requests for clarification and uncodable responses. We additionally examine whether question characteristics such as question length and question complexity moderate the association between interviewer speaking pace and respondent behaviors. Preliminary results indicate that questions asked more quickly have lower rates of uncodable responses (log odds=-0.078, p=0.034), but higher rates of respondent requests for clarification (log odds=0.097, p=0.079). We conclude with implications for interviewer training.

Exploring Interviewer Impressions and Interaction Behaviors in the Survey of Income and Program Participation

Rodney Terry, *U.S. Census Bureau*

Erica Yu, *U.S. Bureau of Labor Statistics*

Holly Fee, *U.S. Census Bureau*

Alina Kline, *U.S. Census Bureau*

Robin Kaplan, *U.S. Bureau of Labor Statistics*

Interviewers play a crucial role in the outcome of an interview, and information about the doorstep interaction from the interviewer's perspective provides insight into their mindset while administering the interview. Typically recorded after the fact via written notes or electronically, the interviewer's early impressions of the physical interview environment and respondent doorstep behavior (which together form the interview "context") are likely being formed starting with the interviewer's first contact attempt. These impressions influence the interviewer's behavior throughout the contact, including how interviewers control the interview interaction and why interviewers deviate from scripted survey questions, particularly when probing respondents for questions that may be cognitively burdensome or sensitive. This study explores the relationship between interviewer impressions and respondent behavior using several data sources from the 2014 SIPP Panel. These data sources include computer audio-recorded interview (CARI) recordings of the interview itself; interviewer perceptions of the physical interview environment (Neighborhood Observation Instrument, NOI) and respondent doorstep behaviors (Contact History Instrument, CHI); respondent demographics; and interviewer characteristics. To create data of interviewer and respondent behavior, three researchers independently transcribed and behavior coded CARI recordings of the full interviewer-respondent interaction for a sample of targeted questions. Behavior codes included whether the interviewer changed the survey question and type of change made, whether the respondent gave a codeable response, and how the interviewer reacted to the response. We present descriptive results for the targeted SIPP questions, and use data from the various sources to model the relationship between (a) interviewer impressions (using data from the CHI and NOI) and (b) interviewer and respondent behaviors (using data from SIPP CARI recordings).

Impact of Rapport and Interviewer Professionalism on Disclosure of Sensitive Information in a Victimization Study

Alexis Kokoska, *Westat*

Ting Yan, *Westat*

David Cantor, *Westat*

Hanyu Sun, *Westat*

Interviewers are often encouraged to establish rapport with respondents, in part because it might motivate respondents to provide higher quality data. Indeed, some research has shown that interviewers who display rapport-promoting behaviors receive more informative and detailed answers to survey items. However, high rapport might not be beneficial for data quality when a respondent is asked a sensitive item, as a respondent might not wish to risk an established positive relationship with the interviewer by disclosing information that could be uncomfortable, awkward, or present the respondent in an unfavorable light. A recent study by Sun, Conrad, and Kreuter (2019) found via a laboratory experiment that high rapport increased disclosure for highly sensitive survey items but not for moderately sensitive items, contrary to expectations. For the current presentation, we add to this research by continuing to investigate the relationship between rapport and disclosure of sensitive information but with data that comes from a national survey. Further, we will explore the effect of the perceived professionalism of an interviewer on disclosure of sensitive information. Interviewers who display a high degree of professionalism might have a more formal and less personal interviewing style, which could impact disclosure in a different way compared to interviews with high rapport. However, it is not clear that a high degree of professionalism and high rapport are mutually exclusive, and we plan to investigate this relationship as it relates to disclosure. We use data from a nationally representative crime victimization survey, which included both moderately and highly sensitive items. Trained coders listened to the recordings of these in-person interviews and rated the interviewer-respondent rapport and interviewer professionalism. For our analysis, we use the coders' ratings of rapport and interviewer professionalism to examine the impact of both factors on the answers respondents provided to moderately and highly sensitive items.

5/13/2021, 10:00 AM - 11:30 AM
Concurrent Session E: Live Sessions

**Mini-Conference: Data Collection in Developing and Middle-Income Countries
During and After COVID-19**

Data Collection in Developing and Middle-Income Countries During and After COVID-19
Matthew "Clark" Letterman, *Pew Research Center*

This panel brings together international research practitioners with extensive experience conducting public opinion polls, impact evaluations, and other forms of population-based research in developing and middle-income countries around the globe. The discussion will focus on how the COVID-19 pandemic has radically transformed the environment in which these practitioners work, and how researchers are using innovative technologies and methodological approaches to continue to collect data during this turbulent time.

Specific topics that will be explored include; effects of the pandemic on on-going work, shifting to new modes of data collection (CATI, IVR, SMS), project management and supervision of dispersed workforces, ethical considerations for data collection during a pandemic, and the ways that the pandemic may be acting as a catalyst for long-term change in the research methods employed in developing and middle-income countries.

The panel includes presenters from Gallup, RTI International, D3 Systems and GeoPoll. It is moderated by Clark Letterman, Senior Survey Manager for International Research at the Pew Research Center.

Global Polling in the Time of Corona
Rajesh Srinivasan, *Gallup*

Since 2005, Gallup has conducted nationally representative surveys in more than 140 countries annually as part of its' World Poll initiative. In addition, Gallup has also been involved in several custom client-related global studies. Even in the best of conditions, conducting these global surveys with strict adherence to research protocols and timelines is a significant challenge. The uncertainty surrounding the COVID-19 pandemic has presented us with some unprecedented challenges in carrying out our global surveys. However, our mission to keep listening and reporting on the thoughts, feelings and opinions of people worldwide has never been more important. The questions we ask about people's health, wellbeing, and food security are ones that leaders and policymakers desperately need the answers to.

With over 100 of the 140 countries historically collecting data using face-to-face methods, finding a solution that is methodologically rigorous, has the least impact on historical trends, is cost effective, within the reach of the skillset and capacity of our partner network, and can be executed within prescribed timelines is a monumental challenge. Over the past five months, the Gallup World Poll team has worked diligently to come up with a set of solutions that address these challenges. Beginning in September with the rollout of the new initiatives, there are significant learnings around every aspect of the data collection and measurement process. Challenges and solutions we will specifically address as part of this presentation cover infrastructural/logistical, recruitment/training and motivation, monitoring performance and productivity, impact of COVID on substantive measures, such as employment and health behaviors, and what potential modifications to the process may be required to enable a more robust execution in 2021 given the continued uncertainty around face-to-face data collection next year.

Decentralized Agent CATI for International Research

Matthew Warshaw, *D3 – Designs Data Decisions*

Danielle Cuddington, *D3 – Designs, Data, Decisions*

D3 has a history of leveraging advances in technology to improve or supplement operationalizing research methodology in traditionally insecure and/or data poor environments. For example, D3 invests in the continued development of its own CAPI software Research Control Solutions (RCS) to enable customized data collection, sample implementation, and quality control for a variety of clients with specific requests. Most recently, with the global pandemic, we have experimented with several methods that focus on CATI but in a decentralized agent setting, allowing interviewers to work from the safety of their homes. We will share some of the lessons learned through these experiments in this presentation, in particular an efforts to connect interviewers from their homes in Afghanistan, Kenya, and Tunisia.

Examining Mobile Modes for Survey Research in Africa, Asia, and Latin America

Caitlin van Niekerk, *GeoPoll*

Roxana Elliott, *GeoPoll*

Remote, mobile-based survey methodologies including self-administered modes such as 2-way SMS, and interviewer-administered modes including Computer Assisted Telephone Interviewing (CATI) and Interactive Voice Response (IVR), have been available in low and middle-income countries for several years. However, while mobile-based surveys had been used for crisis situations (Hoogeveen & Ndip, 2017), they had not become widely adopted as a viable alternative to in-person research when COVID-19 dramatically shifted the research landscape in countries throughout Africa, Asia, and Latin America. In March 2020, COVID-19 related travel restrictions and safety precautions forced in-person data collection to largely cease, and left organizations searching for solutions they could quickly adapt to in order to continue research projects. Many of these organizations were faced with questions they had not previously considered, such as: How to adapt face-to-face questionnaires to voice call methodologies; what biases are seen in surveys administered to mobile-owning populations; how to create sample frames for remote surveys; and how does survey length impact response rates for remote survey methodologies.

There is limited research on these topics and studies that have been done have sometimes found contradictory results, especially when considering research for multiple countries: For example, one study found that IVR was best at reaching less educated populations, but CATI had higher response rates and reached older age groups (Lau et al, 2019). In this paper, mobile survey provider GeoPoll will review its own findings around response rates, questionnaire length, and question types from studies conducted via multiple mobile modes in low and middle-income studies both before and during the coronavirus outbreak. We will demonstrate the challenges of shifting from face-to-face to remote research without design changes, and discuss how researchers can leverage this short-term shift to remote methodologies to increase the knowledgebase around how to most effectively conduct research remotely.

Data for Policymaking on COVID-19: A Case Study of a CATI Survey in Bolivia

Charles Lau, *RTI International*

In Latin America and the Caribbean, public health data are based largely on face-to-face surveys and administrative records. The inability to conduct face-to-face surveys during COVID-19 presents challenges to governments and international health organizations, who need representative and timely data on COVID-19 and its impacts. Governments have a variety of information needs about COVID-19. One need is to understand how COVID-19 is affecting mental health and access to health services for non-communicable diseases such as diabetes and hypertension.

In August 2020, we conducted a computer-assisted telephone interviewing (CATI) survey of the general adult population in Bolivia over a 10-day period. We used a random-digit dialing (RDD) sample, stratified by mobile network operator and geography, to conduct interviews with 1,200 people age 18 and over in Bolivia. The 15-minute questionnaire was developed by the Bolivian Ministry of Health, the Pan-American Health Organization, the U.S. Centers for Disease Control and Prevention, and RTI International. The questionnaire asked about behaviors to mitigate COVID-19 risk, risk communications, and access to health services. Training and data collection were remote, with the 22 interviewers placing calls from their homes. In this presentation, we will review the methodology and emerging best practices from this study. There are three parts. First, we will review the response rate and representativeness of the survey by comparing the survey data to benchmark data sources from Bolivia. Second, we will review operational best practices in quickly pivoting to CATI – in particular, how to train and monitor data collection quality remotely. Third, we will discuss best practices in adopting a consortium approach (comprised of governments, international health organizations, and survey researcher implementers) to conduct COVID-19 surveys. Our presentation will draw from our project experience, but speculate about potentially longer-lasting changes in survey research due to COVID-19.

Mini-Conference: Spending and Payment Use During the Pandemic: How Survey Data Can Inform Policy

Spending and Payment Use During the Pandemic: How Survey Data Can Inform Policy

Gradon Nicholls, *Bank of Canada*

The COVID-19 pandemic has had an impact on the way people shop and how they pay. Early in the pandemic, many businesses closed in-store operations and inessential shopping fell, suggesting that use of all methods of payment such as cash and payment cards may have declined. Further, as the point of sale moved online, there may have been a shift away from cash and towards electronic means of payment. On the merchant side, anecdotal evidence from news reports suggested that some merchants stopped accepting cash as a method of payment due to the risk of transmitting the virus. Despite these factors, we find an increase in the total amount of cash in circulation in many countries relative to previous years. This panel brings together researchers from central banks and other institutions around the globe to provide insight into spending and payment use in Canada, Japan, and the U.S using survey data from each country. We study the consumer perspective on payments to see how the pandemic has affected the demand and use of cash, payment cards, and other electronic forms of payment. We show how timely survey data can be an important monitoring tool for policy work, and that it is imperative to continue monitoring as pandemic conditions evolve.

Changes in Consumer Spending Behaviors During the Coronavirus Pandemic

Adam Safir, *U.S. Bureau of Labor Statistics*

Thesia Garner, *Bureau of Labor Statistics*

Jake Schild, *Bureau of Labor Statistics*

Measuring the social, health, and economic impact of the coronavirus pandemic, along with policy efforts to address those effects, is of critical national importance. Responding to this need, several federal agencies, including the Bureau of Labor Statistics (BLS), developed questions for the rapid response Household Pulse Survey (HPS). The survey includes questions on respondent demographics, employment, food security, health, housing, education, and spending. During the first phase of the survey, which was fielded from April 23 through July 21, 2020, the BLS contributed questions related to the receipt and actual or expected use of Economic Impact Payments (EIPs), as well as sources of income being used to meet spending needs during the pandemic. For the second phase of the survey, which began on August 19, 2020, the BLS questions shifted focus from EIP use, to the potential long term impacts of the coronavirus pandemic and the related policies that influence consumer buying behavior. While many consumer behavioral changes throughout the pandemic have been protective, aligning with avoidance strategies, others were more relaxed, as the duration of the pandemic lengthened. We present findings from the HPS with a focus on analyzing the variations in spending behavior over time, and in particular on pandemic avoidance strategies such as increasing the use of credit cards or smartphone apps for purchases, instead of using cash. The effects of the coronavirus pandemic on spending behaviors appear to depend heavily on individual characteristics and the local pandemic response.

COVID-19 and Cash Use in the United States

Kevin Foster, *Federal Reserve Bank of Atlanta*

During the early weeks of the coronavirus pandemic, there were fears that handling cash could help spread the disease. These concerns were picked up in the national media, which caused the ideas to spread amongst the population. The Federal Reserve Bank of Atlanta added questions to the Understanding America Study Coronavirus Tracking Survey to find out if US consumers changed their cash behavior because of coronavirus. The survey results reveal that aspects of consumer cash activity during the early pandemic align closely with similar behavior observed before major hurricanes in the US. In addition, because the Coronavirus Tracking Survey contains a wealth of additional information on panelists' labor force status and household financial situation, we are able to conclude that enhanced unemployment benefits worked as intended, and put cash into the pockets of those who needed it most.

Digitalization and Changes in Financial Services for Households: Evidence From the Financial Radar 2020

Hiroshi Fujiki, *Chuo University*

We examined the effects of COVID-19 infection on the choice of consumer financial service in Japan. We used the Financial RADAR 2020 special survey conducted at the end of June 2020, which covers 3,026 men and women aged 20-74 years within a 40 km radius of the Tokyo metropolitan area, Japan. Financial RADAR2020 asks respondents on two points. First, whether respondents' preferences for face-to-face and non-face-to-face financial services have changed since COVID-19 occurred. Second, what financial services did respondents want to increase or decrease their use of after COVID-19 occurred? Financial RADAR 2020 also contains questions to identify the demographic characteristics of respondents as well as financial literacy. Using the response to these questions, we run multinomial probit models and obtained the two results below. First, the request for social distance following the COVID-19 outbreak has led to the desire of around 15% of respondents to use non-face-to-face financial services. Respondents with a higher probability of preferring to use non-face-to-face financial services tend to have a higher amount of financial assets, are younger, more financially literate, and do not have an increase in total annual income and spending, including taxes, compared to the previous year. Second, the services that respondents wanted to increase their use of after the spread of COVID-19 infection included cashless payments, stocks, term and fixed-term savings accounts, and mutual funds. Respondents who were more likely to intend to increase their use of cashless payments tended to have a higher amount of financial assets and were more financially literate.

Cash and COVID-19: The Effect of the Pandemic on Cash Demand and Use in Canada

Gradon Nicholls, *Bank of Canada*

Heng Chen, *Bank of Canada*

Walter Engert, *Bank of Canada*

Kim P Huynh, *Bank of Canada*

Julia Zhu, *Bank of Canada*

A recent Bank of Canada study (Chen et al 2020), drawing on data from the Bank Note Distribution System (BNDS) and a consumer survey, analyzed the impact of the COVID-19 on cash demand and methods of payment during the first months of the pandemic (April 2020). This paper updates that analysis using more recent data from the BNDS and from a follow-up survey conducted in July 2020. That survey coincided with the relaxation of lockdown protocols and so provides insight into how the use of cash and digital payments changed as businesses reopened and as people began to move around more freely. The Bank of Canada will continue to monitor these developments, with additional surveys in late 2020 and Spring 2021.

Mini-Conference: Design Changes of In-Person Surveys During the Pandemic

Design Changes of In-Person Surveys During the Pandemic

Andy Peytchev, *RTI International*

Surveys with in-person data collection have been impacted in a severe way due to the nature of the COVID-19 pandemic. Field data collections have been stopped for some surveys indefinitely, temporarily halted to redesign the data collection to include other modes, continued seamlessly with available alternative modes, or involved a combination of the latter two. This panel includes presentations on four large-scale national surveys with in-person data collection that have adapted in very different ways based on their current designs, objectives, and identified available options. These adaptations include changes in sampling frames, modes, data collection procedures to optimize for modes other than face to face, incorporation of in-person interviewing that is limited to select areas, and modifications to operations.

Panel presentations: National Health Interview Survey (Stephen Blumberg) National Survey on Drug Use and Health (Doug Currivan) Current Expenditure Survey (Yezzi Angi Lee) Mental and Substance Use Disorders Prevalence Study (Heidi Guyer)

Impact of the Pandemic on National Health Interview Survey Data Collection

Stephen Blumberg, *National Center for Health Statistics*

The National Health Interview Survey is the longest-running household-based health survey in the US, and one of the National Center for Health Statistics' primary sources of data on the health of the US population. NHIS is an in-person household survey of the civilian noninstitutionalized population, with interviews conducted by Census Bureau field representatives. Like all in-person surveys, NHIS was impacted by the COVID-19 pandemic. However, NHIS never stopped collecting data. On March 19, 2020, NHIS temporarily became a telephone survey. Where possible, sample addresses were matched to telephone numbers using commercial lists and additional searches. Response rates declined from 60% to 40%, and the sample skewed toward older and more affluent households. Personal visits to households resumed in selected areas in July and in all areas of the country in September. However, cases were still attempted by telephone first. Personal visits were used only to follow-up on nonresponse, to deliver recruitment materials, and to conduct interviews when telephone numbers were unknown. Response rates returned nearly to normal in September.

Nevertheless, NCHS chose to also field the 2020 NHIS questionnaire with a parallel sample that had known representativeness and nearly complete telephone contact information. That is, starting in August, about 20,000 adult respondents who completed the NHIS in 2019 were recontacted by phone and asked to participate again. These follow-up interviews will also permit NHIS to look at health, health care, and well-being from before and during the pandemic, for the same individuals. As a result, NHIS fielded four designs in 2020: Normal production in Quarter 1, telephone-only in Quarter 2, telephone-first in Quarters 3 and 4, and the longitudinal follow-up. Ongoing challenges include how to use weighting and estimation techniques to produce official 2020 estimates from the disparate pieces, each with its own coverage and nonresponse issues.

Changing Modes During COVID-19 on the National Survey on Drug Use and Health

Doug Currivan, *RTI International*

Chris Jewett, *RTI International*

Megan Livengood, *RTI International*

Gretchen McHenry, *RTI International*

Allison McKamey, *RTI International*

Katie Morton, *RTI International*

Scott Payne, *RTI International*

Gilbert Rodriguez, *RTI International*

Nan Ganapathi, *RTI International*

The National Survey on Drug Use and Health (NSDUH) provides national estimates of substance use and mental health among the civilian, noninstitutionalized population aged 12 or older in the United States. NSDUH has collected data in-person only since its initiation in 1971. In March 2020 in-person data collection was suspended in all states, due to rapid increases in COVID-19 infection rates in multiple states. With a goal to collect as much data in-person as feasible in 2020, the NSDUH team began simultaneously planning both returning to safe areas for in-person data collection and developing a web mode option for sample units in areas that did not meet safety standards for in-person data collection. The team monitored infection rate metrics closely to determine the sample units in areas likely to be safe for in-person data collection on October 1 and the number of sample units needed in areas unlikely to be safe where the web option would be offered. This presentation will discuss key considerations in planning simultaneous in-person and web data collection in different areas, challenges in adapting in-person protocols and instruments for web data collection, significant changes to preparations for data collection, and lessons learned along the way in each of these realms. Results will include operational outcomes such as web and face to face survey response rates to the household screening and individual interviews.

Consumer Expenditure Survey Response to COVID Pandemic

Yezzi Lee, *U.S. Bureau of Labor Statistics*

David Biagas, *U.S. Bureau of Labor Statistics*

Barry Steinberg, *U.S. Bureau of Labor Statistics*

Jared Ogden, *U.S. Bureau of Labor Statistics*

Sharon Pham, *U.S. Bureau of Labor Statistics*

The Bureau of Labor Statistics' (BLS) Consumer Expenditure Interview and Diary Surveys (CE), administered on behalf of BLS by the Census Bureau, are designed to collect data primarily through personal visits. Although telephone interviews are allowed in some cases to accommodate the respondents' situation and to avoid refusals, interviews collected by telephone are generally limited to ensure better data quality. In response to the COVID-19 pandemic, the Census Bureau stopped all personal visits for the health and safety of both interviewers and respondents in late March 2020 and for the CE surveys shifted to 100 percent telephone interviewing, adopted additional protocol changes for the Diary Survey, and suspended the mailing of advance letters for a period of time. In addition, CE provided the option to use an online diary to respondents in place of the existing paper diaries in the beginning of July 2020. BLS worked closely with the Census Bureau to modify protocols for the Interview and Diary surveys and provide guidance to field interviewers. BLS also needed to reevaluate some of its existing processes in order to reclassify the eligibility (for cases of unknown eligibility) and participation status of households in order to better calculate accurate response rates and measure spending during the pandemic. This presentation will provide an overview of the protocol changes to the interview and diary surveys and describe the processing interventions needed to maintain the best possible quality of data collected during the pandemic. It will further discuss ongoing analyses on the data quality.

The Mental and Substance Use Disorders Prevalence Study

Heidi Guyer, *RTI International*

Andy Peytchev, *RTI International*

Heather Ringeisen, *RTI International*

The Mental and Substance Use Disorders Prevalence Study (MDPS) is a large-scale pilot program designed to estimate the prevalence of mental and substance use disorders among U.S. adults ages 18 to 65. Mental and substance use disorders of interest include schizophrenia, bipolar disorder, depression, and post-traumatic stress disorder. The design includes four study populations: the household population, the prison population, homeless population in shelters, and patients in state psychiatric hospitals. The design is tailored to each population. Due to the need for skilled clinical interviewers and the low prevalence rate of certain conditions in the household population, the design for the household population—by far the largest of the four—includes three stages: household rostering, mental health screening, and clinical interviews. The mental health screening is instrumental in this approach in order to classify sample members into risk strata and increase the efficiency of the clinical interview stage. A key responsive design feature is a test of two different screening instruments—one based on a traditional survey instrument with rule-based scoring and another based on computerized adaptive testing with interactive item generation and scoring—to inform later sample releases and potential future studies. Each stage of data collection includes a mixed mode design. Due to COVID-19, the mixed mode design was substantially revised to reduce the reliance on in-person data collection, including augmentation with contact information, adding further modes, and changes to contacting sample members. The focus of this presentation is threefold: to describe how the design was altered to address COVID-19 implications, how data collection procedures such as training were adapted in this context, and to present completion rates by mode and design stage based on interim data that may be of benefit to other studies.

Sirken Award Lecture in Interdisciplinary Survey Research Methods Research: One Size (or Mode) Does Not Fit All

One Size (or Mode) Does Not Fit All

Edith de Leeuw, Utrecht University

Different modes of data collection have been 'in fashion' in different time periods. For instance, telephone surveys were seen as the heir apparent to face-to-face in the 1990ies. In the new millennium, online surveys and online panels became more and more the thing to do. In cross-national and cross-cultural research, survey experts soon realized that one mode did not fit all, and that much depended on differences in (telephone or internet) coverage, sampling frames, and social customs. Within countries, reducing nonresponse and limiting survey costs are the two most important reasons to employ multiple modes in survey research. Multiple modes of contact and multiple modes of response may be combined into one mixed-mode survey design to reach these goals.

The present COVID-19 pandemic changed and will continue to change survey research, in particular the data collection modes employed to conduct surveys. Social distancing implies limiting interviewer impact and more creative use of self-administration modes. Mixed mode surveys are becoming necessary in the current situation. Especially mixed mode surveys that include an online component offer many advantages, but also challenges. One of the challenges is the use of mixed-devices (smartphone, tablet) to complete an online questionnaire. Other challenges are an increased administrative and logistic burden, and especially a potential for mode specific measurement error.

From a Total Survey Error perspective, one wants the best of all worlds and to reduce overall error. I will discuss the major steps (i.e., design, diagnosis, analysis) necessary in high quality mixed mode studies.

5/13/2021, 1:00 PM - 3:00 PM
Concurrent Session F: Prerecorded Sessions

Mini-Conference: Measuring the Impact of COVID-19 on Consumers and Businesses

Understanding the Intersection of Grocery and Sustainability During a Pandemic

Hannah Borenstein, *Murphy Research*

Maggie Bright, *Murphy Research*

Stacy McCoy, *Murphy Research*

In 2019, the Coca Cola Retailing Research Council (CCRRC) prioritized understanding sustainability through the lens of grocery consumers. They partnered with Murphy Research to launch a multiphase research study focused on understanding how consumers contextualize and engage with sustainability and the associated expectations from grocery retailers and manufacturers. The spread of COVID-19 altered the social contract between consumers and retailers, reflecting a new reality in which local grocery is paramount to daily life. As such, the CCRRC study objectives expanded to include the role of grocery during COVID-19 and explore the likelihood of long-term behavioral changes to shopping habits and their associated impact on sustainability.

In addition to conducting an environmental scan, consumer research and activation sessions were conducted, including the following steps:

- Step 1: Quantitative tracker with a baseline wave of n=3,500 US consumers, additional biweekly pulse checks of n=300 over 8 weeks, and n=45 qualitative in-depth interviews
- Step 2: Development of activation strategies by robust council of retailers to address consumer receptivity to grocery-led sustainability innovation
- Step 3: Consumer validation and prioritization of strategies through quantitative research leveraging a tournament design among n=1,500 consumers

Initial findings indicate consumers are appreciative of grocery retailers and their ability to adapt to COVID related challenges. However, the pandemic has made grocery shopping more stressful. Despite deep division across the country on many issues, consumers align on the role of grocery; it should make a profit and be a force for good within the community where it operates. Consumers think about the environment while shopping, but most would like to do better. They welcome additional help from grocery retailers, especially around recycling, offering local products, and ending food insecurity. Currently in field, Step 3 will provide a roadmap to retailers for prioritizing solutions around local community empowerment and sustainability.

Time for Me to Fly...in Late 2021: A Review of Travel Attitudes During the COVID-19 Pandemic

James Dayton, *ICF*

Ned Shugrue, *ICF*

Thomas Brassell, *ICF*

Guy Cierzan, *ICF*

The COVID-19 pandemic has had a significant health and economic impact throughout the United States, with the death toll exceeding 230,000 at the beginning of November 2020 and over 11 million people unemployed as of October 2020. In particular, the airline industry has been hard hit by the pandemic, with several major airlines recently announcing tens of thousands of layoffs due to reduced travel. Starting in March, we administered a national survey using a Census-balanced non-probability mobile panel to collect 1,000 completed surveys a month. In May we added questions on respondent travel. This current study examines frequent traveler attitudes and behaviors towards traveling during the pandemic since May. It is unique in that it analyses attitudes that pertain to a particular industry heavily impacted by COVID-19.

The effects of COVID-19 are likely to be long-lasting in the airline industry as a growing proportion of travelers look to push off flying. We found that while in May, frequent travelers showed cautious optimism for future air travel in 2020, by September these feelings had shifted substantially. Specifically, in May, 42% of frequent travelers expressed a willingness to travel domestically by air in 2020. However, in October, this percent had dropped to 37%. More concerning for the travel and hospitality industry, the percent of frequent travelers not willing to travel domestically until later than summer 2021 grew from 22% in May to 31% in October. Additional overlaps between COVID-19 and travel attitudes will be discussed.

Consumer Response to the Economic Stimulus Payment and Life in the Time of COVID-19

Jake Schild, *U.S. Bureau of Labor Statistics*

Thesia Garner, *Bureau of Labor Statistics*

Adam Safir, *Bureau of Labor Statistics*

Measuring the social, health, and economic impact of the coronavirus pandemic, along with policy efforts to address those effects, is of critical national importance. Responding to this need, several federal agencies, including the Bureau of Labor Statistics (BLS), developed questions for the rapid response Household Pulse Survey (HPS). The survey includes questions on respondent demographics, employment, food security, health, housing, education, and spending. During the first phase of the survey, which was fielded from April 23 through July 21, 2020, the BLS contributed questions related to the receipt and actual or expected use of Economic Impact Payments (EIP), as well as sources of income used to meet spending needs during the pandemic. Responses to the HPS during Week 7 of the first phase, July 11-16, revealed 84 percent of respondents received or expected to receive an EIP, with a majority indicating they would use the EIP for expenses. However, there was variation in reported use across generations. Generation X reported the highest rate of using the EIP for expenses (74 percent of respondents in this group), whereas the Silent generation had the highest rate of savings (27 percent of respondents in this group). For the second phase of the survey, which began on August 19, 2020, the BLS questions shifted focus from the use of the EIP, to the potential long term impacts of the coronavirus pandemic and the related policies that influence consumer buying behavior. Findings from the first week of the second phase, August 19-31, showed respondents participating in both relaxing and restrictive behaviors; however, an overwhelming majority of the reported changes in behavior were restrictive. Results from this phase also show significant variation in behavior changes across generations and states. Findings from the HPS suggest the effects of the EIPs depend heavily individual characteristics and geographic location.

Assessing The Potential of Online Deliberation During COVID-19: Findings From the HK and Tokyo Deliberative Pollings on Solar Futures

Victor Lam, *University of British Columbia*

Daphne Ngar-Yin Mah, *Hong Kong Baptist University*

Darren Man-wai Cheung, *Hong Kong Baptist University*

Alice Siu, *Stanford University*

Benjamin McLellan, *Kyoto University*

Shinya Wakao, *College of the Mainland*

Catherine Luo, *Stanford University*

The COVID-19 reality and recovery present significant logical and public health challenges that continue to inhibit organization of conventional, in-person public engagements. Once considered an alternative to traditional modes of engagement, online deliberative participatory methods have been rapidly adopted by policy-makers to sustain public engagement during the COVID-19 pandemic. While new online forms of participatory mechanisms have gained increasing scholarly and policy attention, empirical research has been lacking particularly in evaluating the effectiveness of these new approaches. This paper seeks to partially fill this gap by evaluating the effectiveness of online Deliberative Polls (DPs) in the fields of sustainable energy transitions. We drew insights from two DPs on solar futures conducted in Tokyo and Hong Kong. Based on data derived from 136 Tokyo and 174 Hong Kong citizens, this paper preliminarily finds the usefulness of DPs in enhancing representativeness and inclusiveness of deliberative events under COVID-19. The DPs facilitated dialogues among participants in small-group settings which were essential to participant formation of their considered views on solar futures. The DPs also effectively

combined quantitative (derived from pre- and post-workshop questionnaires) and qualitative data (derived from transcriptions of DP discussions) that integrated approaches to understand the various themes arising from engaged citizens' deliberations in sustainable energy transitions. Moreover, while the DPs offered potential for effective deliberative mechanisms for engaging citizens on complex matters which are contentious in nature, they required careful project design, planning, execution, evaluation, and other practical considerations.

A New Ship, for a New Sea, to Anchor in a New World of B2B Research

Cynthia Miller, *Braun Research Inc*

A host of reasons exists as to why marketers measure opinions of CSuiters, small business owners, healthcare executives/providers, and any number of others occupying industry space. Until early 2020 it was common knowledge that reaching B2B respondents required a significantly longer field period – regardless of methodology – than would trying to reach a GenPop audience. And this is within an 'all things being equal' scenario, not taking account of specific screening criteria. An online methodology is helpful with B2B respondents, and we had always experienced some distinct challenges whether via a panel or client-provided sample (with emails). A phone study with B2B respondents was understandably more difficult than was online given that interviewers needed a week to build awareness and familiarity, as well as develop and massage relationships with gatekeepers and respondents. With the ebb and flow of the ever-present pandemic and its challenges, researchers must rechart a 'new world' of approaches to navigate the seas and spot land to anchor B2B completes. A typical phone schedule of five weeks must be doubled. An online schedule, while apparently preferable with many working from home, requires its own plan. B2B respondents have new priorities, and it does not necessarily follow that B2B respondents will answer a survey via a link. Online schedules for a project must be tripled or even quadrupled. This can be supplemented with a LinkedIn effort, but even that endeavor falls into a black hole. How do we best reach B2B respondents in the 'new normal?' This paper will look at these approaches and ideas for success, reached via a larger, more strategically-built ship, on calmer oceans. The information provided in this paper/talk will hope to benefit others who need to run B2B studies – whether qualitative or quantitative, international or domestic – in this new world.

Mini-Conference: It's All About Timing: Lessons Learned From Adapting to COVID-19 at Various Points in Federal Data Collections

It's All About Timing: Lessons Learned From Adapting to COVID-19 at Various Points in Federal Data Collections

Carolyn Hronis, *Energy Information Administration*

This panel will explore the effect of the COVID-19 pandemic on federal data collections at various stages in the survey cycle. Surveys that were still in the planning phase had to decide whether to make changes to the questionnaire design or data collection procedures, which could entail changes to the mode, timing of data collection, or changes in questionnaire content or wording. For surveys that were already in the field, agencies had to determine how to manage changes in mode from in-person to other modes that could be administered while adhering to social distancing guidelines and stay-at-home orders. To make such decisions, agencies had to predict the impact of these changes on response rates, data quality, and costs. This panel assembles presentations about four different data collections fielded during 2020. These data collections represent a diverse set of federal surveys of household, student, and professional populations. The speakers will provide insight into practical considerations, agency decision-making processes, and how project teams anticipated and mitigated potential risks. Panelists will share results and lessons learned from designing and implementing data collections during these unprecedented times.

Working and Schooling at Home: How Does it Impact Household Survey Response?

Danielle Mayclin, *Westat*

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

The Residential Energy Consumption Survey (RECS) is a national household survey about home energy characteristics and usage conducted by the U.S. Energy Information Administration (EIA). When the coronavirus pandemic emerged in early 2020, EIA was in the midst of finalizing the questionnaire content and data collection design for the 2020 RECS. A late decision was made to add a few new items about teleworking and online education to the questionnaire, since the increased time Americans were spending at home for such activities was expected to result in increased home energy use. Otherwise, the 2020 RECS proceeded as planned with data collection starting in September 2020, relying on a mixed-mode Choice+ protocol (Biemer, et al. 2017) with web and mail survey modes, as well as a nonresponse follow-up phase. This analysis will focus on survey response patterns for households that reported various levels of teleworking and/or online education. Paradata, response data, and nonresponse follow-up results will allow for robust comparisons between responding households that were regularly teleworking or schooling online vs. those that were not. The analysis will include multiple aspects of the survey response experience, such as: the selection of web or mail response mode, type of device used by web respondents, web survey duration, likelihood to respond early or late in the field period, and response to the primary survey request as compared to the nonresponse follow-up request. Household characteristics that may be related to the likelihood of teleworking or schooling online will also be considered, including the number and age of household members, household income and education levels, and region of the country. The presentation will conclude with lessons learned and recommendations about questionnaire design and data collection modes for surveys conducted in anomalous data collection environment.

What is the Average Daily Attendance at Your (Virtual) School? Understanding Data From Schools, Principals, and Teachers During a Pandemic

Maura Spiegelman, *National Center for Education Statistics*

Kathleen Kephart, *U.S. Census Bureau*

Jonathan Katz, *U.S. Census Bureau*

The U.S. Department of Education's National Teacher and Principal Survey (NTPS) collects data from public and private schools, principals, and teachers about the state of education. The NTPS is a repeated cross-sectional survey, allowing for analysis of changes over time for questions that have been asked in multiple administrations. The collection is self-administered, and respondents are primarily contacted by mail and e-mail with in-person follow up. For the 2020-21 NTPS, the coronavirus pandemic introduced challenges in collecting data from respondents, asking relevant questions, interpreting survey items, and reporting meaningful results. Because many schools had staff working remotely, data collection approaches had to be rethought. Several new items were added to each questionnaire to capture school and teacher experiences during the spring of 2020, when schools began shutting down due to the coronavirus. The questionnaires also include concepts that are typically well-understood by respondents, but which may have different interpretations for schools operating in virtual or hybrid environments, for example, attendance or instructional time may be complicated concepts for schools that have changed their operating procedures during the 2020-21 school year. As a result, one additional item was added to each questionnaire to capture the respondent's status at the time of data collection (fully in-person, hybrid, or fully virtual), and cognitive interviews were conducted during data collection to assess how interpretations of otherwise standard items may have changed. This presentation will focus on the findings of the cognitive interviews and discuss how this information, combined with the new status question, will allow analysts to better contextualize changes over time and to better understand the educational impacts of the coronavirus pandemic.

Sample Retention During COVID – Lessons Learned from the Middle Grades Longitudinal Study of 2017-18

Carolyn Fidelman, *National Center for Education Statistics*

The Middle Grades Longitudinal Study of 2017-18 (MGLS:2017) was impacted by school disruptions midway through its final round of data collection in 2020 because of the coronavirus pandemic, COVID-19. With no unitary response by schools, the project was forced to adjust on a district-by-district, school-by-school, or parent-by-parent basis. MGLS's approach to salvaging response rates evolved at each week from March through July. This is a description of the team's and other stakeholders' processes as they occurred in real time and the risks presented by given decisions at certain points in time. While a report of solutions that worked and results is provided, the emphasis is on recommendations for data collection disaster preparedness.

Assembling a Large-Scale Hybrid Sample to Achieve Population Representativeness During the Coronavirus Pandemic: Best Practices and Lesson Learned

Y. Patrick Hsieh, *RTI International*

Lynn Langton, *RTI International*

Michael Dennis, *NORC at the University of Chicago*

Herschel Sanders, *RTI International*

Christopher Krebs, *RTI International*

David Dutwin, *NORC at the University of Chicago*

For survey researchers, the necessity of social distancing for infection risk mitigation amid the COVID-19 pandemic has accelerated the pursuit of resource-efficient and robust-quality online survey options. Although nonprobability online panels and conversion to web survey mode seem to be good alternative research design strategies in response to the disruption caused by the pandemic, online data collection can often experience low response rates, questionable data quality, and limited sample representativeness. How can survey researchers effectively achieve sufficient survey response rates, data quality, and sample representativeness amid a pandemic? In this presentation, we seek to discuss

the lessons learned from a mixed-mode survey project that assembled a hybrid sample of more than 31,000 US adults, drawing from both probability and non-probability online panels and from Amazon Mechanical Turk to complete data collection in three weeks.

We investigated the implications of our hybrid design on respondent recruitment, data quality assessment, and weighting analysis for achieving population representativeness. Our analysis found that, as compared to the online panelists recruited via panel aggregators, MTurk workers on average spent more time answering the survey and had a lower rate of item nonresponse. As for the implications for the weighting adjustments of the sample sources, the data collected from MTurk workers introduced more bias in constructing the estimates of demographic characteristics than those collected from online panelists. Not surprisingly, respondents from the probability online panels exhibit desirable respondent behaviors similar to MTurk workers while they were contributing the least bias to the estimates. We examine the data quality and weighting analysis further by examining the differences between early and late respondents and reflect on the potential link between the timing of data collection and the data quality. Finally, we present some best practices recommendations for using various sources of survey participants.

Effects of Mode Change on Survey Measurement Due to the Coronavirus (COVID-19) Pandemic

Beth Newman, *U.S. Census Bureau*

Renee Reeves, *U.S. Census Bureau*

Christine Bottini, *U.S. Census Bureau*

In March 2020, the U.S. Census Bureau no longer allowed Field Representatives (FRs) to visit homes to conduct interviews in an effort to slow the spread of the Coronavirus (COVID-19). This was a dramatic change for many surveys, including the Survey of Income and Program Participation (SIPP), which usually conducts approximately 70 percent of its interviews in person. Following the temporary discontinuation of personal visits seven weeks into data collection, Census Bureau FRs were instructed to call sample cases and conduct interviews over the phone. A change in mode in the middle of data collection has the potential to cause measurement error in the survey. It is possible that a respondent answered questions differently during the phone interview than they would have during a personal visit. Additionally, moving to telephone may have made contacting respondents more difficult or lead to them being more comfortable ending an interview early, thereby reducing response rates and increasing drop-off rates.

This research examines whether changes to the SIPP data collection mode affected data quality. We compare several metrics across multiple survey years to determine fitness for use of the 2020 data. Metrics include overall interview length, most common drop-off points for incomplete interviews, overall response rates, partial completion rates, response distributions, and item nonresponse rates. Additionally, we develop regression models to determine whether the survey mode was associated with respondents' answers to specific survey items. Together, these analyses provide insight into potential data quality issues as a result of a change in data collection mode.

Detecting and Improving Data Quality

The Effect of Burdensome Survey Questions on Data Quality in an Omnibus Survey

Rachel Stenger, *University of Nebraska-Lincoln*

Angelica Phillips, *University of Nebraska-Lincoln*

Some survey questions are more burdensome than others due to the increased cognitive burden that they place on respondents. In interviewer-administered omnibus surveys, where researchers have to make decisions about the ordering of questions, it is unclear where in the interview to place these questions as highly burdensome questions may result in lower quality responses (Dillman, Smyth, and Christian 2014; Galesic 2006). One example of these burdensome questions is social network questions, wherein respondents are asked about members of their personal network and details about their relationships with those individuals. Although previous research has explored how social network questions influence respondent burden (e.g., Manfreda et al. 2004) and how respondent burden, in turn, influences response quality to the social network questions (e.g., Galesic 2006; Galesic and Bosnjak 2009), scant attention has been given to how social network questions can influence the response quality of answers to non-social network questions in an omnibus survey. In this study, we answer the research question of how the placement of a social network question is associated with the data quality, such as item nonresponse, on unrelated questions in an omnibus survey. To answer this research question, we use data from the 2010 General Social Survey (GSS) (Smith et al. 2019; AAPOR RR5=70.3%), which included an experiment on the location of social network questions within the interview (i.e., beginning, middle, or end). We find that the location of the social network questions within the interview was not significantly related to item nonresponse rates, nor was this relationship moderated by cognitive capacity of the respondent. We conclude with implications for questionnaire design.

Evaluating Web Survey Respondents' Interaction With Subsidized Marketplace Coverage Questions on the American Community Survey

Adriana Hernandez-Viver, *U.S. Census Bureau*

Due to the implementation of the Patient Protection and Affordable Care Act (ACA), in 2019 the U.S. Census Bureau introduced a two-part health insurance premium and subsidy receipt question to the American Community Survey (ACS) that can be used to produce estimates of the population with subsidized Marketplace coverage. It also updated the VA coverage question since veterans may qualify for subsidized Marketplace coverage and disenroll from VA health care.

ACS data collection follows a sequential mixed-mode design with three modes: web, mail, and in-person interviewing. Sampled households first receive a mailed request to respond via web, followed later by an option to complete a paper questionnaire and return it by mail. If no response is received by mail or web, the address may be selected for in-person interviewing

To evaluate respondents' interaction with the health insurance coverage question and the two-part premium and subsidy question, and identify potential problems that might impact data quality, we analyze the 2019 ACS web survey paradata internal file. Web survey paradata are automatically generated when respondents answer the web questionnaire. Web paradata analysis is a tool that can be used to assess the quality of a web survey and improve its design. We provide a variety of screen-level measures such as error messages rendered, clicks on non-question links, and changed answers, among others. We further compare the characteristics of respondents that had trouble answering the questions with those who answered without error. Results will identify issues respondents had interacting with and understanding the health insurance questions, whether some groups of respondents have more difficulty answering these questions than others, and inform recommendations to improve design.

Can You Commit to This Survey? Examining Whether Committed Web Probing Respondents Provide Responses of Higher Data Quality

Jonathan Katz, *U.S. Census Bureau*

Matt Virgile, *U.S. Census Bureau*

Mandi Martinez, *U.S. Census Bureau*

Jessica Holzberg, *U.S. Census Bureau*

In recent years, web probing has been used to supplement more traditional methods of pretesting such as in-person interviewing. One advantage of web probing is the ability to collect qualitative data from a large sample to evaluate survey questions. However, one disadvantage is the absence of a live interviewer, which may reduce respondent motivation to answer probing questions with the same level of effort.

One possible method to motivate respondents is a prompt at the beginning of a survey called a “commitment device”. This prompt may ask the respondent to be thoughtful and complete the survey to the best of their ability. Work by Charles Cannell and others in the 1970s/1980s found that asking respondents in face-to-face interviews to commit to putting in effort or reporting as accurately as possible can lead to higher data quality. Although there has been research of commitment in in-person and telephone interviewing, there has not been as much of this research conducted in web probing.

In this paper, we conduct split ballot testing of a commitment device prompt in a web probing survey, in which half of the respondents received such a prompt and the other half did not. We will compare respondents who received the prompt and respondents who did not, as seeing the prompt may motivate them to complete the survey with more effort, regardless if they commit. We will also compare responses for those who committed and those who did not, if possible. We will analyze open-ended and closed-ended probing questions and explore differences in measures of data quality between the two groups, such as the rate of non-substantive answers, length of characters in open-ended questions, time on page, perceived difficulty, and consistency of responses. These findings will inform whether a commitment device prompt can increase data quality in web probing.

Leave the Low-Quality Data at the Door: Using Screening Surveys to Increase Data Quality From Online Non-Probability Panels

Erica Yu, *U.S. Bureau of Labor Statistics*

As researchers shift toward collecting data from online non-probability panels, we must also consider methods for evaluating the quality of the data being collected from these sources. Data quality checks embedded in the survey, such as attention check or “trap” questions, are limited in their utility because they often do not identify bogus participants trying to give acceptable responses and may be perceived by participants as tricks and lower participant trust. They can also be costly because they are used to exclude data after responses have been collected, resulting in exclusion of upwards of 25% of the sample.

To reduce the amount of low-quality data collected in surveys, researchers can implement initial screening surveys to identify participants providing low-quality data and exclude them from further participation. In this study, we assess the effectiveness of using open-ended questions as a data quality screening method. The data were collected from Amazon Mechanical Turk. In an initial screening survey, participants were asked to provide demographic information and respond to open-ended questions. The responses to the open-ended questions were then evaluated and both high and low data quality participants were retained and invited to complete a follow-up survey. In this presentation, we explore whether the open-ended screening method is predictive of data quality on the follow-up survey and share recommendations for developing effective data quality screeners.

How Do Survey Interviewers Handle Respondents' Satisficing Tendencies? An Analysis Based on Audio-Recordings of Face-to-Face Interviews

Silvia Schwanhäuser, *Institute for Employment Research*

Bettina Müller, *Institute for Employment Research (IAB)*

Non-differentiation, extreme responding, or item nonresponse are commonly attributed to respondents' satisficing behavior. In interviewer-administered surveys, the effect of such response styles on survey outcomes may additionally depend on interviewer behavior (Loosveldt & Beullens 2017). While interviewers' assistance of the answering process and adherence to standardized interviewing may counteract respondent satisficing (Heerwegh 2008), interviewers may as well attune to individual response styles in a way that promotes satisficing, e.g., by simplifying the presentation of response scales. Previous research indicates that differential interviewer behavior is indeed an important contributing factor to measurement error (e.g., Loosveldt & Beullens 2017; Olson & Bilgen 2011). However, this research so far lacks direct measures of interviewer behavior necessary to disclose the underlying mechanisms.

We contribute to this research with a more detailed assessment of interviewer influences based on direct measures of interviewer behavior from audio-recordings of face-to-face interviews in the German panel study "Labour Market and Social Security" (PASS). The PASS survey is administered to a sample of welfare recipients, comprising a high proportion of individuals with migration background and low education. This population may face greater difficulties in answering survey questions and thus be more prone both to satisficing and interviewer influence as compared to the general population. Interviewers on the other hand may be more inclined to deviate from standardized interviewing.

Applying multilevel random intercept models, we find significant interviewer effects on non-differentiation, extreme responding, and item-nonresponse in the PASS study. To discern the sources of these influences, we analyze audio-recordings of interviews. Specifically, we apply behavior coding to compare exceptional interviewers with positive and negative effects on the data quality indicators under study. Ultimately, this research provides practical guidance for interviewer training and monitoring.

Survey Speeding

Carla VanBeselaere, *Mount Allison University*

Completing a survey rapidly compared to the norm, known colloquially as survey speeding, is considered problematic survey behavior because respondents are likely not providing thoughtful, accurate answers (Rao, Wells, & Luo, 2014; Zhang, 2003). But how can one determine what constitutes survey speeding? Most studies of speeding look at the time it takes respondents to answer the survey and compare this to some baseline but this does not capture the fact that the time to respond to a question varies by question and respondent; task difficulty, respondent ability, and respondent motivation all mediate how thoughtfully respondents answer questions (Krosnick, 1991). Using data from the 1998-1999 Multi-Investigator Study (MI-1998), which collected response times for several questions in addition to demographic characteristics of respondents and a measure of how difficult respondents found the question, it is possible to model how long we expect respondents to take in order to answer a question. From this we can examine which respondents appear to be speeding on the survey. Furthermore we can compare the behaviour of these speeders with those of other respondents in an attempt to determine whether responses from speeders are accurate and reliable.

Web Survey Data Quality Assessment

Ioannis Andreadis, *Aristotle University Thessaloniki*

In this paper, we present a series of R functions that can be used for the assessment of the quality of web survey data. According to related survey methodology literature, web survey respondents may get tired or lose their interest and motivation and then they tend to respond less carefully to the questions of a survey (i.e. with minimal cognitive effort). This may result in lower response quality in terms of speeding, providing non-substantive answers (i.e. no opinion, "don't know", "none of the above"), non-differentiation in grid questions, or choosing midpoint responses (i.e. "neither/nor"). Survey methodology scholars have used many methods to measure response quality. Based on the theory of satisficing we use a series of indicators to estimate the level of engagement of web survey participants in answering the questionnaires: i) item-nonresponse (skipping). ii) mid-point responses in Likert-type scale items, because respondents may choose mid-point responses when they do not process a question with the required effort. iii) the time spent on questionnaire items (speeding): e.g. using the minimum time needed to read and answer an attitudinal question given the length of the question text. iv) various patterns in answering grid questions, such as straight-lining. Using the aforementioned response quality indicators we can create an innovative multidimensional estimation of response quality that can help us identify less engaged web survey respondents.

Cross-National and Cross-Cultural Research

A Collaborative Approach to Developing a Maternal and Child Health Survey for the U.S. Jurisdictions

Anna Wiencrot, *NORC at the University of Chicago*

Clare Davidson, *NORC at the University of Chicago*

Brian Kirchhoff, *NORC at the University of Chicago*

Michael Stern, *Michigan State University*

Erin Fordyce, *NORC at the University of Chicago*

Lisa Lee, *NORC at the University of Chicago*

Benjamin Skalland, *NORC at the University of Chicago*

Michael Kogan, *Maternal and Child Health Bureau, Health Resources & Services Administration*

The mission of the Maternal and Child Health Bureau (MCHB) is to improve the health and well-being of mothers, children, and families. To serve this mission, their Title V State Block Grant program emphasizes MCH data available to the 59 states and jurisdictions for their selection of performance measures and identification of priority areas. However, the eight U.S. Jurisdictions (Puerto Rico, the U.S. Virgin Islands, Palau, Marshall Islands, Guam, American Samoa, Northern Mariana Islands, and Federated States of Micronesia) have had limited data compared to the states. With the goal of meeting this need for data, in a way that was standardized across all jurisdictions, the MCH Jurisdictional Survey was developed through a collaborative effort between NORC, MCHB and the jurisdictional MCH leadership.

To develop a survey tailored to each jurisdictions' unique data needs, priority areas were identified through background research; interviews with Centers for Disease Control and Prevention (CDC) staff, jurisdictional MCH leads, and other experts; and, reviews of existing surveys. The results of this assessment were used to adapt the National Survey of Children's Health (NSCH) for use in the jurisdictions, with the addition of maternal health and jurisdiction-specific questions. Working closely with the jurisdictional MCH leads, the questionnaire was pretested in summer 2018. The results of this pretest informed the final questionnaire. Further collaboration with the jurisdictional MCH leads and data collection experts in these and similar areas informed the development of the field plan and sample design. Standardized training protocols, training materials, and random walk procedures helped ensure data were collected in a uniform manner. Data collection was conducted once in each jurisdiction between May 2019 and February 2020, with over 200 surveys completed in each location. These data are now being used by MCHB and the jurisdictions to improve planning and reporting on MCH priorities.

Revisiting the Need for ACASI in Face to Face Surveys: The Role of Culture and Language

Emilia Peytcheva, *RTI International*

Andy Peytchev, *RTI International*

Many national surveys that collect data on sensitive or socially undesirable behaviors employ Audio Computer Assisted Self-Interviewing (ACASI) to improve report accuracy. Indeed, ACASI has been found to increase reporting of sensitive behaviors (e.g., Tourangeau and Smith, 1996), attributed to the more private setting it offers. However, recent research demonstrates that ACASI may not be as beneficial as originally considered, and that it can either have systematic negative effects on the estimates, or no effect at all. Such findings raise the question under what circumstances ACASI may not yield the expected benefits and why.

One possibility is that ACASI is particularly valuable when there is cultural stigma associated with certain sensitive behaviors, suggesting that for particular subgroups of the target population, ACASI may contribute disproportionately to more accurate reporting than for others. For example, alcohol use, mental health problems and abortion are more stigmatized among Hispanics than other ethnic groups. Historically, a higher percent of Hispanics have opposed abortion (Pew, 2009), suggesting higher topic sensitivity and possibly higher underreporting of this sensitive behavior.

We explore this possibility by analyzing data from the National Survey of Family Growth. We compare reports of abortion in CAPI and ACASI modes by respondent ethnicity and language. We expected to find larger differences in abortion reports between the two modes for Hispanics and those interviewed in Spanish relative to non-Hispanics and those interviewed in English. An alternative hypothesis is that since abortion is so stigmatized, it is always underreported among Hispanics and a self-administered mode will not make a difference. The effect of Hispanic origin on the need for ACASI was unexpectedly large—over three-fold. The results of this research have important implications for data collection protocols and survey costs, as well as major implications for bias in estimates of health disparities.

Calibrated Multilevel Regression With Poststratification for SMS Survey Data in Low-Income Countries

Sarah Hughes, *Mathematica*

Jon Gellar, *Mathematica*

Constance Delannoy, *University of Colorado at Boulder*

Shirley Jeffreys-Leach, *FinMark Trust*

Bobby Berkowitz, *FinMark Trust*

Grant Robertson, *FinMark Trust*

Face-to-face (FTF) surveys have been the traditional method to gather nationally representative data and remain the dominant data collection mode in resource-poor countries. Conducting these surveys is expensive and time consuming. With the rapid expansion of mobile phone use, Short Message Service (SMS) presents an opportunity to conduct inexpensive, fast, and scalable surveys. However, using SMS surveys to obtain nationally representative estimates can lead to two types of bias: selection bias and survey mode effects. Standard adjustments to correct for nonrepresentative sampling can be insufficient to remove all biases. We introduce calibrated multilevel regression with poststratification (calibrated MRP), a procedure that corrects for residual bias by incorporating a relatively small sample of FTF data that is known to be unbiased. We apply this method to the problem of estimating financial inclusion (access to formal banking systems) in eight countries in Africa and Asia; we focus on Uganda in the interest of clarity. We find that our calibrated MRP approach is effective in replicating estimates from a larger and much more expensive F2F survey. This paper includes a description of our methods as well as results from the financial inclusion study and a discussion of limitations and future areas for research.

Trends in Response Rates to Health Surveys in Sub-Saharan Africa

Rachel Bray, *ICAP at Columbia University*

Shannon Farley, *ICAP at Columbia University*

It is important to examine changes and trends in response rates over time for project planning, especially as cultural factors influence survey response and trends in response rates may differ between survey contexts. Studies from higher income countries have found that response rates are declining (De Heer and De Leeuw, 2002; Koen et. al, 2018; Williams and Brick, 2018). However, the literature on whether these the declining trends in response rates hold in developing countries is sparse. A review of more than 20 national population-based HIV prevalence surveys in sub-Saharan Africa in 2006 found that response rates in general were high, with only three of the surveys having household response rates under 90%, yet did not examine trends over time (García-Calleja, Gouws and Ghys, 2006). We are compiling response rates from published reports of health surveys in Sub-Saharan Africa to examine whether response rates are declining over time in this context. Most of the data for this metanalysis comes from three nationally representative groups of household based-surveys: the Multiple Indicator Cluster Surveys (MICS) conducted by UNICEF, the Demographic and Health Surveys (DHS) conducted by ICF International, and the Population Based HIV Impact Surveys (PHIAs) conducted by ICAP. Together, these surveys represent nearly 30 years of data across over 40 countries. We will examine household and individual response rates in these surveys over time, standardizing definitions of response rate where necessary. Linear regression will be used to identify whether response rates are increasing, decreasing, or holding steady over time. Preliminary results are showing a slight decrease in household response rates over time, and a slight increase in individual interview response rates for men and women (conditional on the household interview being completed). Country and regional effects will be explored through maps and mixed models.

Adapting a U.S. Market Research Study in China and Germany

Christina Tworek, *HarrisX*

Anna Gurevich, *HarrisX*

Marina Taraukhina, *HarrisX*

Dritan Nesho, *HarrisX*

Comparative research is essential to extract meaningful insights that apply across different populations. 3M research (multinational, multiregional, and multicultural) is relatively new and has risen dramatically in popularity only over the last three decades. Researchers in academic and market research settings continue to refine their 3M methodologies, but they face many challenges at both the conceptual and operational level including data collection, translation, and interpretation. One particularly important problem is adapting a methodology designed for one market to an entirely different market.

Here, we present a post-mortem of a set of comparative studies first conducted in the United States (N = 5,018) and then adapted in Germany (N = 5,000) and China (N = 5,041). The original study was designed to examine consumer attitudes regarding the privacy and security of their online information in the US. This study was refreshed in Germany and China to understand market opportunities within each country, as well as differences between countries. Maintaining consistent study methodology was key to adequately understanding market differences when transitioning the US study to Germany and China. All studies were conducted via an online survey, and panelists were recruited via third-party providers. Respondents from each country completed the questionnaire in their country's native language. The original questionnaire was written in English for US respondents and translated to German and Mandarin for German and Chinese respondents, respectively. We discuss the challenges we faced during translation, data collection, analysis, and reporting, emphasizing the unique methodological concerns for each country. For example, we found between-country differences in both recruitment speed and scale interpretation. We also report on the strategies we used to overcome these challenges and discuss implications for future research.

The Devil is in the Details: A Randomised Experiment to Assess the Effect of Providing Examples on a Question About Confidence in Social Media

Eva Aizpurua, *European Social Survey HQ - City, University of London*

Gianmaria Bottoni, *ESS HQ - City, University of London*

Rory Fitzgerald, *ESS HQ - City, University of London*

Despite the widespread use of examples in survey questions, very few studies have examined their impact on survey responses. These studies (Martin, 2002; Tourangeau, Sun, Conrad, & Couper, 2016) have found that respondents tend to echo the level of generality of the examples, supporting their value clarifying the scope of the questions. However, the tendency to answer using the examples provided in the questions has raised concerns regarding the risk of respondents focussing excessively on the examples. This has been referred to as the 'focusing hypothesis' (Tourangeau et al., 2016). In addition to the scarcity of research in this area, to date the evidence is based on data collected in the United States using questionnaires in English. This study builds on previous research by examining the effects of providing examples using data from CRONOS, a cross-national, probability-based web panel implemented in Estonia (n = 730), Great Britain (n = 685), and Slovenia (n = 529). Respondents were randomly assigned to two experimental groups. The first group received a question assessing confidence in social media using Facebook and Twitter as examples, while the second group answered the same question with no examples. To analyse how the examples influence the retrieval stage of the response process, respondents were then asked about the social media platforms considered when answering the confidence question. We compare the confidence placed in social media and the platforms taken into consideration by experimental condition, analysing whether the findings are consistent across the three countries. The implications of these findings and how they fit in with previous work are discussed.

Cross-Cultural Researchers' Positionality in Immigrant Health Research: Reflections on Research Examining Chinese Immigrants' Experiences of Exclusion

Lei Chen, *University of California, Los Angeles*

Hin Wing Tse, *UCLA Center for Health Policy Research*

Deborah Wu, *Johns Hopkins University, School of Nursing*

Maria-Elena Young, *University of California, Merced*

While a growing body of research has examined immigrants' health generally, less is known specifically about the impact of immigration policy on the health of Chinese immigrants, the second-largest foreign-born population in the U.S. This is due, in part, to the lack of methodologically well-trained, cross-cultural researchers who have both the cultural and linguistic expertise and health knowledge to engage with Chinese immigrant populations. This paper fills this gap by examining Chinese cross-cultural researchers' roles in the qualitative phase of the Research on Immigrant Health and State Policy Study (RIGHTS), which sought to assess how immigration policies shaped Chinese and other immigrants' experiences in healthcare and other related sectors. The authors use self-reflexivity to assess Chinese cross-cultural researchers' positionality of insiderness and outsiderhood and its influence on the process of data collection (i.e., recruitment, conducting interviews, transcription and translation) and communication with the Chinese community, research participants, and the broader research team. Discussion of self-reflexivity and researchers' positionality guides the assessment of the opportunities (e.g., expanding the recruitment pool, engaging participants more effectively in interviews, ensuring data integrity) and challenges (e.g., hard to recruit low income and undocumented immigrants, time-consuming transcription and translation) in conducting immigrant health research with the Chinese community. These results highlight the need for cross-cultural researchers to help build trusting relationships with ethnic-minority communities, thus gaining new insights and advancing knowledge within the field of ethnic minority health research. These insights can guide future investigations of Chinese immigrant communities and other communities of color as research on immigration policy and health continues to expand.

Measuring Religious Non-Affiliation in China: A Comparison of Major National Surveys

Chan Zhang, *Zhejiang University*

Philip S. Brenner, *University of Massachusetts*

Lirui He, *Jinan University*

China is one of the least religious countries in the world, with most Chinese typically reporting no religious affiliation on surveys. However, measuring the rate of non-affiliation in China may be as challenging as measuring it in the US and elsewhere. An examination of the major national surveys in China reveals that a variety of survey design features differ across the surveys measuring religious non-affiliation. This study aims to evaluate the trend of the religious non-affiliation after accounting for the survey design differences and identify the critical survey design features that fundamentally impact estimates. We combined the data from one international or four domestic survey programs. In total, our dataset contained survey responses from over 200 thousand respondents across 22 surveys from 1990 to 2018. We found that two surveys asking beliefs in a mix of gods and religions yielded dramatically lower estimates of non-affiliation compared to the others that measured affiliation explicitly. Among the latter, allowing multiple choices and putting "no religion" toward the end of the lists of response options, moderately decreased the reports of non-affiliation. After taking into the question design and demographic differences across the surveys, we found a significant decline in religious non-affiliation since 1990, but the trend is less unclear after 2005.

Innovative Approaches in Targeting and Adapting Data Collection Efforts

An Experimental Evaluation of Alternative Methods for Case Prioritization in Responsive Survey Design

Brady West, *University of Michigan*

Wen Chang, *University of Michigan*

Alexa Zmich, *University of Michigan*

Growing reluctance of households to participate in surveys has led to a variety of methodological efforts to combat this phenomenon. Several organizations that use interviewers to collect survey data employ case prioritization in a responsive survey design framework, asking interviewers to dedicate increased effort to specific subgroups of sampled cases during certain data collection phases. For example, some surveys may prioritize cases from specific subgroups defined by age and/or race/ethnicity if response rates in those subgroups are lagging and balanced response rates across all subgroups are important for minimizing nonresponse bias. Unfortunately, no methodological studies to date have identified optimal approaches for applying this increased effort to prioritized cases. We will describe results from a study that experimentally examined three alternative methods for case prioritization in the National Survey of Family Growth: 1) simply flagging the cases to receive increased effort in the sample management system (a “standard” method), 2) developing tailored approaches to working the prioritized cases, and 3) no prioritization (a “control” method). In the “tailored” method, which was designed to provide the interviewers with more guidance than simple “flagging”, interviewers worked with their supervisors to develop tailored strategies for how to best work each case within the prioritized subgroup. We found that both prioritization methods improved response rates and led to significant reductions in calling effort per completed case, with the “flagging” approach working particularly well. Given the additional costs associated with the “tailored” method, the results of our experiment provide support for a “hybrid” case prioritization approach that combines optimal features of these two methods.

Using R-Indicators to Make Case-Level Decisions for GSS 2020

Steven Pedlow, *NORC at the University of Chicago*

NORC uses R-indicators (Representativity Indicators from Schouten et al. 2009) on projects to monitor overall representativeness and to make operational adjustments to prioritize under-represented groups through future outreach in real time. R-indicators come from regression modeling on a response variable and include an overall R-score (overall representativeness), unconditional partial R-scores for each predictor variable subgroup, and a conditional partial R-indicator for each variable. The overall R-indicator score measures the variation in response propensities among all cases while the partial R-indicators split the R-indicator score into between (unconditional) and within (conditional) variation.

In 2020, the General Social Survey (GSS) applied the case-level sum of the unconditional partial R-indicators to make decisions on whether to target a case with extra effort and higher incentives, continue working a case, or stop attempts to recruit a case. A panel sample of 2016 and 2018 GSS respondents was collected in autumn 2020 while a new cross-sectional sample is starting in November, 2020. The panel was a unique opportunity to use R-indicators with known information about each and every sample member. We used demographics and some opinion variables as well. The cross-section is a more typical application where we can only use information based on where they live.

We will show the models developed, the decisions made about cases, and our R-indicators results for GSS 2020. One lesson learned is that the overall R-score is not a good measure of representativeness in isolation. If you consider fewer variables and leave out better predictors, you will have a higher overall R-score. By using the sum of the R-scores, we were able to achieve a higher response rate among the cases we most needed to improve our representativeness.

Combining Propensity Models and Qualitative Analysis to Tailor Respondent Approach for a National Building Survey

Jennifer Taub, *Energy Information Administration*

Zachary Marohl, *Energy Information Administration*

One of the challenges to obtaining a high response rate in an establishment survey is the difficult process of finding a willing and knowledgeable respondent. In this paper, we analyze both quantitative and qualitative paradata available from Commercial Energy Consumption Survey (CECS). High response rate is an important aspect to a successful survey leading to the study of paradata. The CECS is unique in that it is the only nationally-representative survey of U.S. commercial building stock and its corresponding energy use. This leads to some distinct difficulties in finding a knowledgeable respondent, due to the large differences in use, size, and bureaucratic set up of the sampled buildings. Using the 2018 CECS we developed a response propensity model, finding that the region, building use, and size all had significant effects on a respondent's propensity to respond. The CECS also has a wealth of qualitative data from interviewer notes. We took a sample of the interviewer notes from the 3 of the most ubiquitous building uses (Education, Office, and Warehouse) and qualitatively analyzed this sample to give us a fuller picture of how to gain cooperation from building respondents. Then from the qualitative analysis we created tags, such as 'security' or 'receptionist', which we then plugged into a new propensity model to see if they helped to predict a respondent's likelihood to respond. We hope our findings can help maximize response rates for the next CECS cycle, and may also shed some light on maximizing response rates for other surveys wherein finding the respondent may pose a difficulty such as establishment surveys.

Two Novel Approaches for Estimating Population Size in the Absence of Census Data

Nathaniel Poland, *NORC at the University of Chicago*

Lee Fiorio, *NORC at the University of Chicago*

Gregory Haugan, *NORC at the University of Chicago*

Brian Kirchhoff, *NORC at the University of Chicago*

Anna Wiencrot, *NORC at the University of Chicago*

In ideal settings for large, population-based surveys, survey practitioners start from a sample frame of enumeration areas with population figures from national census data. In practice, in many international locations, these enumeration areas and national census data may not exist, are of unsuitable quality for use in rigorous sampling designs, or otherwise may not be available to researchers. We describe two methods for creating enumeration areas and obtaining population estimates using publicly available data which allow for replicable, systematic sampling of populations. For work in Greater Accra, Ghana, we describe a method for producing an estimated number of businesses per 100 meter grid cells and then aggregate grid cells into compact enumeration areas for use during fieldwork. This method uses open source VIIRS night-time light data available from WorldPop.org and incorporates street blocks from Open Street Maps to create enumeration areas with easily identifiable borders and compact shapes. This allowed field teams to conduct a complete listing of businesses in selected enumeration areas. For Pohnpei and Weno Islands in the Federated States of Micronesia, we describe a method for producing estimates of households and household population using 30 meter grid cells derived from satellite imagery using open source data from Facebook's Data for Good project. The households are grouped together to form enumeration areas with sizes and shapes designed to take into account natural geography in order to ease interviewer burden when conducting a random walk of the enumeration area to select households. These methods provide useful alternatives to researchers about population distributions when more traditional methods are not available and these solutions can be adapted to other uses depending on the available public information.

More for Less? Targeting Specific Households for FedEx Mailing Service

Mahi Megra, *American Institutes for Research*

Rebecca Medway, *American Institutes for Research*

Mickey Jackson, *American Institutes for Research*

The National Household Education Surveys program (NHES), a national cross-sectional survey, has used mailed invitations to contact households since 2012. Nonrespondents to the initial invitation receive up to three reminder mailings. Typically, one of the reminders is sent using FedEx courier service. In the 2019 administration, an experiment was included that varied when this FedEx mailing was sent to assess the ideal timing for sending this costly reminder. As part of this experiment, 70,000 sampled cases were randomly assigned to one of three experimental conditions: FedEx second, FedEx fourth and Modeled FedEx conditions. Cases that were assigned to FedEx second or fourth conditions, received FedEx as the second or fourth mailing, respectively. In the modeled FedEx condition, a cost-weighted response propensity model based on available frame data was used to identify cases that were the most likely to respond to the screener, and for which the FedEx mailing is the costliest—these cases received FedEx as a fourth mailing while the remaining sample received FedEx as a second mailing.

In the current study, the effectiveness of the model used to assign cases to receive FedEx as the second or fourth mailing will be evaluated by comparing response rates, respondent representativeness, and cost outcome indicators amongst the three experimental conditions. Since FedEx mailing timing is based on a pre-defined model, results from this study will provide insight into whether FedEx mail timing is an effective adaptive design intervention in address-based sample surveys. Findings from the study will also guide future NHES designs and inform other researchers seeking to hold-off more expensive methods of reminders to targeted sample members.

Is There an Ideal Shift Length That Maximizes Interviewer Productivity?

Nikki Gohring, *University of Nebraska-Lincoln*

Lindsey Witt-Swanson, *University of Nebraska-Lincoln*

Jolene D. Smyth, *University of Nebraska-Lincoln*

Kristen Olson, *University of Nebraska-Lincoln*

Interviewers are often evaluated based on their productivity levels. Although organizations vary in the types of productivity measures used, two common metrics include the number of dials (call attempts) per hour and the number of completes per hour (Tarnai & Moore, 2008, Gwartney, 2007). Higher productivity levels can help decrease survey costs (Tarnai & Moore, 2008). As such, organizations may try to maximize interviewer productivity levels by setting targets for these metrics. In order to help achieve these targets, organizations may adjust interviewers' schedules to help yield more productive shifts (Tarnai & Moore, 2008, Kreuter & Olson, 2013, Gwartney, 2007).

An open question is whether the length of an interviewer's shift is associated with their productivity levels. Although some organizations have a fixed schedule for interviewers, other organizations provide more flexible scheduling, allowing interviewers to identify the hours that they want to work. We examine data from the Bureau of Sociological Research (BOSR) interviewing pool (n=69 interviewers), which has offered flexible scheduling to the interviewing staff for several years. Interviewers schedule themselves for between one to eight hours per workday, the majority working three to four hours on average. One hypothesis is that interviewers who work shorter shifts are able to stay productive throughout the shift, whereas interviewers who work longer shifts may tire more easily and be less productive. An alternative hypothesis is that interviewers are less productive during shorter shifts because it takes interviewers time during a day to "hit their stride." We examine the association between shift length and productivity measures of dials per hour and completes per hour among BOSR's interviewing staff (n=69) for the 2019 calendar year. We evaluate weekday days, weekday evenings, and weekend shifts separately. We conclude with implications for survey managers on shift length and interviewer productivity.

The Impact of Callback Attempts on Sample Profile and Substantive Outcomes: A Cross-National, European Review

Alexandra Castillo, *Pew Research Center*

To optimize the balance between methodological quality and operational costs in CATI projects, there are a host of design decisions practitioners may consider. One consideration is whether it's advisable to reduce the amount of effort to expend on each piece of sample — that is, to lower the maximum number of call attempts per phone number. There are critical tradeoffs, of course. Costs and timeline pressures are likely relieved with less required effort to close sample. Less clear are the potential implications for both sample performance and substantive outcomes. One concern is that reducing the callback design could bias the sample toward easier-to-reach populations. For example, our probability-based, nationally representative CATI surveys in Europe tend to skew toward older and more educated populations, meaning a change in call design could worsen this trend. Another concern is that a reduced call design could change substantive point estimates for populations via differential item nonresponse and attitudinal variation across subgroups.

Pew Research Center has primarily employed a 7-callback design for its annual Global Attitudes Project. Focusing on a subset of European countries since 2019, this presentation will explore the implications of a reduced callback design cross-nationally, across time and by phone type. Key questions of interest include whether the impacts of a reduced callback design are stable across time and space, whether the deviations from sample profile are within a tolerable range, and whether substantive outcomes are impacted by design decisions. The presentation will focus on key demographic components of sample performance, like gender, age and educational attainment, along with valuable substantive outcomes. Additional results will speak to gained efficiencies for each call design, such as the number of calls needed to achieve an interview, the effect of different callback designs on contact and cooperation rates, and consequences of fielding during the pandemic.

Novel Use of Digital Trace Data From Social Media, Mobile Apps or Sensors in Surveys

When Survey Science Met Online Tracking: An Error Framework for Metered Data

Oriol J. Bosch, *The London School of Economics and Political Science*

Melanie Revilla, *Research and Expertise Centre for Survey Methodology - Universitat Pompeu Fabra*

Metered data (also called web log data, digital trace data, web-tracking data, or passive data) is a type of Big Data obtained from a meter willingly installed or configured by participants on their devices (PCs, tablets and/or smartphones). A meter refers to a heterogeneous group of tracking technologies that allow sharing with the researchers, at least, information about the URLs of the web pages visited by the participants. Depending on the technology used, HTML, time or device information can also be collected. Metered data is objective, free of human memory limitations and produced in real-time. Therefore, metered data has a great potential to replace part of survey data or to be combined with survey data to obtain a more complete picture of reality. Metered data, nevertheless, needs to be used properly. It is crucial to understand its limitations to mitigate potential errors. To date, some research has explored potential error causes of metered data. However, a systematic categorization and conceptualization of these errors is missing.

Therefore, in this presentation, we present a framework of all errors that can occur when using metered data. We adapt the Total Survey Error (TSE) framework (Groves et al., 2009) to accommodate it to the specific error generating processes and error causes of metered data. The adapted error framework shows how the unique characteristics of metered data can affect data quality, but also allows comparing metered data errors with survey errors, since it is based on the TSE framework. Hence, the adapted framework can be useful when using metered data alone or in combination with surveys, to choose the best design options for metered data, but also to make better informed decisions while planning when and how to supplement or replace survey data with metered data.

Who Produces Twitter Posts? Are They Different From Non-Posters and Non-Users?

Zeina Mneimneh, *University of Michigan*

The use of social media among the US population has been increasing over time. The real time and organic nature of the data shared by users of these platforms have provided researchers an unprecedented opportunity to investigate human behaviors and attitudes in a cost-effective manner (Pfeffer, 2019). Twitter, in particular, provides publicly available and accessible data to researchers. While the public and accessible nature of the data are strengths of this platform for researchers, the lack of clear understanding of who is represented in a specific sample of tweets poses challenges for interpretation of findings and obstacles to their inference and replication.

In this presentation we investigate: 1) How individuals who have a personal Twitter account differ from those who don't with respect to socio-demographics, religious, racial and political attitudes, and health conditions; and more specifically 2) How individuals who have a personal account and produce Twitter data (i.e. tweet or retweet), and how those who have a personal account but don't produce Twitter data, differ from those who don't have a personal Twitter account. The second question is important given that the majority of Twitter posts generated in the US come from a small fraction of Twitter users (Wojcik S and Hughes, 2019). To investigate these research questions we use data from a web survey that was conducted in 2020 on a probability sample of Twitter users and non-users. The sample was selected from an online probability panel that represents the adult US population. The survey completion rate was 76%. Twitter users were oversampled and a total number of 1709 users and 854 non-users were interviewed. Results from this survey will be presented shedding light on the inference properties of Twitter data as they relate to the socio-demographics of posters, their religious, racial and political attitudes and health conditions.

Sweet Tweets! Exploring a New Method for Probability-Based Twitter Sampling

Trent Buskirk, *Bowling Green State University*

Brian Blakely, *Bowling Green State University*

Adam Eck, *Oberlin College*

Ravinder Singh, *Bowling Green State University*

Youzhi Yu, *Bowling Green State University*

As survey costs continue to rise and response rates decline, researchers are seeking more cost-effective ways to collect, analyze and process social and public opinion data. As we navigate a global pandemic, data collection using traditional modes such as in-person interviewing and to some extent, telephone interviewing, are also facing challenging barriers with respect to interviewer and respondent safety. Altogether, these issues create an opportunity to expand the fit-for-purpose paradigm to include alternate sources such as passively collected sensor data and social media data. But methods for accessing, sourcing and sampling social media data are just now being developed. In fact, there has been a small but growing body of literature investigating how the Twitter platform can be used to source data for understanding social phenomena and public opinion. These studies have focused on sampling Twitter users, keywords, or collections of tweets gathered from specific users. Other studies have focused on comparing different Twitter data access methods through either the elaborate firehose or the free Twitter search or streaming APIs. In this study, we introduce a new method we recently developed to select a probability-based sample of tweets based on estimated tweet volume from a given day and region. Specifically, we describe the Velocity-Based Estimation for Sampling Tweets (VBEST) algorithm that creates a collection of primary sampling units (PSUs) of time blocks for a given day. These PSUs each have approximately equal number of expected tweets across a given day and are sampled using a systematic random sample of tweets that is representative of Twitter volume. We will compare the performance of VBEST sample estimates of the distribution of total tweets as well as COVID-19 keyword incidence to those obtained from the full firehose, as well as to others derived from uniform time sampling and Twitter search API samples.

Validation of Survey-Based Health Indicators Through Health App Data

Evgenia Kapousouz, *University of Illinois Chicago*

Christoph Beuthner, *GESIS - Leibniz Institute for the Social Sciences*

Florian Keusch, *University of Mannheim*

Henning Silber, *GESIS - Leibniz Institute for the Social Sciences*

Bernd Weiß, *GESIS - Leibniz Institute for the Social Sciences*

Timothy P. Johnson, *University of Illinois Chicago*

Digitalization has opened entirely new possibilities concerning available data sources for empirical social research. This study makes use of the fact that an increasing number of people use smartphone devices for daily communication and many other activities. Against this background, our study explores the linkage of survey and health app data and the benefits of doing so. The main research question is whether health app data can be successfully used to validate survey-based health indicators. Data were collected in December 2019 as part of a German non-probability web survey, in which 1,085 respondents who reported having an iPhone were asked whether they were willing to share their health data. Data sharing could be completed by first downloading the data from one's smartphone and then uploading them via a data-sharing platform. This data-sharing method is often referred to as data donation because the respondents give their data away freely in support of the research. A randomized incentive experiment with four different groups (conditional incentive of 0, 2.5, 5, or 10 Euros for uploading the data) was included to test optimal monetary motivation for the data sharing request. In total, we were able to obtain 104 health app data sets, which resulted in a data-sharing rate of 9.9 percent. To address our research questions, we explore associations between survey-based and app-based measures such as between the survey question "How many times they have been physically active for 30 minutes or more over the past week?" and the number of minutes that a respondent walked the week before the survey, measured via the respondent's Apple device (smartphone or smartwatch). In addition, we also examine the amount and quality of the obtained health app data more generally. We conclude by discussing a set of recommendations for health app data collection.

Affective Language on Facebook and Issue Polarization of Public Opinion

Yujin Kim, *University of Texas at Austin*

Substantial studies have contributed to our understanding of political polarization, especially how affect addresses the current divisions in both political and everyday life settings. This study suggests affective language as a potential lead in political polarization among the public. This study examines the effects of language that appears on postings on Facebook with negativity and incivility. Based on the evidence of how the two political parties in the U.S. speak different languages in various political domains (e.g., Gentzkow & Shapiro, 2010; Jensen et al., 2012), this study focuses on how Facebook pages with left- or right-leanings contribute to the increased opinion polarization of the public. With applications of topic modeling and sentiment analysis of negativity and incivility with about 968,000 postings on Facebook that correspond to public opinions on issues from the four ANES studies between 2012 and 2018, findings show increases in political polarization over time and in partisan asymmetric effects of the Facebook pages between affective language use and mass polarization on issues. Implications of this study address the role of language use between toxic communication on social media and political polarization and manifestations of affective political language and subsequent effects on issue polarization of the public with applications of mixed methods.

Do Proposals of Anti-LGBT Legislation Harm Teenagers Mental Health?

Claire Kelley, *Child Trends*

Dominique Parris, *Child Trends*

Emily Fulks, *Child Trends*

In the last few years, even as federal legislation has largely been in favor of LGBTQ rights, state legislatures have passed laws that may be harmful to the LGBTQ community. These include laws restricting participation of trans youth in athletics, allowing religious exemption from same-sex marriages, or allowing religious exceptions to health care provision. We hypothesize that these laws may harm the LGBTQ community not just through the direct impact of the legislation, but also through signals that these laws send about LGBTQ acceptance. Anecdotal reports from LGBTQ support organizations, such as the Trevor Project, suggest that passage of such laws may create short spikes in LGBTQ mental health crisis, particularly among teenagers. And organizations running LGBTQ-focused suicide prevention hotlines and text lines have long pointed to sharp increases in calls during times when anti-LGBTQ state policies have been proposed (Perone, 2015). To study this, we draw on a policy scan, done using PolitcoPro, to collect all state level anti-LGBT legislation introduced from 2015-2020. In this systematic search we found 221 proposed laws (including laws in state houses of representatives and senates) that can be classified as anti-LGBT. To examine how these proposed laws impacted youth mental health, we draw on data about text volume from the Crisis Text Line, a non-profit organization that provides text-based counseling. Crisis Text Line indicates that approximately 75% of their texters are under the age of 25, and in a subset of texts, participants also fill out a survey indicating if they are LGBT. Using this data, we perform an exploratory analysis to test the hypothesis that texts to the crisis text line will increase in states where anti-LGBT laws are proposed.

Using Novel Data To Understand Youth Suicidality

Sarah Kelley, *Child Trends*

Claire Kelley, *Child Trends*

There is no single perfect measure of youth suicidality: surveys are vulnerable to social desirability bias, suicide attempts and/or hospitalizations don't capture suicidal ideation that doesn't lead to action, and suicide rates are influenced by the methods used. This paper augments our understanding of youth suicidality by testing whether two novel sources of data – texts to the Crisis Text Line and Google searches for suicide (or other mental health related) content – can help understand the full picture of youth suicidality and provide an early warning systems to identify areas of rising mental health issues. Both sources of data provide geographically specific information, allowing us to identify geo-spatial correlations between both metrics.

Our use of Google search data builds on the growing interest in using search engine data to understand a wide variety of social-science relevant trends from predicting outbreaks of the flu, to forecasting unemployment, to identifying rates of tobacco use or HIV infection to improving estimates of child maltreatment rates (respectively Ginsberg et. al. 2009; Askitas and Zimmermann 2009; Google search data has the strong advantage that it represents people's organic, unprompted information seeking (and information seeking about issues people may not be willing to discuss with a survey researcher). We supplement this data with data about texts to Crisis Text Line, a text-based counselling service that focuses on youth under 25. Text volume is another window into the level of mental distress and help seeking that youth in a given area are experiencing. Combining these diverse (and differently biased) sources allows us to build a fuller picture of a difficult-to-measure issue.

Recruitment Methods for Push-to-Web Surveys

Evaluating The Potential Of Mode Choice And Contingent Incentives As An Adaptive Design Intervention In Self-Administered Mixed-Mode Studies

Michael Jackson, *American Institutes for Research*

Adaptive survey designs often aim to mitigate nonresponse bias by targeting interventions (such as incentives) to “high priority” subsets of the sample. A challenge in developing adaptive designs is identifying effective interventions for low-response-propensity sample members—who, by definition, are the most difficult to persuade to respond, but whose responses are most needed to reduce nonresponse bias. This challenge is particularly salient in self-administered surveys, in which effective adaptive design interventions for interviewer-administered surveys—such as case prioritization and interviewer bonuses—are not available. Using an experiment incorporated into the 2019 National Household Education Survey (NHES:2019), this paper will evaluate whether a “choice plus” design (Biemer et al. 2018)—in which respondents are offered a choice of modes along with a promised incentive contingent on response by web—holds promise as an adaptive design intervention in address-based, self-administered studies. Preliminary NHES:2019 evaluations found that the choice plus design achieved a higher overall response rate than a sequential mixed-mode control; and, furthermore, that low-response-propensity subgroups showed higher-than-average increases in response rates, relative to higher-response-propensity subgroups. This suggests that using choice plus for only lower-propensity cases could improve sample representativeness to a similar degree as using choice plus for all cases. Building upon these findings, this paper will use data from the NHES:2019 experiment to project the impact of an adaptive design using choice plus as a targeted intervention for low-propensity cases. A Monte Carlo simulation will be used to evaluate the potential for reduced nonresponse bias, reduced weighting variance, and/or a shortened field period, compared to non-adaptive designs that use the sequential mixed-mode control or the choice plus treatment for all cases. By evaluating the utility of the choice plus protocol as an adaptive design intervention, this paper will contribute to ongoing efforts to incorporate adaptive design principles into address-based studies.

Experiments to Inform Refinement to the Choice+ Protocol for the 2020 Residential Energy Consumption Survey

Kristine Wiant, *RTI International*

Paul Biemer, *RTI International*

Joe Murphy, *RTI International*

S. Grace Deng, *U.S. Energy Information Administration*

The Residential Energy Consumption Survey (RECS), collects housing characteristics and householder behavior data integral for understanding energy usage trends across the United States. Conducted by the U.S. Energy Information Administration and initiated in 1978, the fourteenth iteration (2015 RECS) was the first to include self-administered modes, namely mail and web. The 2020 RECS is the largest in the program’s history, and the first iteration to use entirely self-administered modes. The 2020 RECS employs an adaptive, phased approach to data collection. The sample of approximately 18,000 completed interviews is stratified by state and then partitioned into three phases. Phase 1 consists of about 3,600 completed interviews and includes several experiments designed to inform potential improvements to the Choice+ protocol used in the 2015 RECS (Biemer et al., 2017). Phase 2 of data collection, accounting for 80% of the Phase 1 and 2 sample, will then use an improved version of the protocol, as determined by Phase 1 results. Three experiments will be conducted during Phase 1 data collection to determine 1) whether a \$30 promised incentive for web response (instead of \$10 for paper) generates significantly more web responses than a \$20 promised incentive for web response, 2) whether “official” looking black and white postcards yield a higher submission rate than postcards with color and graphics, and 3) whether a 4 page nonresponse follow-up questionnaire can achieve a submission rate on par with a 2-page questionnaire alternative. We will first discuss the motivation for the experiments in terms of optimizing data quality, representativity, proportion responding via web vs. paper, and costs. We will then report the results of the Phase 1 experiments and the decisions and outcomes from Phase 2. We will conclude with the implications of these findings for other large-scale web and mail data collection efforts.

You've Got Mail: Estimating the Effects of Pre-Notification Postcards and Monetary Incentives on Respondent Recruitment

Graham Jones, *U.S. Bureau of Labor Statistics*

Daniel Dorfman, *U.S. Bureau of Labor Statistics*

Laura Erhard, *U.S. Bureau of Labor Statistics*

Parvati Krishnamurty, *U.S. Bureau of Labor Statistics*

Brandon Kopp, *U.S. Bureau of Labor Statistics*

Brett McBride, *U.S. Bureau of Labor Statistics*

From October 2019 to March 2020, the Bureau of Labor Statistics' Consumer Expenditure Survey (CE) program fielded a large-scale online diary feasibility test (LSF) as part of the multi-year Gemini Project to redesign the survey. One of the goals of the LSF was to examine the effectiveness of respondent recruitment methods, specifically pre-notification postcards and monetary incentives. As part of an embedded experiment, a \$5 cash incentive was included with the advance letter for a randomly selected half of the sample and sent out by Priority Mail. The entire incentive group and a random half of the control group were mailed a pre-notification postcard, about a week prior to the advance letter mail out. The postcard was meant to give respondents advance notice about the survey and prevent them from discarding the advance letter that contained the cash incentive upon arrival. Monetary incentives were tested in an earlier phase of the Gemini Project, and these prepaid cash incentives showed promise in increasing participation likelihood in the CE Interview Survey (Elkin, McBride, Steinberg, 2018). Pre-notification postcards have also been found to increase participation rates in other surveys (M. Koitsalu, et al. 2018). To estimate the effects of both methods, the participation outcomes of the consumer units (CUs) assigned to the postcard group (n=1,332) and the monetary incentive group (n=879) were analyzed against the outcomes of untreated CUs, using LSF data collected through February 2020. The resulting analysis showed that neither the postcard nor the monetary incentive group exhibited statistically different participation outcomes from their respective controls, indicating that both methods were an ineffective means of increasing participation. These results will inform decisions regarding recruitment methods to be used in the CE and may be of interest to survey methodologists considering the use of incentives and postcards in other household surveys.

Effect of Contact Method on Response Rates in a Mixed-Mode Physician Survey

Geraldine Haile, *Mathematica*

Jared Coopersmith, *Mathematica*

Alexandra Saunders, *Mathematica*

Hanzhi Zhou, *Mathematica*

Karen Bogen, *Mathematica*

Nancy Duda, *Mathematica*

Deborah Peikes, *Mathematica*

While recent survey literature supports mixed-mode survey administration to improve physician survey response rates, there is limited evidence on how contact method affects physician choice of survey mode. The ability to push physicians to complete a survey on the web has the potential to save on mailing, quality control, and data entry costs.

We examined the effect of having physician email addresses on response rate for the 2019 Comprehensive Primary Care Plus (CPC+) Physician Survey, which was fielded to 8,235 physicians from CPC+ and comparison practices to support the evaluation of CPC+, a Centers for Medicare & Medicaid Services initiative designed to improve primary care. We also examined response rate differences by selected physician subgroups to further assess the impact of contact method on survey completion.

Mathematica designed and administered the instrument as a 20-week web survey with 3 hard copy questionnaire mailings. Physicians with email addresses received six reminder emails, while those without email addresses received five postcard mailings. We obtained email addresses for 82% of the physicians from a health care data organization. Respondents with an email address had higher rates of overall response (60% vs 54%) and web response (52% vs 44%) than respondents with no email address. Similarly, respondents with no email address completed hard copy surveys at a slightly higher

rate than those with an email address (11% vs 9%). Hard copy survey returns represented 15% of all completed surveys. This demonstrates that hard copy surveys are still an important aspect of data collection among physicians, even when email contact is an option.

Will Web Response Ever Catch Up to Paper?

Teresa Koenig, *Westat*

Sarah Bennett-Harper, *Westat*

Martha Stapleton, *Westat*

Jocelyn Newsome, *Westat*

Kerry Levin, *Westat*

Scott Leary, *Internal Revenue Service*

Rizwan Javaid, *Internal Revenue Service*

Pat Langetieg, *Internal Revenue Service*

Brenda Schafer, *Internal Revenue Service*

The use of multiple modes of data collection has become standard practice in survey research (Dillman, Smyth, & Christian, 2014). Within the multi-mode environment, researchers continue to seek ways to encourage response via web, since it speeds up data collection and results in significant cost savings (de Leeuw & Toepoel, 2018; Tourangeau, 2017). However, pushing respondents to the web can be challenging and runs the risk of depressing overall response (Messer & Dillman, 2011; Sun et al., 2020). The Internal Revenue Service (IRS) routinely surveys taxpayers to gather information about the time and money spent complying with tax reporting responsibilities. These burden surveys have typically used a concurrent multi-mode approach that emphasizes a paper questionnaire while simultaneously offering a web option. Though web responses have steadily increased over the years, most respondents still choose to respond by paper (Stapleton et al., 2020). The IRS continually explores options to encourage web response using a multi-mode design. This paper presents findings from an experiment with the IRS Taxpayer Compliance Burden (TCB) Survey, one of the burden household surveys. The experiment compared two methods of presenting a web option to respondents. In the sequential web-push method, respondents were initially presented with a web option at early contacts. Those who chose not to respond by web were later sent a paper survey. In the concurrent method, respondents were sent paper surveys with the web option mentioned in a call-out box on the inside cover of the questionnaire. In addition, web login instructions were provided on the back cover. Our analysis will compare response rates, percentage of web response, and any impact on nonresponse bias.

Evaluating the Effectiveness of Advance Mailings in a Web-Push Design

Jiashan Cui, *American Institutes for Research*

Mahi Megra, *American Institutes for Research*

Tzu-Jou (Carol) Wan, *American Institutes for Research*

Advance letters have been shown to increase survey response rates by alerting potential respondents about the upcoming survey request (de Leeuw et al, 2007; Spruyt and Droogenbroeck, 2014). In line with this expectation, the National Household Education Survey (NHES), an address-based household study that predominately uses a web-push design, typically sends households an advance letter a week before sending the initial survey request. In the 2019 administration, an experiment was included (1) to assess whether sample member engagement could be increased by using a more intensive advance mailing campaign that increases familiarity with the NHES prior to notifying the household about the survey request; and (2) to confirm whether there is a benefit to sending an advance letter in an NHES web-push design. For this experiment, sample members were randomly assigned to an advance mailing campaign condition, an advance letter-only condition, or a no-advance mailings condition. Within the advance mailing campaign condition, sample members received two oversized, color postcards prior to the advance letter (for a total of three advance mailings). These postcards presented interesting statistics from past NHES administrations, but they did not mention that the household had been sampled for NHES:2019. It was hypothesized that the advance mailing campaign would increase the response rate to earlier mailings as compared to both the advance letter-only and no-advance mailings condition by increasing awareness about the NHES (Lavrakas et al, 2004).

We will present, within the context of a web-push design, how the advance letter and mailing campaign motivated survey participation at different stages of administration. We will also assess whether the effect on response rates and response mode varied by selected household characteristics. Our presentation will provide insights on the effectiveness of advance mailings in a web-push design and inform decisions on the ideal number and types of advance mailings.

Experimenting With QR Codes in a Statewide Push-to-Web Survey

Kyle Endres, *University of Northern Iowa*

Erin O. Heiden, *University of Northern Iowa*

Mary O. Losch, *University of Northern Iowa*

Survey researchers are continually evaluating mechanisms to increase response rates and reduce respondent burden while maintaining data quality. Quick Response (QR) codes are a design feature that can easily be incorporated into mailed invitations for push-to-web surveys to potentially facilitate participation. In this study, we randomize the inclusion of a personalized QR code in the mailed invitation for a statewide push-to-web survey. We evaluate differences in response rates, data quality, device type used by participants, and the perceived respondent burden between subjects who were mailed invitations with a personalized URL and subjects who were mailed invitations with both a personalized URL and personalized QR code.

Survey Deduplication: Assessment and Decision-Making From a Multi-Mode Longitudinal Cohort Survey

Laura DiGrande, *RTI International*

Rita Thissen, *RTI International*

Mai Nguyen, *RTI International*

Taylor Abernathy, *RTI International*

David Wu, *NYC Department of Health and Mental Hygiene*

Junaid Maqsood, *NYC Department of Health and Mental Hygiene*

Shengchao Yu, *NYC Department of Health and Mental Hygiene*

Multi-mode scientific surveys face a challenge with multiple submissions by some respondents. Regardless of the care taken in advance, vigilance is required to detect multiple survey responses for a single sample member and determine which responses to keep. Such multiples are often called duplicates, although the data may not match exactly.

The World Trade Center Health Registry supports a longitudinal survey that monitors health status among people highly exposed to the September 11, 2001, terrorist attacks. Since 2003, the New York City Department of Health and Mental Hygiene has collected information from registrants through multi-mode surveys to obtain information on exposures and health status. RTI International supported the health department with the fifth wave of data collection, which was launched in April 2020 and continues through January 2021.

The Registry's multi-mode survey format has the potential for unnoticed duplicates, which can artificially increase statistical power, decrease variance, and result in smaller confidence intervals. This presentation shares how Wave 5 protocols detected, evaluated, and resolved multiple survey submissions about the same sample member. To prevent and detect duplicates, we tracked and cross-referenced survey returns for each mode. We categorize duplicate surveys by several characteristics including dates of receipt, survey mode, demographics (i.e., age, gender, language), and registrant exposure group (i.e., lower Manhattan residents, rescue/recovery workers). We analyze discordant responses across duplicates to determine which types of questions are most subject to potential measurement error. Finally, we present methods to identify duplicates completed by a non-registrant (proxy), different registrants, or the same registrant twice, as well as decision-making rules for determining the survey of record.

Because duplicates are a common challenge to survey research, the information shared in this presentation will offer practical information that can be applied by others to ameliorate similar problems, reduce error, and increase measurement quality.

5/13/2021, 1:30 PM - 3:00 PM
Concurrent Session F: Live Sessions

Mini-Conference: A Meeting Place and More...: A History of the American Association for Public Opinion Research

A Meeting Place and More...: A History of the American Association for Public Opinion Research

Tom Smith, *NORC at the University of Chicago*

Kathleen A. Frankovic, *Consultant*

Scott Keeter, *Pew Research Center*

Jeff Hackett, *NORC at the University of Chicago*

Timothy Johnson, *University of Illinois Chicago*

This panel will feature many of the 18 authors of the updated AAPOR history. It will do three things: 1) introduce the book to the AAPOR community, 2) discuss some of the main themes of change within AAPOR that have been covered within and across the 16 chapters, and 3) consider what our history tells us about our present state and possible futures.

Mini-Conference: Understanding the Electorate Through Pre-election Polls

Improving Likely Voter Models With Cultural Worldview Data

Natalie Jackson, *PRRI*

Pre-election polls in 2020 systematically underestimated Donald Trump's vote share. National polls that PRRI conducted in September and October were no different; both polls underestimated Trump's vote share, and to a lesser degree overestimated Joe Biden's vote share. In evaluating why this occurred, we looked to both the sample composition and likely voter modeling. While sample issues should be attended to, we show that relying less on self-reported propensity to turn out to vote and more on individual respondent characteristics can greatly improve likely voter modeling. This paper will show the results of multiple tests of constructing probabilistic likely voter models that incorporate respondents' cultural worldviews - measures of how respondents think the world should work and how they want to relate to other people. One such exploration finds that the fatalist cultural worldview is particularly useful in this exercise, as fatalists believe everything happens by chance and luck, and that nothing can change what will happen. Most fatalists say they will turn out to vote, but their worldview is one that does not comport with going out of their way to vote. Adjusting fatalists' propensity to vote results in a sample more in line with preliminary exit poll data and puts both Trump and Biden vote shares very near actual vote counts. Other models will investigate how to use cultural worldview data to inform our likely voter models and explain exactly what self-reported likelihood to vote fails to show: how likely someone is to go out of their way to vote in an election.

Latinos in the 2020 Presidential Election: Was There a Trump Effect?

Ana Gonzalez-Barrera, *Pew Research Center*

Luis Noe-Bustamante, *Pew Research Center*

The 2020 election placed renewed focus on the record 32 million Hispanic eligible voters and their political values and attitudes, and the reactions the current presidential administration could have had in engaging them electorally or not. At the beginning of his presidential campaign, President Donald Trump made provocative comments about Mexican and Muslim immigrants and Hispanics in general. The issue of immigration was also a focal point of Trump's administration from the beginning, an issue that is relevant to the Latino community. The 2020 presidential election drew increased interest in the election among Latinos and increased enthusiasm for voting, among those already registered to do so, but did Trump's policies and comments motivate Latino voters more than in past presidential election years? This analysis will be based on a number of Pew Research Center surveys done prior to the election, media exit polls done during the election and the latest data from BLS and Census Bureau on civic engagement. Findings highlight that most Hispanics were downbeat about their situation during the first year of the coronavirus outbreak, and uncertain about their place in America under Trump, long before the pandemic hit. A large majority of them disapproved of the way Trump handled his job as president, and there was an increased interest in the election from all political sides of the spectrum. Early analysis of exit poll results shows a strong effect on voter engagement or voter turnout, unlike the 2016 election. Despite issues with exit polling this time around, it is almost certain that in 2020 the number of Hispanic voters reached a record number, and voter turnout rate might have increased to record levels of participation.

Shyness? Mode Effects? Self-Selection? Lessons From a Large Mixed-Mode Pre-Election Study

Eran Ben-Porath, *SSRS*

Ashley Kirzinger, *Kaiser Family Foundation*

Audrey Kearney, *Kaiser Family Foundation*

Emily Hachey, *SSRS*

The debate about self-reported or misreported voting intention has become intertwined with a discussion of mode effects. The basic logic of the argument is that if respondents are embarrassed about their vote-choice, this shyness is likelier to be present when interviewers ask about voting intention than when the question is self-administered. In this paper we test the shy-voter proposition by looking at a large (N=3,479) three-state study carried out before the 2020 general election, in which data were collected from respondents either by phone (n=363) or online (n=3,116). Response-mode was self-selected, meaning, respondents could choose to go online or to call into a dedicated phone line, and some were ultimately called by interviewers. In our analysis we address this type of self-selection. We find that controlling for factors associated with self-selection, respondents completing the survey by phone were likelier likely to indicate an intention to vote for Joe Biden. While this finding is consistent with the shy-voter framework, it is far from definitive. We consider alternative explanations to our findings, and evidence that runs contrary to this framework.

VoteCast Explains: The 2020 Electorate

Emily Swanson, *The Associated Press*

Hannah Fingerhut, *The Associated Press*

With unparalleled depth, AP VoteCast provides a rich and nuanced portrait of U.S. voters. Conducted for The Associated Press by NORC at the University of Chicago, VoteCast combines probability-based interviews, drawn from state voter lists, with non-probability interviews from opt-in online panels using a cutting-edge calibration method. The survey included more than 133,000 registered voters nationwide, with more than 110,000 likely voters. (VoteCast methodologies will be explained in a separate paper.) The resulting research allows for a deep understanding of voter attitudes as they cast their ballots, as well as subgroups in the electorate that are rarely captured, both nationally and at a state level. This paper will offer insight into the opinions and voting behavior of the American electorate in the 2020 presidential election. Voters for President Donald Trump and Democratic rival Joe Biden had fundamentally different views of the nation after an untraditional four years under Trump, of the worst public health crisis in a century, and even of voting. The paper will dissect how Biden narrowly edged out Trump in key states, relying on dissatisfaction among voters driven by the pandemic. But the data will also show why Trump remains a force to be reckoned with in American politics. This paper also will analyze the Georgia runoff electorate to understand how Democrats were able to mobilize voters in the state, particularly Black voters, to eek out a narrow Senate majority. It will also help explain dynamics of the transition between the Trump and Biden administrations, and Republican loyalty to Trump.

How Could Polling Have Better Told the Story of the Red Mirage and the Blue Wave in the 2020 Election?

Laura Wronski, *SurveyMonkey*

Jon Cohen, *SurveyMonkey*

Jack Chen, *SurveyMonkey*

Xinyue Zhang, *SurveyMonkey*

The final tally in 2020 turned out to be something of a landslide for Joe Biden, who received 306 Electoral College votes to Donald Trump's 232 and more than 78 million of the 152 million votes cast--a record high. Biden had a consistent lead over Trump in the polling averages, both at the national level and among the states that were up for grabs from the time he first emerged from the Democratic primary--and yet, there was widespread sense on Election Night and the subsequent four days before Biden declared victory that the polls had been misleading. This paper will use SurveyMonkey's 50-state plus DC election polling data, spanning from June 2020 through the post-election period, to examine how the horse race polls could have been better presented to demonstrate the uncertainty and the time-lapse driven by mail-in vote counting that they accurately predicted but that failed to be conveyed to the public. In particular, we'll focus on the partisan disparities in mail-in versus in-person voting in several states—Florida, Georgia, North Carolina, Pennsylvania, California, Arizona, Nevada, Wisconsin, and Michigan—that we had been tracking in the lead-up to the election. We'll create probability models to estimate the likelihood of a Biden vs. Trump win at each stage of vote counting in each state, and we'll demonstrate new data visualization tools that would have made this election more digestible for the public.

Mini-Conference: COVID-19 and the Decennial Census

Addressing Analytical Challenges Due to COVID-19 Impacts on Data Collection for 2020 Census Experiments

Jennifer Hutnick, *U.S. Census Bureau*

Michael Shaw, *U.S. Census Bureau*

Julia Coombs, *U.S. Census Bureau*

Kelly Mathews, *U.S. Census Bureau*

Fred Lestina, *U.S. Census Bureau*

Laura Norena, *U.S. Census Bureau*

Austin Schwoegl, *U.S. Census Bureau*

The COVID-19 pandemic caused a dramatic change in the operational plans for the 2020 Census. Embedded within the 2020 Census operations were a series of experiments and evaluations designed to test innovations from the 2010 Census and for the 2030 Census. These experiments and evaluations were planned to take advantage of the unique environment of the 2020 Census. As the 2020 Census adapted in response to the COVID-19 pandemic, so did the experiments and evaluations. The experiments and evaluations were impacted in a variety of ways including changes in the mailing strategy, to the environment being leveraged, to the research designs, and in analytical approaches. This presentation summarizes the changes made to the 2020 Census experiments and evaluations due to COVID-19. It highlights the analytical challenges faced due to these changes, how changes to experiments and evaluations are expected to impact final analysis, and proposed paths forward.

Managing Communications During the 2020 Census Advertising Campaign

Anna Sandoval, *U.S. Census Bureau*

Maria Olmedo-Malagon, *U.S. Census Bureau*

Jennifer Diaz, *Deloitte*

The Integrated Partnership and Communications operation communicates the importance of participating in the 2020 Census to the entire population of the 50 states, the District of Columbia, and United States territories to: Engage and motivate people to self-respond, preferably through the internet. Raise and keep awareness high throughout the entire 2020 Census to encourage response. Support field recruitment efforts for a diverse, qualified census workforce. Support dissemination of census data to stakeholders and the public. In June 2017, the U.S. Census Bureau made public Version 1.0 of the 2020 Census Integrated Communications Plan, identifying the path forward for future communications efforts. This laid out foundational steps to develop a comprehensive communications campaign while providing a high-level overview of anticipated communications strategies. The Census Bureau planned to promote self-response by taking advantage of new and improving technologies. Internet tools, increasing cellular coverage, and smartphone use have opened new communications channels and media. This plan accounted for a number of challenges, however it did not account for the impact of a global pandemic. In March 2020 the COVID pandemic triggered changes to the communications plan. In this paper we summarize the adjustments that took place as a result of the pandemic on 2020 Census operations, the ways in which the existing research was used, and the lessons learned from this experience.

Adapting a Mailing Experiment in Real Time: Changes to Mailing Strategies for 2020 Census Experiments Due to COVID-19

Julia Coombs, *U.S. Census Bureau*

Jennifer Hutnick, *U.S. Census Bureau*

Kelly Mathews, *U.S. Census Bureau*

Austin Schwoegl, *U.S. Census Bureau*

During the 2020 Census, a series of experiments were planned to test changes to the mailing materials sent to households to increase self-response rates. The experiments were designed to mirror the mailing strategy of the primary 2020 Census operations where applicable. During the middle of the mailing operations for these experiments, the COVID-19 pandemic caused the facility used to print, assemble, and mail the materials for these experiments to shut down, which prevented the mailing of experimental

packages. This presentation describes how COVID-19 affected the mailout plans for the 2020 Census experiments. It includes a discussion on the efforts taken to adapt the mailing strategy in real time to maintain as much of the original experimental design as possible and challenges encountered as the 2020 Census operation adapted in response to COVID-19.

2020 Census Bilingual Enumerator Experiences in the Field

Patricia Goerman, *U.S. Census Bureau*

The U.S. Census Bureau has been engaged in research about multilingual doorstep interview messaging for many years. As a part of the 2020 Census, we planned field observations with census enumerators who spoke six different languages to understand their messaging and interactions with non-English speaking respondents. We also planned to do 5-minute debriefings with respondents after they completed the census interview. This work was designed to advance research into how to improve and enhance training and non-English materials for interviewers who speak different languages. The COVID-19 pandemic changed our plans dramatically and we quickly pivoted our methodology and the focus of our study. We instead conducted telephone focus groups with enumerators who spoke English alone or one of six additional languages: Spanish, Chinese, Vietnamese, Korean, Russian or Arabic. We asked interviewers about the success of doorstep messaging, use of translated census materials at the doorstep, undercount of children, respondent confidentiality concerns, and any impacts of COVID-19 on respondent participation. This presentation will focus on methodological considerations and lessons learned related to conducting telephone debriefing focus groups with field interviewers. In addition, we will share some preliminary findings from the focus groups during COVID-19 and discuss similarities and differences across interviewers who spoke the different languages included in our study.

Studying Privacy and Confidentiality in the 2020 Census During the COVID-19 Pandemic

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

Rodney Terry, *U.S. Census Bureau*

Matthew Virgile, *U.S. Census Bureau*

As part of the 2020 Census Experiments and Evaluation program, we conducted a mixed-mode evaluation of the privacy and confidentiality concerns of respondents around the 2020 Census. Beginning in August 2020, we fielded a mail-out survey to about 110,000 households who had self-responded to the 2020 Census. A sub-sample of non-respondents were followed-up with a telephone interview. We also sampled 10,000 households who had responded to the census with an enumerator. Our original plan was to conduct face-to-face interviews with this group, however many were conducted over the phone due to the covid-19 pandemic. We also sought to capture the opinions of hard-to-count groups but we knew they would be less likely to respond to a follow-up survey about their census experience. Thus, we turned to another source – the Partnership Program in which Census enlists trusted partners in the local communities to conduct outreach. We conducted focus groups and in-depth interviews with partners associated with traditionally hard-to-count populations. In this presentation, we describe how our research plans were affected by the covid-19 pandemic and how we adapted our design to the situation. Using qualitative data, we explore how the pandemic context might have affected broader opinions on privacy and confidentiality in the 2020 Census. Drawing on both the survey and partnership data, we present preliminary findings that paint a picture of respondent concerns about privacy and confidentiality in the 2020 Census. We conclude with implications for continuing exploration of privacy and confidentiality concerns about government data collection with emphasis on lessons learned for data collection in the COVID-19-era.

5/13/2021, 3:00 PM - 5:00 PM
Concurrent Session G: Prerecorded Sessions

Mini-Conference: Creative Adaptations for Data Collection During the Pandemic

Responding to the Challenges Conducting Cross-National Fieldwork in a Pandemic: A 3 Country Push to Web Experiment in Europe

Rory Fitzgerald, *City University London*
Don A Dillman, *Washington State University*
Peter Grand, *Institute for Advanced Studies*
Vera Messing, *Hungarian Academy of Sciences*
Victoria Ratti, *NatCen Social Research*
Bence Sagvari, *Hungarian Academy of Sciences*
Elissa Sibley, *City, University of London*
Dragan Stanojevic, *University of Belgrade*
Bojan Todosijevic, *University of Belgrade*

The European Social Survey (ESS) is a biennial, large scale, cross-national general social survey that has been conducted using in-home face-to-face interviews with a representative sample of the general population since 2002. ESS face-to-face interviewing has been paused in all countries for the remainder of 2020 following increased restrictions on movement and household mixing. Whilst it is hoped fieldwork might restart in 2021 there remains considerable uncertainty if face-to-face interviewing will be possible in all ESS countries in the short term.

In response to this the ESS has launched a push to web survey in 3 countries where there is little experience to date of using such an approach. Whilst push to web surveys are fairly common in countries with named person samples (eg Norway, Netherlands) or well developed survey infrastructures (eg UK), there are parts of Europe where push to web surveys and other self-completion are rarely if ever used for the general population. This paper will outline the design approach being used in Hungary (which uses an individual named sample but where there is limited experience in this method) and Austria and Serbia (which use address samples and require within household selection). In the countries using address samples the on-line questionnaire will check against the household roster to see if the target person has done the survey and if not the target respondent will also be invited using a novel approach. Once respondents have been offered the chance to do the survey by web, a paper self-completion is offered to non-respondents. The authors will summarise the effectiveness of the experiment in terms of cooperation and demographic representativeness. The particular design issues of conducting such a survey cross-nationally will be highlighted.

NSHAP COVID-19 Study: Transitioning an In-Person Survey of Older Adults to Multi-Mode Data Collection

Jennifer Satorius, *NORC at the University of Chicago*
Lauren Sedlak, *NORC at the University of Chicago*
Katie O'Doherty, *NORC at the University of Chicago*
Anna Wiencrot, *NORC at the University of Chicago*
Ned English, *NORC at the University of Chicago*
Erin Burgess, *NORC at the University of Chicago*
Becki Curtis, *NORC at the University of Chicago*
Evelyn Zepeda, *NORC at the University of Chicago*
Colm O'Muircheartaigh, *NORC at the University of Chicago*

The National Social Life, Health, and Aging Project (NSHAP) is a longitudinal, population-based field study of adults born between 1920 and 1965 and their romantic partners. NSHAP's overarching aim is to examine the relationship between social life and health among older adults. Data are traditionally collected every five years in the respondent's home using trained interviewers, with a smaller subset of data collected via a post-interview self-administered questionnaire. The fourth wave of in-person data

collection was scheduled for 2020, but preparations were paused due to the coronavirus pandemic. In the interim, NSHAP conducted a multi-mode survey about NSHAP respondents' experiences during the pandemic. The impetus for the COVID-19 Study was twofold: assess the impact of the pandemic on older adults' lives and evaluate the feasibility and effectiveness of using different modes of data collection with older adults. From September to December 2020, a combination of web, phone, and paper was used to administer a 20-minute survey to all NSHAP respondents who had participated in the most recent wave in 2015-16. The survey included a broad range of topics, such as COVID-19 diagnoses among respondents and social networks, health care, social activity, social unrest, and health behaviors. An experiment to test the effectiveness of incentive type (monetary v. non-monetary), amount, and timing (pre- v. post-data collection) was conducted. In this paper, we will describe the multi-mode design, examining response to the survey across modes, by various respondent characteristics (e.g., age, gender, race and ethnicity), and by incentive treatment group. We will also assess survey non-response follow-up and the effect of mode on item non-response and response to sensitive items. Finally, we will evaluate how our experiences may inform the collection of NSHAP data in future rounds and in other panel studies of older adults considering the adoption of multi-mode data collection.

Population-Based Estimates of COVID-19 Incidence in New York City: A Comparison of Two Surveillance Systems

Kathryn Peebles, *Centers for Disease Control and Prevention*

Michael Witt, *Abt Associates*

Michael L. Sanderson, *New York City Department of Health and Mental Hygiene*

Sharon Perlman, *New York City Department of Health and Mental Hygiene*

Steven Fernandez, *New York City Department of Health and Mental Hygiene*

Sarah Dumas, *New York City Department of Health and Mental Hygiene*

Karen A. Alroy, *Centers for Disease Control and Prevention*

Andrew Burkey, *Abt Associates*

Nicholas Ruther, *Abt Associates*

COVID-19 cases have been underestimated throughout the SARS-CoV-2 pandemic due to limited testing capacity and inadequate identification of cases with only mild symptoms. Population-based estimates of COVID-19 incidence are needed; approaches that rely solely on laboratory testing are logistically challenging and costly to conduct on an ongoing basis. Survey-based elicitation of COVID-19-like illness (CLI) symptoms, adjusted for emergency department-based estimates of the fraction of CLI cases that represent true SARS-CoV-2 infections, offers a feasible alternative for obtaining ongoing population-based estimates of COVID-19 incidence. The New York City (NYC) Department of Health and Mental Hygiene identified two approaches to collecting these data. First, we modified data collection for the Community Health Survey (CHS), an annual population-based random digit dial telephone survey of non-institutionalized adults in NYC, to add items assessing COVID-19 signs and symptoms from March through August 2020. Second, we accelerated the establishment of Healthy NYC, a representative panel of NYC residents for population health research, to collect CLI symptom data during months not covered by CHS. Panel members were recruited from an address-based sample frame and from participants in other Health Department surveys, and questionnaires were administered via web and telephone. Data for both surveys were weighted to account for the probability of selection and raked to population control totals from the American Community Survey, a United States Census Bureau survey. Data collection via CHS and Healthy NYC overlapped during the month of August 2020 to evaluate the effect of the different survey sampling frames and modes on estimates of COVID-19 incidence. Findings will indicate whether comparable population-based estimates of COVID-19 incidence can be obtained from existing and planned survey data collection despite different survey sampling frames and modes.

Effects of Severe Weather and COVID-19 pandemic on Survey Response Rates and Survey Methods for the Behavioral Risk Factor Surveillance System (BRFSS)

Gary Lerner, *Abt Associates*

Dennis Daly, *Abt Associates*

Matthew Anderson, *Abt Associates*

Since 2016, several historic and natural events have created a substantial impact on survey response rates, interviewing, sample management, staffing and other components of survey research. First, dangerous weather systems caused by climate change have increased in intensity and become more frequent. Secondly, the COVID-19 pandemic in 2020 affected survey research projects by adapting to new methods of interviewing. The increased frequency of severe weather events and public health emergencies have required survey data collectors to be adaptable to extreme conditions that impact the ability to reach and engage respondents. This methodological brief presents information on how Abt Associates (Abt) modified data collection protocols on the Behavioral Risk Factor Surveillance System (BRFSS) in response to multiple hurricanes and the COVID-19 pandemic.

Hurricane Matthew (October 2016), Hurricane Irma (September 2017), Hurricane Michael (October 2018), and Hurricane Dorian (September 2019) devastated parts of Florida and resulted in millions of dollars in damage. In lieu of these impending hurricanes, Abt used NOAA tracking tools to inform the suspension of dialing samples for the counties in the paths of the storms. After assessing the damage and reviewing local power outage and cell coverage information and consulting the Florida Department of Health, Abt resumed data collection when reasonable and in the best interest of citizens and staff. During COVID-19 pandemic, in an effort to ensure the continuity of data collection protocols, Abt transitioned on-site interviewers to remote interviewing and quickly stood-up a virtual call center (VCC) allowing interviewers to safely and securely from home. We discuss the decision making process behind suspending and resuming interviewing during extreme weather, and present information on our efforts on managing the VCC. To examine the impact of these decisions and adjustments, we will examine the results of our efforts by exploring response rates and other indicators of data quality

USPS Mail Delivery Patterns for a National Address-Based Sample in the Time of COVID-19

Raphael Nishimura, *University of Michigan*

Rebecca Gatward, *University of Michigan*

Brady West, *University of Michigan*

Htay-Wah Saw, *University of Michigan*

Many survey data collection operations have changed modes as a result of the COVID-19 pandemic. Surveys using mailed letters and/or reminders at any point in a data collection protocol may have been adversely affected by staff shortages and other equipment issues affecting the delivery times of the USPS. In this study, we analyzed daily USPS mail delivery reports for each of 21 National Distribution Centers (NDCs) for two specific mailings that were sent to a national sample of addresses as part of the American Family Health Study. We linked concurrent aggregate rates of COVID-19 infection and mortality to each of the NDCs, and then examined the variability in cumulative delivery rates across the NDCs as a function of these aggregate rates. We also collected paradata on whether any of the mailings sent to a sampled address were returned as undeliverable, and then modeled the probability of any undeliverable mailing as a function of auxiliary variables available for the full national sample, including the aforementioned COVID-19 rates. Considering the cumulative delivery trajectories for the 21 NDCs, we found that some NDCs lagged behind others in delivery rates, but we did not find any evidence of consistent variability in the cumulative delivery rates as a function of the COVID-19 case and mortality rates. In our models of undeliverable mail, we found that counties with a higher COVID-19 mortality rate actually had marginally reduced odds of undeliverable mailings, and that addresses flagged as vacant or seasonal or missing linked commercial information were significantly more likely to have their mail returned as undelivered. Collectively, our results suggest that area-level rates of COVID-19 infection and mortality are not strong predictors of mail delivery problems when taking other relevant auxiliary variables into account.

Hospital Heroes: How Honoring Physicians Through Tailored Messaging and Materials Impacts Response Rates

Rebekah Torcasso Sanchez, *RTI International*

Rebecca J. Powell, *RTI International*

Marshica Stanley, *RTI International*

Murrey G. Olmsted, *RTI International*

The principle argument of leverage-salience theory is that survey design attributes will influence cooperation decisions differently for different people. Survey research validates the efficacy of tailored design strategies to increase response rates and encourage higher quality participation by survey respondents. Researchers have conducted experiments varying messaging techniques to gain cooperation through specific appeals to values such as expertise (Leece et al, 2006), social exchange (Dillman, Smyth, and Christian 2014), or in-group membership (Lynn, 2016). We aim to expand the existing literature by exploring the impact of tailored messaging and imagery on materials in a survey of physicians.

Due to the overwhelming impact of COVID-19 on the healthcare industry in 2020, we wanted to appeal to the physicians' status as "heroes". We conducted an experimental comparison between our standard survey letters and token incentives and a version incorporating the "Hospital Heroes" messaging and imagery that the survey sponsor uses in other products. In order to isolate the affect of the "heroes" language on our outcome variables, we updated our standard messaging to also include an acknowledgment of the COVID-19 pandemic but refrained from wording that labeled participants as "heroes". We hypothesized that the "Hospital Heroes" theme would yield higher response rates than an appeal solely to their expertise. The experiment was implemented in the 2021 Physician Survey, a mail survey of physicians and a component of the U.S. News & World Report Best Hospitals and Best Children's Hospitals ranking. Physicians were randomly assigned to either condition (n = 4,850). To test our hypothesis, we compare response rates and time to completion both overall and by physician specialty subgroups. This work contributes to the literature on leverage-saliency theory and how best to encourage participation in physician surveys by appealing to physicians' societal value.

Sampling, Recruiting and Frame Construction for Telephone Surveys

The Mixing Bowl: Analyzing Blends of Mixed-Mode Studies to Reduce Design Effect and Maintain Accuracy

David Taylor, *Roanoke College*
Cesar Bonilla, *Roanoke College*
Jordan Harris, *Roanoke College*
Alice Kassens, *Roanoke College*
Harry Wilson, *Roanoke College*

As more survey organizations consider and use mixed-mode studies, the importance of studying design effect and accuracy in those studies rises. For the last two years, we have tried a variety of mixed-mode studies in order to reduce survey costs due to the increasing costs associated with CATI. In particular, push-to-web via address-based sampling (P2WABS), Redirected Inbound Call Sampling (RICS), and Ask Your Target Market (AYTM) have been used by Roanoke College to compliment CATI methodology in February 2018, February 2019, and November 2019; a subset of the questions, from the University of Michigan, used to compute the index of consumer sentiment, index of current economic conditions, and the index of consumer expectations were used for each study. In addition to reviewing raw and weighted statistics from each of these months, we use bootstrap techniques to resample varying blends of these studies to analyze the increase and decrease of design effect and accuracy when the ratio between modes is adjusted; for example, we look at how one-third CATI and two-thirds RICS performs as compared to two-thirds CATI and one-third RICS (along with a sliding scale between those benchmarks) or how mixtures of P2WABS and CATI complement each other to provide cost-effective and accurate data collection.

A Comparison of Address-Based Sampling and Redirected Inbound Calling Sample for General Population Surveys

Michelle Cavazos, *RTI International*
Burton Levine, *RTI International*
Ellen Coats, *RTI International*
Matthew Farrelly, *RTI International*

Redirected Inbound Calling Sample (RICS) is an emerging nonprobability sampling methodology. Calls that dial into a bank of otherwise not-in-use toll-free numbers are redirected for screening, recruitment, and data collection via interactive voice response (IVR) or live interviewer. Prior studies have compared RICS IVR demographics and study outcomes to dual-frame random digit dialing (DFRDD) and in-person household surveys. However, there has been no published research comparing RICS IVR to an address-based sample (ABS). This presentation addresses a knowledge gap in quantifying the bias of RICS IVR compared to an ABS with respect to basic respondent demographics and study outcomes. We fielded questionnaires designed for surveillance of health and tobacco-related outcomes to a national sample using two data collection methodologies: (1) an ABS with a mail recruitment and web or paper-and-pencil interview (PAPI) data collection (N=1,500); and (2) a RICS with IVR recruitment and data collection (N=1,800). We examine the representativeness of the respondents by comparing the demographic distributions of the U.S. population with the unweighted demographic distributions of the ABS and RICS respondents. Additionally, we examine bias by quantifying the differences in weighted study outcomes between the ABS data and the RICS data. Findings can be used to inform when RICS methodology is fit for purpose and inform the optimal allocation of resources when fielding two survey methodologies simultaneously.

Assessing the Importance of Out of Area Cell Phones in State Surveys

Seth Brohinsky, *Abt Associates*

As public opinion researchers deal with increasing reliance on cell phone sample in telephone-based surveys, evaluating the inclusion of out of area cell phone numbers (i.e. cell phones with an out-of-state area code that have a billing zip code in the targeted geography) in state-level surveys is critical to ensuring accurate sample representation. As cell phone mobility continues to increase, further complicated by the COVID-19 pandemic, understanding the bias implications of exclusion is critical for ensuring the representativeness of future state-level telephone surveys. Abt Associates has been at the forefront in developing strategies to deal with increased cell phone mobility. We recently conducted a series of state-level polls on behalf of ABC News and The Washington Post in seven battleground states across the country. To address the potential coverage bias associated with cell phone mobility, each survey included a sample of cell phones with out-of-state area codes, in addition to the traditional in-state cell and landline RDD samples. Therefore, in theory, each state sample covered the entire state adult population, including those residents with an out-of-state area code. This presentation will analyze the potential coverage bias implications of state-level polls that do not include a sample of out-of-state cell phones. This analysis will explore whether demographic and substantive differences exist between respondents with an out-of-state cell phone number and those with an in-state number, based on area code. If differences are found, it may still be argued that inclusion of out-of-state cell phones is unnecessary, that most of these respondents are covered by the landline frame, as many households have both a cell and landline phone. In response, this presentation will also evaluate the household phone status of out-of-state cell phone respondents, providing an additional key measure of the importance of their inclusion in state-level surveys.

The Variability in Sample Accuracy, Coverage And Efficiency of “Pulsing” Sample for European CATI Over Time

Alexandra Castillo, *Pew Research Center*

Martha McRoy, *Pew Research Center*

Given the rising costs of international CATI surveys, researchers look to increase efficiencies while maintaining data integrity. One approach to is to reduce the volume of unproductive numbers dialed by sending a signal or ‘pulse’ to identify and remove nonworking numbers from samples. Though efficiency may improve, it’s possible that this process of differentiating working from nonworking numbers may introduce error—that is, actual nonworking numbers are identified as working (which may decrease dialing efficiency) or actual working numbers are identified as nonworking (which may decrease population coverage and increase the risk of noncoverage bias). The variation of these efficiencies and errors over time is an added concern for annual studies valuing consistent designs. To evaluate the implications of pulsing accuracy over time, landline samples from seven European countries were initially pulsed and numbers flagged as either “working” or “not working”, and then all numbers were dialed using the full call design regardless of the pulsing flag outcome. Based on initial analyses of numbers dialed, the accuracy rate of pulsing — that is, pulsing correctly identifying working numbers as working and nonworking as nonworking when verified by an interviewer — was 80% in spring 2019 but 74% in summer 2020. While 21% of landline interviews were achieved on numbers that were flagged as nonworking by pulsing in 2019, only 10% were achieved on such numbers in summer 2020. Efficiency disagreements varied by nearly 10 percentage points across the years, while noncoverage disagreements were relatively stable. In addition to the 2019 and summer 2020 data, an additional fall 2020 data point is available for three European countries – France, Germany and the UK – facilitating further analyses of country-specific pulsing trends before and during the covid-19 pandemic, which will be a key feature of our presentation.

Telephone Sampling Frame Differences: Lessons Learned From an Experiment in Lebanon

Salma Al-Shami, *Princeton University*

Julie A. de Jong, *University of Michigan*

Zeina Mneimneh, *University of Michigan*

Michael Robbins, *Princeton University*

In recent years, the cost of face-to-face interviews has risen significantly across much of the world and telephone interviews and other modes have become increasingly appealing. This trend has been accelerated by the COVID pandemic, forcing many face-to-face surveys to adopt a telephone mode. While the availability of telephone sampling frames and the selection of probability samples have been less of an issue in many western countries, challenges remain in drawing a representative probability telephone sample in many lower-resource countries. Such challenges are driven by the feasibility to generate a random digit dial (RDD) sample, the difficulty of using external data to make the sample more efficient, and the lack of information on the coverage differences between landline and cell phones. Given these challenges, many researchers revert to using existing telephone panels whose coverage, selection and non-response properties are unknown. In this paper, we use data from an experiment conducted by the Arab Barometer in Lebanon. We first select and interview respondents by telephone using a pre-screened RDD sample of mobile phones. We also select and interview a second sample from an existing telephone panel maintained by a local firm, using the same instrument. The panel members consist of respondents who have previously participated in at least one face-to-face survey conducted by the firm within the previous 18 months, with representation of the different religious sects in Lebanon. While the list-assisted sample is hypothesized to be more representative of the telephone population in Lebanon, there are concerns that the sample will suffer from differential nonresponse by religious sects. Using data from this experiment, we will compare sociodemographic differences from the two samples to available benchmark data and assess their relative representation. We will also compare responses to substantive questions controlling for any demographic differences.

Effect of Using Multiple Data Collection Methods With an Address Based Sample

Michelle Cantave, *University of Baltimore*

Elias S. Nader, *University of Baltimore*

P. Ann Cotten, *University of Baltimore*

In June of 2020 the Schaefer Center for public policy conducted a survey of Baltimore City residents using an Address Based Sample. The sample was selected from high and low crime Census Tracts in the City. The sample of telephone numbers was phone matched by the sample vendor. Telephone numbers were obtained for about 65% of the sample. The records that were matched to a phone number were called to complete the survey and the numbers that were not phone matched, were sent a letter inviting them the either call into to complete the survey or to go online and completed the survey on the web. Of the phone matched records, those that resulted in a "bad number" call disposition, such as disconnected numbers or fax machines. were mailed the invitation letters. This methodology resulted it three groups: Group 1: Outbound Calls Only, Group 2: Mail Invitation Only Sample, and Group 3: Mail Invitation and Outbound Phone Calls. This presentation will analyze the demographic differences in the three groups and will identify if the method of contacting potential respondents can be used to target specific demographic groups. In addition, we will look at differences in opinions from the difference contact groups.

Applying Data Science to Analyze Text Data

Evaluating Political Parties: Criterion Validity of Open Questions With Requests for Text and Voice Answers

Jan Höhne, *University of Mannheim*

Konstantin Gavras, *University of Mannheim*

The rise of smartphone surveys, coupled with technological advancements, provide new ways for measuring respondents' political attitudes. For instance, smartphones offer the collection of voice instead of text answers by using the built-in microphone. This may facilitate answering questions with open answer formats and provide high-quality data. In addition, new developments in Automatic Speech Recognition (ASR) and text-as-data methods allow the proper handling of open answers from large-scale surveys. However, there is almost no research investigating the measurement quality of text and voice answers to open questions. In this study, we therefore compare the criterion validity between text and voice answers to open questions on four German political parties. For this purpose, we conducted an experiment in a smartphone survey (N = 2,402) in Germany and randomly assigned respondents to an answer format condition (text or voice). Voice answers were collected using the JavaScript- and PHP-based "SurveyVoice (SVoice)" tool that resembles the voice function of Instant-Messaging Services. We initially transcribed voice answers into text using Google's "Cloud Speech-to-Text API". We then conducted sentiment analyses to extract sentiment scores for each respondent and each target question on German political parties. This was done to obtain adequate measures that can be correlated with the measures of the criterion question on party preference. The results show that the open questions with requests for text and voice answers differ in terms of criterion validity. More specifically, they show that voice answers result in a somewhat higher criterion validity than their text counterparts. This is an important finding justifying open questions with requests for voice answers as a promising new method for measuring respondents' political attitudes in smartphone surveys. However, more refined research on the measurement quality of text and voice answers is required to draw robust conclusions.

Automated Emotion Recognition With Voice Data in Smartphone Surveys

Christoph Kern, *University of Mannheim*

Jan Karem Höhne, *University of Mannheim*

Konstantin Gavras, *University of Mannheim*

Stephan Schlosser, *University of Göttingen*

Technological developments for surveys on mobile devices, particularly smartphones, offer a variety of new forms and formats for collecting data for social science research. A recent example is the collection of voice answers via the built-in microphone of smartphones instead of text answers via the keyboard when asking open questions. Voice answer formats, compared to text answers, may decrease respondent burden by being less time-consuming and may provide richer data. Furthermore, advances in natural language processing and voice recognition do not only facilitate to process and analyze voice data, but also allow to utilize tonal cues, such as voice pitch, that can be used to study new research questions. Specifically, we explore the usage of pre-trained emotion recognition models for predicting the emotional states of respondents based on voice answers.

In this study, we use data from a smartphone survey (N = 1,200) that was collected in Germany in December 2019 and January 2020. To collect respondents' voice answers, we developed the JavaScript- and PHP-based "SurveyVoice (SVoice)" tool that records voice answers via the microphone of smartphones. We make use of the openEAR toolkit (Eyben et al. 2009) which allows us to predict respondents' emotional states based on voice data using pre-trained NLP models for emotion recognition. On this basis, we analyze the effects of (predicted) emotions on survey responses with respect to both content (e.g., sentiment of response) and quality (e.g., number of words, lexical complexity). This study exemplifies how to utilize new meta-information that can be extracted from voice data and adds to the research on the merits and limits of collecting voice answers in smartphone surveys.

Automatic Coding of Open-Ended Questions Into Multiple Classes: Whether and How to Use Double Coded Data

Matthias Schonlau, *University of Waterloo*

Zhoushanyue He, *University of Waterloo*

Responses to open-ended questions in surveys are usually coded into pre-specified classes, manually or automatically using a statistical learning algorithm. Automatic coding of open-ended responses relies on a set of manually coded responses, based on which a statistical learning model is fitted. In this paper, we investigate whether and how double coding can help improve the automatic classification of open-ended responses. We evaluate four strategies for training the statistical algorithm on double coded data, using experiments on simulated and real data. We find that, when the data are already double coded (i.e. double coding does not incur additional costs), double coding where an expert resolves intercoder disagreement leads to the greatest classification accuracy. However, when we have a fixed budget for manually coding, single coding is preferable unless the coding error is very large.

The Effect of Question Characteristics on Recurrent Neural Network Coding of Interviewer Question-Asking Behaviors

Jerry Timbrook, *RTI International*

Survey researchers commonly use behavior coding to identify whether interviewers read questions exactly as worded in the questionnaire (a potential source of interviewer variance). Recurrent Neural Networks (RNNs; a machine learning technique) can partially automate this coding with reliability levels comparable to humans (Timbrook and Eck 2019). RNNs learn to categorize sequential data (e.g., speech) based on patterns learned from previously categorized examples. However, it is unknown whether the reliability of RNN-based coding of question-asking behaviors differs across question characteristics. For example, because interviewers are more likely to misread longer, more complex questions, RNNs may need to learn more patterns in question-asking for these questions (e.g., interviewers adding/omitting words) versus shorter, easier-to-read questions; this may cause differences in RNN coding reliability across questions.

In this paper, I compare human and RNN coding of interviewer question-asking behaviors across four question characteristics: length, Flesch-Kincaid reading level, type (demographic, attitude/opinion, behavior), and comprehensibility using the Question Understanding Aid (QUAID). Humans transcribe and manually behavior code 33 questions from each interview in the Work and Leisure Today II telephone survey at the conversational turn level (25,442 question-asking turns) to identify when interviewers asked questions exactly as worded or with changes. With a random subset of interview transcripts as learning examples, I train RNNs to classify question-asking behaviors into these same categories. A random 10% subsample of transcripts (n=94) were also coded by experienced master coders to evaluate inter-coder reliability. I compare the reliability of human coding (versus the master coders) with the reliability of RNN coding (versus the master coders) across the four question characteristics. Preliminary results indicate that RNNs were more reliable when coding questions with shorter stems ($\kappa=.604$) versus longer stems ($\kappa=.401$). I conclude with implications for using RNNs to code question-asking behaviors across different question characteristics.

An Evaluation of IBM Debater: Assessment of Quantitative Text Summarization to Analyze Survey Verbatim Responses

Mika Jugovich, *IBM*

Bill Dusch, *IBM*

Tyler Waite, *IBM*

Jason Telner, *IBM*

Randal Ries, *IBM*

Is there anything that would improve your ability to feel more comfortable in the workplace during COVID-19 pandemic? Is there anything that would improve your experience or make you feel more comfortable when using the above clearance process? What one change would make you more likely to recommend? These are a small sample of open-end questions our team collects across hundreds of thousands of IBM employees each year. These questions generate tens of thousands of responses across a wide array of topics. With business leaders eager for insights and limited resources for conducting text analysis, a sound methodology and cutting-edge tools are needed to deliver impactful insights that help drive business decisions. This paper will focus on how IBM's Project Debater can be leverage for the rapid analysis of verbatim comments. First, we will begin with an overview of the IBM Debater tool, particularly its approach for utilizing text summarization to highlight the key points within a corpus of text documents. Second, we will present our methodology for text analysis and how we have integrated Debater. Finally, we will share the results of an evaluation of Debater, in terms of accuracy and amount of time to complete analysis. Accuracy was assessed using inter-rater reliability metrics to observe how closely the themes identified by a human researcher aligned with the output of the automated Debater analysis. The conclusion will discuss whether we feel Debater analysis is sufficient as a stand alone evaluation method, areas where we feel the algorithm may have room to improve, and suggestions to other researchers on how best to utilize this AI tool to speed up their ability to quickly identify keep pain points that their users, customers, and employees want them to know about.

Understanding Student COVID-19 Response From Reddit Data: A Structure Topic Modeling Approach

Jiaxi Li, *University of Michigan*

Chendi Zhao, *University of Michigan*

The COVID-19 pandemic has spread worldwide, impacting every aspect of daily life. In an emergency event of such kind that persists, it becomes increasingly important to have reliable, up-to-date emergency management assessments. The ability to accurately capture public needs and attitudes are critical to the success of emergency response plans. The traditional methods, such as surveys, have several pitfalls to understanding these issues. Most surveys introduce hypothetical bias since subjects' responses are measured under hypothetical scenarios. Besides, as response rates fall, costs rise, and the desire for timely data increases, social media are becoming increasingly popular as an alternative source of data for social science research. Reddit, the self-proclaimed "front page of the Internet" with its unique feature of subreddits, opens up potentials to study smaller geographies and rarer subpopulations. This study focuses on college students' responses to COVID-19 emergency management and the variability across time and groups in Michigan. Using Reddit data, we collect organic submissions in text formats from corresponding subreddits of the ten largest colleges and universities in Michigan for the period of March 1st, 2020, to October 31st, 2020. Taking a structure topic modeling approach, we derive COVID-19 related topics of our interest and present the student group's social media landscape. An analysis of topic prevalence and topic content indicates unmet needs on course housekeeping and needs beyond educational requirements. Our study gives school administrators validity to understand student concerns without validating privacy and sheds light on the implications of school emergency management plans on student responses.

Partisanship and Polarization

Just Like the Others: Party Differences, Perception, and Satisfaction With Democracy

Hannah Ridge, *Duke University*

A robust literature on citizens' satisfaction with democracy argues that system satisfaction is based on the policy outcomes citizens anticipate from electoral results. A tacit assumption in this research is that there are differences between the parties and that citizens are aware of the ideological and preference diversity in their political environment. Some citizens, however, fail to perceive these differences. Using a multi-national set of post-election surveys, regression analysis, and propensity score matching, perceived party difference is shown to substantially impact citizens' systemic satisfaction. Those who believe all parties are the same are substantially less satisfied with the functioning of their democracy. The negative effect of perceived party homogeneity is mitigated by closeness to a political party and sense of representation. By manipulating the public's reserve of democratic satisfaction, this perception of party homogeneity threatens democratic stability.

Partisanship in the Public's Mind: Analyzing Self-Reported Party Identification

Heysung Lee, *University of Wisconsin-Madison*

The self-reported scale ranging from strong Democrat to strong Republican has been consistently used in surveys to measure Americans' party identification, since Campbell et al. (1960)'s American Voters. Although there is an academic consensus that this item can predict political attitudes, vote choice, and political participation, not many studies have explored the public's perceptions of their own partisanship. Based on an online survey with 500 respondents to represent the U.S. adult population, the current study aims to capture how Americans think about their own partisanship. To do this we started from an open-ended question asking people to explain what is the meaning of their own partisanship, as a follow-up for the classic 7-point party identification question. For analysis, we employed STM (structural topic modeling), which enables clustering texts in the open-ended question as well as testing covariation with other variables in the survey. Results show that both Democrats and Republicans identify themselves as partisans for different reasons. Vote choice is the most prevalent topic for both partisans. But Republicans answered the main reason is they support Trump and specific issues like abortion or taxes, while Democrats did not specify issues or political elites, but mention the importance of more abstract values such as equality. Regression analyses is conducted to test the effects of topic proportion in the open-ended question on feeling thermometers, to understand their contribution to affective polarization. Results indicate that Democrats who mentioned "policy" tend to experience a greater affective gap between parties as well as Republicans who answered "supporting Trump." Our findings suggest a nuanced picture of how people understand their own partisanship and how such bottom-up definitions lead to the different self-image of political parties and affective discrepancies.

Rethinking Party ID: Interpretation of Political Polls Affected by Party ID Construction

Ryan Tully, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

Frances M. Barlas, *Ipsos Public Affairs*

Political events over the past four years demonstrate that the United States is in a period of deep partisan polarization. Party ID is increasingly critical to interpreting political polls. Perceived failures in the 2016 and 2020 presidential elections provide an opportunity to reevaluate all aspects of the political polling process. This paper greatly expands on our previous research demonstrating how different methods of combining response categories for Party ID are associated with varying levels of validity. In our earlier research, we identified a misordering of the traditional 7-point scale from 'Strong Democrat' to 'Strong Republican'. In this study, we greatly expanded upon our previous findings using a large volume of domestic and foreign policy issues from dozens of polls conducted on KnowledgePanel®, the largest probability-based online panel in the U.S. Fielded from 2002 to present, these studies evaluated opinions on a wide range of topical U.S. political issues, including the ongoing COVID-19 pandemic and the recent presidential election. We evaluated the impact of using different Party ID measures (e.g., linear/pooled 3-point scale, full range 7-point scales with normal response ordering, and our revised empirically ordered 7-point scale) in interpreting political polling results. We once again confirm that our revised 7-point category scale had the greatest predictive utility across an array of political issues. In addition, we found that the use of various Party ID measures produced significant differences in the interpretation of observed polling results. Furthermore, we detected clear evidence of ongoing partisan polarization and misrepresentation of 'Independents' in these polling results. We discuss our results in the context of the ideological sorting people go through when identifying by party and how this currently may be leading to greater partisan realignment and polarization in political attitudes and beliefs.

How Response Categories Affect Political Party Identification

Ge Tang, *Ipsos Public Affairs*

Frances M Barlas, *Ipsos Public Affairs*

Maya Grosul, *Ipsos Public Affairs*

Researchers are keenly aware of the importance of question wording in public opinion polling because small differences in question wording, and response options can affect a study's results. In the lead up to any Presidential election, there are numerous political polls administered to assess public opinion. In such polls, party identification is a major predictor of outcomes of interest – vote choice, candidate favorability ratings, importance of issues in the election, etc. There are no gold standard measures of party identification and different response categories are used across studies, often making it difficult to compare distributions. Most polls do not offer a more explicit "no preference" option – opting instead for a vaguer category like "something else". In this study, we compared two methods of asking respondents for their party identification – one with 4 categories (Republican, Democrat, independent, or something else) and one with 5 categories (Republican, Democrat, independent, another party, or no preference). Both versions were asked of the same respondents with the five-category version in the first study and the four-category version in a follow-up study. We examined how the change in options affected the response. The studies were conducted using KnowledgePanel®, the largest probability-based online panel in the US. A large difference was found in the percentage of independents between the two versions. Among those who chose "No preference" in the 5-category version, around 32% of them changed their status to "independent" when asked the 4-category question. We further investigated the demographics of this subset of respondents to better understand who they were. These findings demonstrate the importance of carefully understanding this group of people, since they often constitute a large portion of swing voters in close elections and results can significantly be affected if they are underrepresented or used for weighting or sampling.

Affect and Ideology: Investigation of the Origins of Affective Polarization in the US Using Open-Ended Survey Questions 1984-2016

Rongbo Jin, *University of Arizona*

Recent years witnessed the surging scholarly interests in affective polarization, which refers to the growing partisan hostility among Democrats and Republicans (Iyengar et al 2012; 2019). Many research contributed to the causes (Iyengar 2018; Mason 2015), consequences (Nicholson et al 2016; Iyengar et al 2018) and constraints (Wojcieszak and Warner 2020; Levendusky 2017; Lelkes and Westwood 2016) of affective polarization but we should be very cautious to investigate how to contain affective polarization without a clear understanding of its mechanism and origins. Extant research basically falls into two camps: first group of scholars inherit the tradition of social identity theory in social psychology, arguing that partisanship as a salient social identity drives partisan hostility (Iyengar et al 2012; Iyengar and Westwood 2014; Mason 2015; 2018), whereas the other group argues that affective polarization has ideological foundation (Rogowski and Sutherland 2016; Webster and Abramowitz 2018; Orr and Huber 2019). However, these two views are not necessarily conflicting to each other. Therefore, this research intends to revisit the puzzle what are the origins of affective polarization. I am using large-scale text analysis open-ended survey questions in American National Elections Studies (1984-2016). This research mainly focus on four open-ended questions “What do you like [dislike] about Democratic [Republican] Party?” The major method employed in this paper is Latent Dirichlet Allocation (LDA) which is a unsupervised topic model. Based on the topics generated by LDA model, people’s like and dislike about two parties falls into three categories: economic ideology, social ideology and political identity. The major hypothesis in this paper is that higher level of affective polarization measured by feeling thermometer is closely associated with political identity

The Four Faces of Partisan Prejudice

Joseph Phillips, *University of Exeter*

A recurring question in the literature on the rise of affective polarization is whether it represents in-party favoritism or out-party hatred. A proliferation of behavioral experiments and observational treatments have not added up to a clear answer. In contrast to previous work, I argue that which form partisan prejudice takes on is heterogeneous within populations and over time. To examine configurations of partisan affect, I employ a latent profile analysis on American National Election Study data and find that four configurations of partisan affect exist: neutrality, in-party favoritism, out-party hatred, and both. I further find that existing theories of affective polarization only explain people who display both in-party warmth and out-party hatred.

Party Identification and Social Relationship: Exploring the Causal Mechanisms

Ikuma Ogura, *Georgetown University*

The literature on American mass party identification has found that US citizens prefer co-partisans in building social relationships. According to the American National Election Study (ANES) conducted in 2016, nearly 80% of the respondents answered that they voted for the co-partisan candidates at the presidential election that year. Previous studies have also pointed out that Americans prefer co-partisans even for non-political social relationships, such as marriage and friendship. However, it is not clear why we see such patterns. Is this because people consider that co-partisans tend to have similar socioeconomic backgrounds? Or, is it because they think co-partisans hold similar policy positions? To answer this research question, this paper conducts original survey experiments with American voters. In the experiments, respondents are first asked to read a vignette of a hypothetical individual describing his/her partisanship, group traits, and issue attitudes. Respondents are then asked whether they want to build various social relationships with the hypothetical individual. By suitably designing the experiments for identifying causal mechanisms (Acharya, Blackwell, and Sen 2018), this paper aims to decompose the mechanisms lying behind partisanship and social relationships empirically. This paper will contribute to the literature by offering insights into the long-standing debate over the meaning of mass party identification.

Impact of Popular Vote Margins on Accepting 2020 Presidential Election Results

Brendan Hartnett, *Tufts University*

Alexandra Haver, *Tufts University*

The results of the 2020 Presidential Election were greatly challenged both preemptively and after the fact. Due to the growing uncertainty that preceded the election, we chose to analyze support among voters for their preferred candidate resisting the results of the election. To do this, we asked respondents to the Tufts University 'Polling the 2020 Election Seminar' National Election poll about what their preferred candidate should do if they lost the Electoral College across various popular vote margins. We provide insight into the impact of popular vote margins on election legitimacy, into the favorability of the Electoral College, and into views that would warrant voters supporting their candidate challenging the results. Through the end of 2020, we offer analysis of what the presidential election results and the actions of political elites mean for U.S. democracy. Our analysis does not include events that transpired in 2021.

Qualitative Methods for Questionnaire Development and Pretesting

How Do You Translate “Homeschooling” Into Spanish, and What Does It Mean in 2020?

Daniela Glusberg, *Research Support Services Inc*
Maria Payri, *American Institutes for Research (AIR)*
Erika Martinez Picazo, *Research Support Services Inc.*
Alisú Schoua-Glusberg, *Research Support Services Inc.*
Danielle Battle, *American Institutes for Research (AIR)*

Everyday life under Covid-19 has, undoubtedly, influenced how respondents in the US interpret certain survey items. Questions about homeschooling, we hypothesize, may be among them. With schools across the country having shifted to remote learning via the internet, is homeschooling being understood as it was before March 2020? Surveys, particularly those having to do with topics related to education, often need to ask respondents if they homeschool their children, rather than sending them to a traditional public or private school, a physical building outside of their home. Translation into Spanish of these types of survey items has always been a challenge because of the lack of familiarity with the concept on the part of less acculturated immigrants and other monolingual Spanish speakers in the US. In the Summer of 2020, we conducted qualitative interviews to cognitively test changes to one of the National Household Education Surveys (NHES) instruments which included questions for parents about homeschooling their children. Our cognitive testing of the Spanish translation of these items can shed some light on these issues. We asked Spanish speakers about their familiarity with the concept of homeschooling, as well as their reactions to multiple definitions and various translations of the term. Despite their exposure to varied formulations and receiving an explicit instruction not to include any schooling at home after March 2020 that resulted from Covid-19 school closings, we still found lack of understanding that led to errors in Spanish speakers’ survey responses. In our presentation we will examine: pre-2020 issues with translating questions about homeschooling; possible confusion of remote learning experiences when we ask about homeschooling in 2020 and beyond; and the intersection of these two issues for monolingual Spanish speakers, through an examination of what we learned in our NHES testing.

Pretesting New Survey Materials in a Remote Environment

Lauren Walton, *Nielsen*
Emily Baker, *Nielsen*
Jennifer Hunsecker, *Nielsen*
Hani Zainulbhai, *Nielsen*

Measurement error occurs when respondents are unable or unwilling to provide accurate and precise answers to our survey questions. Writing a “good” survey question is an art and science, and pretesting survey and communication materials is always a best practice. Outer packaging of mail surveys is a researcher’s chance to engage the potential respondent with images, color, size, or messaging. Private, government, and academic organizations all send out mail surveys and without a successful outer package, they may be doomed to the recycle bin. During Q2-Q3 2020, Nielsen rapidly designed a mixed mode recruitment methodology to supplement in-person methodology due to COVID-19 pandemic. Nielsen’s overall contact strategy includes these phases: (1) Brand awareness postcard, (2) Large mailer with web invite and small cash incentive, (3) Mail reminder for those who did not complete after initial invite, (4) Final reminder for hard to reach demographics (4) Recruitment website Nielsen.com/join (5) Online registration survey, and (6) Mixed mode follow up contacts. In order to deliver new household communications, researchers needed to iteratively pretest mailed communications, emails, online surveys, and telephone follow ups. This paper will present results of moderated and unmoderated online tests in both English and Spanish as well as internal friends and family testing of the new process, and fresh qualitative sample cognitive interviews. Results from 8 separate pretests spanning motivational messages to survey usability will be discussed. Findings include support in the selection of an 8 by 12 Nielsen branded envelope for the survey invitation that is “It’s a very pretty color.” “I’m drawn to the front;” “It looks fun - it looks inviting;” “It’s beautiful. You can tell there’s a lot of thought and energy that went into it so I would definitely open it.”

Refining the Use of Web-Probing as Part of Questionnaire Pretesting

Mandi Martinez, *U.S. Census Bureau*

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

Yazmin Garcia Trejo, *U.S. Census Bureau*

With advances in the administration of web surveys, web-probing has become an increasingly popular tool in questionnaire pretesting. Web-probing is the practice of asking cognitive probes as follow-up questions to web-based survey questions as a method to assess question performance. Previous research has found that web-probing can be an effective complement or supplement to in-person cognitive interviewing. Web-probing has some benefits over traditional face-to-face cognitive interviews including increased sample sizes and geographical diversity, quicker data collection, and reduced costs per participant. However, it is still an evolving method and additional research is needed to refine methods used for web-probing under a variety of circumstances, to identify limitations, and to explore how it complements and/or could supplement more traditional cognitive interviewing. In this study, we examine web-probing data collected using an online panel and responses to probes asked during in-person cognitive interviews. Data were collected during pretesting of two different surveys measuring respondents' knowledge of the 2020 Census, trust in the U.S. Census Bureau, and concerns about privacy and confidentiality. In both the web surveys and in-person interviews, a series of probing questions were asked retrospectively to assess respondent comprehension of the questions and to identify issues in the questionnaires. The data collected allow us to examine the performance of web probes and in-person probes to identify which types of web probes perform well to uncover different types of issues commonly found during pretesting, and to compare the richness or quality of the responses collected with the two methods. Additionally, the inclusion of two different types of probes – process-oriented probes and meaning/comprehension probes -- allows us to assess how effective web-probing is for different types of probes commonly used in traditional cognitive interviewing.

Cognitive Interviewing and Survey Pretesting – State of the Art

Andrew Caporaso, *Joint Program in Survey Methodology, University of Maryland*

Stanley Presser, *Joint Program in Survey Methodology, University of Maryland*

Although it is often recommended that survey questions undergo cognitive testing prior to fielding, little is known about how often this recommendation is followed and what cognitive testing consists of when it is conducted. This paper reports the results of a 2020 survey to address these issues. The questionnaire, administered to a sample of survey organizations, allows us to shed light on.

- How many cognitive interviews are typically conducted to test questions, and who conducts the interviews
- Guidelines used for think-aloud and probing
- Cognitive interview summarization and analysis practices
- Use of other pretesting methods and
- The impact of COVID-19 on cognitive testing and pretesting practices.

Conducting Cognitive Interviews With People With Intellectual and Developmental Disabilities

Amanda Wilmot, *National Center for Health Statistics*

Despite known health disparities between those with and those without Intellectual and Developmental Disabilities (IDD) (Bonardi et al., 2011; Krahn & Fox, 2014), there is a paucity of data about the health of adults with IDD to inform policy (Krahn, 2019). Since people with IDD generally live in the community, household population surveys are an appealing data collection vehicle. However, the prevalence of adults with IDD is not considered adequately captured through current population survey questionnaires (Havercamp et al, 2019), in part because of the way in which survey questions are worded or administered (Krahn, 2019).

Staff at the National Center for Health Statistics are conducting an evaluation of survey questions designed to identify community-dwelling adults with IDD. To see how the questions performed with this particular population the evaluation included the standard short set of six questions developed by the Washington Group on Disability Statistics. These questions ask about the presence of difficulties in six core functional domains: seeing, hearing, mobility, cognition, self-care and communication. Other questions evaluated were derived from those asked previously on surveys or used in assessment tools about learning ability, independent living skills, communication, self-direction and social participation. Cognitive interviews were conducted directly with adults with IDD, and in some case, also with those who may act as proxy respondents in a social survey setting. This presentation discusses the challenges involved in conducting cognitive interviews with this particular population.

The Bigger Picture: Strengths, Limitations and Recommendations From a Photo-Elicitation Study Among Older Adults With a Low Socioeconomic Status

Feline Platzer, *University Medical Center Groningen*

Nardi Steverink, *University of Groningen, University Medical Center Groningen*

Marieke Haan, *University of Groningen*

Mathieu de Greef, *Hanze University of Applied Science*

Martine Goedendorp, *University of Groningen, University Medical Center Groningen*

There is an abundance of research focusing on the health of older adults with a lower socioeconomic status (referred to as SES). However, research with this target group faces several methodological challenges. For instance, due to the cognitive skills and literacy level of the participants. Visual methods are accepted tools in qualitative research and are suitable for research with participants with lower literacy levels. For older adults with a lower SES, the use of photo-elicitation interviews could be a proper method to investigate abstract topics, such as perception of health. Previous studies used photographs to address the health of older adults. However, these studies used mostly participant-driven photographs or focused either on lower socioeconomic status or on older adults. Knowledge about the use of researcher-driven photographs in research with older adults with a lower SES is scarce. In our research, we wanted to investigate the positive health perceptions of older adults with a lower SES with the use of researcher-driven photographs. In addition, we wanted to know whether participants perceived a sense of control on their health. We developed a researcher-driven photo-elicitation study with the use of a developmental, testing and implementation phase which consisted of nine steps. During the developmental phase we gathered the photographs and designed the interview. During the testing phase, we tested the photographs with representatives, two focus group sessions and two individual interviews. During the implementation phase we recruited participants and interviewed 19 older adults with a lower SES. It is the aim of the presentation to address the process, strengths and limitations of our photo-elicitation study and the lessons which we learned along the way. By doing so, we hope to contribute to the understanding of the methodological issues when using the photo-elicitation method with this target group.

Developing Measures for Emergent Topics: The Example of E-Cigarettes

Lauren Creamer, *National Center for Health Statistics*

Paul Scanlon, *National Center for Health Statistics*

Recent research has explored the application of web probing in questionnaire evaluation; however, mixed-method approaches for developing questions on emergent topics has not been described in the literature. This poster will use the example of e-cigarettes to discuss best practices for developing survey questions on emergent-topics. Awareness of e-cigarettes has substantially increased since 2010 and a measure of e-cigarette use was first included on the National Health Interview Survey (NHIS) in 2014. Consistent with questionnaire best practices for new and unfamiliar topics, the NHIS incorporated definition text within the e-cigarette question. Between 2014 and 2018, the National Center for Health Statistics (NCHS) used qualitative and quantitative methods to evaluate whether or not this definition text was still needed.

The development of the e-cigarette measure spanned across multiple projects and utilized numerous approaches including expert review, cognitive interviewing, and close-ended comprehension probes on NCHS' web panel-based Research and Development Survey (RANDS).

Each approach provided unique data points. Subject-matter experts provided insight into the desired constructs to be measured. The cognitive interviews provided rich, but non-generalizable, data on how respondents understood the term "e-cigarette." Specifically, the interpretivist approach was used to analyze cognitive interview data and understand the ways respondents interpreted the e-cigarette use question with and without the definition text. A split sample web survey experiment was developed and a close-ended web probe that explored item comprehension was cognitively tested before being administered on the RANDS web panel. The RANDS data provided a comparability of prevalence of in-scope and out-of-scope interpretations of the term "e-cigarette."

Respondent Compliance and Cooperation

Hello...Good-bye...: Hang-ups and Break-offs by Mode, Geography, and Neighborhood Characteristics in Oregon

Deborah Krug Mangipudi, *ICF*

Matt Jans, *ICF*

Robynne Locke, *ICF*

Stephen Haas, *ICF*

John Boyle, *ICF*

Randy Zuwallack, *ICF*

Lizzie Remrey, *ICF*

Heather Driscoll, *ICF*

Melinda Scott, *ICF*

Survey topic can influence response propensity, even if the topic only becomes salient to respondents once questions are read. We usually design interviews and questionnaires to begin with topic-relevant, simple, nonsensitive questions. However, perceived sensitivity can vary widely by respondent experience and background, producing disproportionate nonresponse. This presentation addresses the following questions: 1) Do specific topics, questions, or screens lead to more hang-ups and break-offs than others, 2) Do topic/question/screen-level hang-up and break-off rates vary by respondent demographics and characteristics of where they live, and 3) Do these rates differ between phone and web modes? Data come from the Oregon Crime Victimization Survey (OCVS), which used dual-frame random-digit-dial (RDD) and address-based samples (ABS). In both samples, adults living in Oregon for at least the past 12 months were eligible. The questionnaire was identical in both modes, and included these sections: 1) eligibility screening, 2) consent, 3) quality-of-life, 4) primary demographics, 5) index crimes, 6) non-index crimes, 7) crime follow-up, and 8) secondary demographics. Among people contacted by phone or who began the web questionnaire, we evaluate the following for each topic section, question, and screen: 1) % hang-up or break-off by 2) geographic stratum, 3) areas with higher and lower levels of poverty or other sociodemographic compositions, and 4) demographics captured early in the questionnaire. Each of these is compared between phone and web. Given the current sociocultural context, and demonstrations about criminal justice reform, we are particularly interested in whether the initial topics (e.g., screening, consent, and neighborhood quality-of-life) obtain differential nonresponse across strata and geographies that represent different neighborhood types and sociodemographic distributions. Results are interpreted in the context of questionnaire design decisions, question sensitivity, mode differences, and overall nonresponse rates.

“I Wasn’t Really Thinking About Myself”: Exploring Who Participants Think About When Answering Survey Questions About Abortion

Kathryn J. LaRoche, *Indiana University*

Katherine R. Haus, *Indiana University*

Julie Maier, *Indiana University*

Alejandra M. Kaplan, *Indiana University*

Kristen N. Jozkowski, *Indiana University*

Abortion attitudes may differ depending on who participants think about when responding to survey questions. In a pilot survey assessing abortion attitudes in English (n=31) and Spanish (n=25), we randomly assigned participants to answer questions that asked them to think about themselves/themselves and their partner, or a hypothetical woman. We conducted follow-up cognitive interviews (CIs) and used content and thematic analysis to assess how participants arrived at their answers. Then, using NORC’s AmeriSpeak panel (N=608), we administered a revised survey, assigning participants to the same conditions, and asked them who they thought about when answering questions. Based on who participants reported thinking about, we categorized participants as fully, somewhat, or not at all compliant. Descriptive statistics and chi-squared tests were used to assess compliance across survey conditions. Regression will be used to examine participants’ demographic characteristics and compliance. In the CIs, participants described using a variety of strategies when responding to the survey questions. Many drew on their own experiences even if they were asked to think about someone else,

and vice versa. Findings from the survey data indicate that 21.5 to 28.6% of participants did not consider wording of the survey at all when responding to items (non-compliance) and 20.8% to 39.9% reported mixed compliance. There was a significant difference in survey condition ($\chi^2 > = 101.24$, $df = 2$, $p = .001$), such that people assigned to the “you” condition were more likely to think of themselves than those assigned to other conditions. Our findings suggest that people draw on various personal experiences and cognitive strategies when answering questions about abortion, and many people are unable to separate thoughts about themselves from other people in this context. Low rates of compliance with survey instructions have implications for the development of polling items related to abortion.

It Is Only One Question– Finding Ways to Encourage Response From Ineligible Households

Danielle Battle, *American Institutes for Research*

Luke Natzke, *American Institutes for Research*

In two-stage surveys, response from sample units that are ineligible for the second stage during the first stage is vital. Even though such sample units do not contribute data, they impact final response rates and can inflate the cost of nonresponse follow-up. In the National Household Education Survey (NHES), a first-stage screener survey is used to find households with children. This paper presents the results of an experiment designed to increase response to the 2019 NHES screener among households without children—households ineligible for the second stage of the survey. The experiment tested a set of contact materials that emphasized the brevity of the survey for households without children paired with a version of the paper questionnaire which placed the screener question about children in the household on the survey cover. Placing the screener item on the cover of the survey allowed households without children to respond without having to move beyond the cover of the paper questionnaire. Preliminary analyses of experimental results indicate that the experimental materials increased the final response rates compared to the control group. Further analyses will examine the impact of the experimental materials on eligible sample units to ensure that increasing response among households without children did not depress the response rates for households with children.

Understanding Respondent Behavior: Toward a Revised Approach to Understanding Who Participates in Surveys

Seth Messinger, *Ipsos Public Affairs*

Robert Petrin, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

Frances M. Barlas, *Ipsos Public Affairs*

Of increasing concern to many researchers is the recruitment and retention of respondents of color. While the existing literature attempts to understand overall (and subgroup specific) respondent rates via reference to infrastructural issues and choices made by potential respondents (Leeper 2019), perceptions of interviewer and style (Dijkstra 1987; Pickery et al 2001), and the role of question formation (Olson et al 2019; Rogelberg et al 2001), we build a broader conceptual critique on research participation that can shed light on racial and socioeconomic differences in participation. More specifically, by augmenting the survey response and social psychology perspectives with those from sociology and anthropology, we argue that attention to survey respondents’ identities represents a lacuna in the research on participant motivation and may be a factor in whether individuals continue to participate or not. In developing this approach, we draw upon two literatures to expand our thinking about survey response behavior: critical race theory and social recognition theory. Critical race theory addresses the social and political disparities that marginalize the accounts of people of color, and by extension women and sexual and gender minorities. (Garcia et al 2018; Sablan 2019). Social recognition theory provides a foundational understanding of social exclusion and inclusion (Honneth 2004; Cherrier 2007). Critical race theory and social recognition theory place participants with varying degrees of social status at the heart of quantitative measurement. We developed differential implications for each theory for survey participation. A series of studies were conducted that asked respondents to consider their survey experiences and concerns about surveys as well as their interests in survey topics. We examined these responses in the context of the theoretical implications we developed as well as consequences for individual completion rates.

Quick Turnaround Probability Panel Surveys

Kyle Berta, *SSRS*
Jennifer Schmidt, *SSRS*
Jennifer Su, *SSRS*
Chintan Turakhia, *SSRS*
Jennifer Agiesta, *CNN*

It is often challenging to conduct representative quick turnaround surveys via self-administered modes (such as via a probability-based online panel) as hard-to-reach populations usually take longer to complete surveys. A prolonged field period gives hard-to-reach populations more time to complete surveys and generally results in improved response rate. In an age where the length of a news cycle keeps shrinking and the next big story is right around the corner, surveys need to be implemented in a short data collection window while preserving representation.

In our presentation, we outline our practices for conducting quick-turnaround probability-based panel surveys, obtaining both a representative sample and a suitable completion rate. Utilizing the probability-based SSRS Opinion Panel, we examine a recent national general population survey on a breaking news event where we obtained a representative sample of over 900 adults with fewer than 36 hours in field.

Does Failing an Attention Check Relate to Survey Mode and Representativeness?

Jessica Daikeler, *GESIS - Leibniz Institute for the Social Sciences*
Joss Rossmann, *GESIS - Leibniz Institute for the Social Sciences*
Tobias Gummer, *GESIS - Leibniz Institute for the Social Sciences*
David Bretsch, *GESIS - Leibniz Institute for the Social Sciences*
Henning Silber, *GESIS, Leibniz Institute for the Social Sciences*

Respondents who inattentively respond to survey questions usually provide answers of lower quality compared to respondents who effortfully comprehend, process, and answer the questions. Previous research has argued that inattentive responding may particularly be a challenge for self-administered surveys in which no interviewer is present who can motivate and guide the respondents through the interview. Accordingly, attention checks were proposed to identify inattentive respondents. However, methodological studies on the properties of these tests are sparse and all of the few existing studies have focused on web-based data collection. So far, mail mode has not received any attention in this line of research. Moreover, previous studies have mainly used data from convenience samples and non-probabilistic access panels. Since the attention of the respondents is particularly related to their motivation to participate in the survey, this research gap is particularly unfortunate. Thus, our study aimed at investigating whether attention checks methods can be used in mail questionnaires and probability-based mixed-mode surveys.

We addressed both research gaps by conducting an experiment in a probability-based general population mixed-mode panel survey in Germany (N=4,777). The respondents were randomly assigned to either answering a grid question including an instructed response item attention check (66.4% of the respondents) or receiving the same grid question without the attention check item (33.6%). In our study, we found that mail respondents are more inattentive than web respondents, even when correcting for mode and attrition using gradient boosting models. Our findings suggest that mail respondents provide more item-nonresponse and web respondents are more inattentive. Furthermore, we are investigating for mail respondents whether this is just a single item problem or if the entire item battery is affected. Our contribution closes with recommendations for the implementation of attention checks in mail mode and probability-based panels.

The Issue of Non-Compliance in Attention Check Questions

Henning Silber, *GESIS - Leibniz Institute for the Social Sciences*

Joss Roßmann, *GESIS - Leibniz Institute for the Social Sciences*

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

Attention checks instruct respondents to perform a task to prove that they are attentive (e.g., provide a specific answer or select a specific response category). However, respondents may properly process the question but choose to not comply with the instructional manipulation on purpose. These respondents are then falsely identified as inattentive by the instrument, which confounds the attentiveness measure. We contribute to the knowledge about attention checks by reporting the findings of two web surveys. In the first study, we measured respondents' attitudes toward attention checks and their self-reported compliance. In the second study, we experimentally manipulated reasons that were given to respondents why they should comply with the instruction in the attention check. Our results showed that most respondents understand why attention checks are conducted but, at the same time, a non-neglectable proportion of respondents evaluated them as controlling and annoying. While most respondents noticed the attention check, among those who failed the test –strikingly– 61% seem to have done so on purpose, reinforcing that non-compliance is a serious issue with these instruments. The results of our experiment showed that more respondents pass the attention check if we give them a comprehensible reason. Altogether, our findings resonate with critical voices about using attention checks and suggest that if used, it should be clearly communicated to the respondents why they are implemented.

In an Era of Enhanced Cybersecurity: The Effect of Confidentiality Pledges and Penalty Language on Survey Response

Cleo Redline, *National Center for Education Statistics*

Alfred D. Tuttle, *U.S. Census Bureau*

Federal surveys often pledge that personally identifiable information (PII) will be used for statistical purposes only and that the willful disclosure of PII by employees who have been sworn to confidentiality are subject to felony penalties. In 2015, the passing of the Cybersecurity Enhancement Act provided the Department of Homeland Security with a mandate to monitor information on federal computers for viruses, malware and other threats (6 U.S.C. §151). Since Homeland Security personnel and their contractors are not bound to the confidentiality pledges used by the statistical agencies, the question became how to inform respondents about these changes without alarming them and harming response rates. This paper reports on an experiment conducted under survey conditions with a probability sample of respondents (n=4122). The experiment examined differences in response rates across three versions of pledges describing cybersecurity monitoring. We also included in our experimental design messages describing penalties for employees and contractors if they willfully disclose PII. Half of the pledges included the penalty language and the other half did not, resulting in a fully crossed factorial design. There were no statistically significant differences between any of the experimental comparisons. We conclude that, of the pledges studied, the Homeland Security version of the pledge with penalty language provides respondents with the most accurate information for deciding whether to participate in the survey without harming response rates.

5/13/2021, 3:30 PM - 5:00 PM
Concurrent Session G: Live Sessions

Mini-Conference: Impacts of the Pandemic on Social Isolation and Mental Health

Social Isolation and COVID: Does Forced Social Isolation Create More Empathy and Understanding?

Terri Guengerich, *AARP*

Social isolation and loneliness pose serious health risks. They can be worse for your health than obesity, and prolonged social isolation can be as bad for your health as smoking 15 cigarettes a day. While older adults may be at a higher risk of experiencing loneliness and social isolation, the COVID pandemic has certainly expanded this experience to many more adults, both young and old. To study the effects of social isolation during this pandemic, AARP commissioned two phases of qualitative research among people living in Kansas who are at least 40 years old to understand the dynamics of social isolation. Phase I included two online focus groups (one with Persons of Color), and Phase II captured the thoughts of forty-three Kansans via online journaling. Using the findings from these online focus groups and 106 pages of journal transcripts, we can better understand the dynamics of social isolation including perceptions, experiences, and barriers to engagement; isolation disparities between White people and People of Color; COVID's influence; and coping mechanisms and approaches to overcome social isolation. The purpose of this poster is to better understand social isolation during this pandemic including disparities among whom it affects and the burden it bears. Given a shared experience of social isolation, will greater empathy and understanding of social isolation replace old stigmas and garner more community support and solutions?

The Effect of COVID-19 Mitigation Measures on the Prevalence of Sexual Misconduct and Intimate Partner Violence Among Young Adults

Reanne Townsend, *Westat*

David Cantor, *Westat*

Susan Chibnall, *Westat*

Vanessa Kranz, *Westat*

Bonnie Fisher, *University of Cincinnati*

Studies measuring sexual assault and intimate partner violence among young adults have found victimization rates to be associated with a number of different behaviors and characteristics. Factors such as being in the first year of college or attending fraternity parties have been associated with sexual assault victimization (Krebs, Lindquist, Warner, Fisher, & Martin, 2007), and cohabitation with a partner is associated with intimate partner violence (Renner & Whitney 2012). Many of these factors associated with victimization were significantly affected by COVID-19 mitigation measures, and some organizations have expressed concern over possible increases in victimization during the pandemic (Evans, Lindauer & Farrell 2020). This presentation will begin to investigate the issue of how rates of sexual misconduct and intimate partner violence among young adults changed during the first months of the COVID-19 pandemic. This study uses data from the National Study of Young Adults, a longitudinal pilot study sponsored by the National Institute of Justice. In the spring of 2019, the study recruited young adults age 18 and 19 from an address-based sample frame matched to a marketing list of high school seniors. Participants were initially recruited into the study through a mail-push-to-web procedure, and invitations to all subsequent surveys were sent via email and text only. After recruitment, all participants were surveyed once in November 2019, and again six months later in May 2020. Between these two surveys, two-thirds of participants were also asked to complete short microassessments every two or four weeks. This presentation will focus primarily on the results of these microassessments. This presentation will discuss victimization rates both before and after the onset of the COVID-19 pandemic, and examine how they changed. We will explore how these trends differ by respondent behaviors and characteristics, including whether they are a student, live on campus, and have a job.

How Right Now: Addressing Mental Health, Coping, and Resilience Amidst COVID-19

Amelia Burke-Garcia, *NORC at the University of Chicago*

Ashani Johnson-Turbes, *NORC at the University of Chicago*

Elizabeth W. Mitchell, *Centers for Disease Control and Prevention*

Jorge M. Vallery Verlenden, *Centers for Disease Control and Prevention*

Richard Puddy, *Centers for Disease Control and Prevention*

Melissa C. Mercado, *Centers for Disease Control and Prevention*

Allison Friedman, *Centers for Disease Control and Prevention*

Katherine Bruss, *Centers for Disease Control and Prevention*

Daniel Pierce Nelson, *CDC Foundation*

Background. How Right Now (HRN) is a mental health initiative grounded in psychological and communication science, which launched in August 2020 to support emotional well-being and resilience amid COVID-19. This presentation will review the formative research that signaled the need for mental health support and informed the initiative.

Methods. From April-July 2020, rapid research was conducted to explore mental health challenges, perceptions of resilience, and support needs among primary audiences: people living with pre-existing health conditions, people experiencing violence, people experiencing economic distress, people aged 65+, and caregivers of older people. This included environmental scans (n=700 publications), social listening (n=1 million posts), listening sessions (n=29), online focus groups (n=58), and a national survey (n=1,000).

Findings. Among key findings, common feelings of grief, depression, and anxiety emerged. Those experiencing economic distress reported lower confidence in both their own (20%) and their communities' (16%) abilities to be resilient. Most (60%) respondents preferred support from family or friends over other forms of support. Additional findings will be shared.

Implications. Targeted communication and partner engagement at the local, community and national levels are needed to support HRN's audiences. Additional research is underway to identify emerging needs that HRN may be able to address.

Personality and Social Distancing During the COVID-19 Pandemic: An Examination of 23andMe's Research Cohort

Daniella Coker, *23andMe*

Background: Social distancing, the practice of keeping at least six feet from others, has become a primary tool for preventing the spread of coronavirus disease 2019 (COVID-19). Despite widespread support for social distancing, adherence is not uniform within the United States. Identifying factors associated with differences in behavior is important for ongoing prevention. This investigation focused on examining social distancing adherence by demographic and personality variables.

Methods: Between April and July 2020, nearly 1 million consented research participants self-reported their adherence to social distancing recommendations via an online survey. Of those research participants, over 100,000 also completed an online survey assessing personality traits (IPIP-NEO-60). Relationships between social distancing and demographics, including sex, age, and state of residence, were examined. Multivariate logistic regression was used to calculate associations between individual personality traits and social distancing adherence.

Results: Overall, 88.7% of participants reported adherence to social distancing. Adherence decreased from April through June, but increased by July. Greater prevalence of social distancing was observed with older age, and residence on the West Coast or Northeast. After controlling for potential confounders (age, sex, education, essential worker status, presence of comorbidities with higher risk of serious illness from COVID-19, relative zip code density, and survey completion month), higher scores on facets of extraversion (gregariousness, excitement seeking, or friendliness) were associated with a reduced likelihood to practice social distancing. Higher scores on facets of openness (liberalism, artistic interests)

and agreeableness (sympathy, altruism, cooperation, trust) were associated with an increased likelihood to practice social distancing.

Conclusions: Our findings show that while a majority of 23andMe research participants practice social distancing, differences persist when adherence is examined by age, time, geography, and personality. Furthermore, these findings suggest that personality may be important for understanding the spread and containment of infectious disease.

Surveying Neighborhood Connections

Brianne Gilbert, *Loyola Marymount University*

Simona Vishnevsky, *Loyola Marymount University*

Trust in one's neighbors and a sense of community are vital to an individual's desire to live and remain in a particular area. However, the COVID pandemic has made developing or maintaining trust and a sense of community much more difficult. In a 2020-2021 study of 2,000 adult city of Los Angeles residents (n=1,000 residents per year), researchers at the Center for the Study of Los Angeles at Loyola Marymount University ask a battery of questions related to these two areas of focus. Residents are queried each January (that is, before the pandemic hit and during the pandemic) in a mixed-mode (phone and online) survey in four languages (English, Spanish, Mandarin, and Korean). Next, results of these survey questions are mapped by geography and then analyzed by year and demographics (e.g., race/ethnicity, age, gender, and household income). Following up on these results, researchers then conduct qualitative interviews with the neighborhood council leaders of the various communities to understand what is being done at the council level to encourage trust and a sense of community during the pandemic. The goal of this research is to understand which and where communities have retained trust and a sense of community in one's neighbors, as well as what are the characteristics of those communities. As the pandemic lingers, this research sheds light on what these communities are doing well in these important areas and how others can learn from best practices.

Mini-Conference: The COVID-19 Household Pulse Survey: Innovation by Necessity

The COVID-19 Household Pulse Survey: Innovation by Necessity

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

The Household Pulse Survey represents an historic effort by the Census Bureau and our partner agencies to design, collect, and release data to inform the nation's response to the COVID-19 pandemic. Since April 2020, more than one million responses have been collected to provide rapid-response data to the public at the national, state, and large metro area level. This panel will cover the survey lifecycle for this groundbreaking data collection. In less than one month's time, this data collection went from an idea into the field, overcoming logistical hurdles that necessitated never-used-before methodology and a truly collaborative interagency effort. From the pretesting methods, on-the-fly experiments and agile questionnaire improvements to innovative contact strategies, analyzing data quality and nonresponse bias, all the way through high-frequency, rapid-response data dissemination, this presentation will provide all you need to know on this high-profile, interagency effort.

Using Web Probing to Rapidly Pretest the Census Bureau Household Pulse Survey

Jessica Holzberg, *U.S. Census Bureau*

Robin Kaplan, *U.S. Bureau of Labor Statistics*

Paul Scanlon, *National Center for Health Statistics*

In April 2020, the Census Bureau launched the Household Pulse Survey (HPS), an experimental data collection designed to collect information on the experiences of American households during the coronavirus pandemic. The accelerated timeline for this Census Bureau data collection was unprecedented, moving from inception to data collection in approximately one month. This left very little time to conduct questionnaire pretesting in a time in which traditional in-person methods such as cognitive interviewing were also not feasible. In order to collect feedback on the HPS questionnaire in this tight timeframe, we quickly pivoted to using web probing. We conducted three rounds of web probing, including two rounds on the HPS Phase 1 questionnaire and one round on the Phase 2 questionnaire. Our primary goals were to ensure that new questions were answerable, coronavirus pandemic terminology was understood, and reference periods were salient. We also asked a question to determine whether there were other impacts of the pandemic about which the Census Bureau might consider asking at a later time. To accomplish these goals, we used a mixture of different probes, including a) open-ended and closed-ended, b) concurrent and retrospective, and c) question-specific and general probes. We also split-ballot pretested some survey questions and probe types to facilitate quick, informed decision-making. Probes were embedded in a copy of the HPS instrument that was then sent to a sample drawn from the Census Bureau's opt-in nonprobability panel, for a total of 2,589 responses across the three rounds. In this paper, we highlight key pretesting findings on new pandemic questions. We discuss innovations implemented in the HPS Phase 2 pretesting based on experiences with Phase 1 and the advantages and methodological tradeoffs of relying on web probing as a rapid pretesting method in lieu of in-person cognitive interviewing.

Bob and Weave: When COVID-19 Made a Federal Survey Nimble

Casey Eggleston, *U.S. Census Bureau*

Mikelyn Meyers, *U.S. Census Bureau*

Jason Fields, *U.S. Census Bureau*

A hallmark feature of the Census Bureau's experimental Household Pulse Survey is its ability to make adjustments in near real time. Week-to-week feedback from respondents, data users, and sponsors, as well as new developments in the pandemic response have necessitated and enabled a process of ongoing adjustment and improvement. Though various design features made large-scale modifications between rounds difficult,* small-scale course corrections were implemented throughout. In this talk, we will discuss several of these within-phase modifications, including split-ballot testing of questionnaire design and wording, development of data quality checks in follow-up questions, the addition of soft edits based on unanticipated response patterns in the data, and text analysis of open-ended responses. For each, we will summarize the information that initiated the evaluation or change, details of the modifications, results, and lessons learned for nimble data collection in a rapid response survey. *But see discussion elsewhere in this panel about major changes by phase of data collection.

Reaching a Household Sample During a Pandemic; Contact Strategies for the Experimental Rapid-Response Household Pulse Survey.

Jason Fields, *U.S. Census Bureau*

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

Casey M. Eggleston, *U.S. Census Bureau*

In March 2020, the U.S. Census Bureau no longer allowed Field Representatives (FRs) to visit homes to conduct interviews and closed mail and telephone center operations to slow the spread of the Coronavirus (COVID-19) and protect the workforce and population. The Household Pulse Survey (HPS) was conceived and developed between March 23 and April 23 to address an absence of high-quality population level data to assist federal agencies in meeting the challenges raised by the pandemic. The absence of our usual contact and interview methods led to the development of an experimental approach for contact and collection. Addresses were sampled from the Census Master Address File (MAF) and supplemented with up to five emails and five mobile-phone numbers from our administrative data Contact Frame. The HPS is conducted online using the Qualtrics collection platform that is currently used at Census for survey R&D and provides the necessary agility to deploy the HPS quickly and securely. Importantly, Qualtrics operates in the Gov Cloud, is FedRAMP authorized at the moderate level, and has an Authority to Operate from the Census Bureau to collect personally identifiable and Title 13-protected data. In phase 1, contact activities for the 6-day collection started on Thursdays, and continued through closeout on Tuesdays each week. In phase 2 the collection period was shifted to 13-days. On the first Wednesday of the data collection period, email invitations are sent in the morning and an SMS invitation in the afternoon. These modes are expected to yield response rates much lower than traditional Census Bureau household surveys. The benefits are implementation efficiency, cost, and timeliness of responses. Along with the mix of contact strategies and success by method, we discuss some of the operational and logistical challenges with implementing these contact methods on the scale necessary in the HPS.

The Census Bureau's Household Pulse Survey: What Do We Know About Data Quality and Nonresponse Bias

David Hornick, *U.S. Census Bureau*

Sandy Peterson, *U.S. Census Bureau*

Anthony Tersine, Jr., *U.S. Census Bureau*

Norilsa Toribio, *U.S. Census Bureau*

In April 2020, the Census Bureau launched the Household Pulse Survey (HPS), an experimental data collection designed to collect information on the experiences of American households during the coronavirus pandemic. A systematic sample design was implemented to select representative samples within each state and the 15 most populous metropolitan areas. Data collection was conducted by internet through invitations sent by email and text messaging, so the sampling frame consisted of only housing units on the Census Bureau's Master Address File for which an email address or cell phone number was known. In this paper we will provide a high-level overview of the sample design to help understand the impacts on data quality between Phase 1 and Phase 2. We will discuss coverage rates for basic demographic groups to show how the weighting adjustments effect data quality. Due to low response rates, we will discuss the results of our ongoing nonresponse bias analysis.

Data From the Census Bureau's Household Pulse Survey

Hyon B. Shin, *U.S. Census Bureau*

In April 2020, the Census Bureau launched the Household Pulse Survey (HPS), an experimental data collection designed to collect information on the experiences of American households during the coronavirus pandemic. The accelerated timeline for this Census Bureau data collection was unprecedented, moving from inception to data collection in approximately one month. The questionnaire content and the final data products from the HPS were developed in collaboration with several different federal agencies. In this presentation, we will discuss the questionnaire content changes over the last year, highlighting some key differences between the phases, the testing of the questions, and other topics. Between the phases, we expanded the content, adjusted existing content, changed the collection period from 6 to 13 days, and adjusted some of the processing. These necessary operational changes led to variations in some of the survey estimates. The degree to which this is a break-in-series will be discussed along with the results. The Household Pulse Survey is an experimental data product; it includes only minimal editing and processing. The presentation will describe some of the ways that this influenced the data, the development of data products, and the communication of how to use the data products to the public. We will present key findings from both phases of the HPS as well as how to access the data through the Interactive Data Tool, the detailed tables, and the Public Use File (PUF). Finally, as the coronavirus pandemic continues, the future of the HPS and potential adaptation to the survey will be presented.

Mini-Conference: Evaluating the 2020 Pre-Election Polls

What Happened in the 2020 Election? Perspectives From Administrative Data

Yair Ghitza, *Catalist*

Jonathan Robinson, *Catalist*

Public opinion research is critical to the interpretation of election results. When an election concludes, survey data informs the media's interpretation of the election, the strategies of elected officials and political parties, and researchers' understanding of electoral trends writ large. Yet major data sources disagree substantially on important questions as basic as, "what percent of voters had a college degree?" Problems in estimating the composition of the electorate inevitably lead to problems in estimating candidate support, because the system as a whole adds up to the (true) election result. We use multiple administrative datasets – namely, Catalist's voter registration database, matched to precinct and county election returns – to estimate the composition of the electorate and voting preferences of different demographic and geographic subgroups. We compare our estimates to traditional design-based approaches, including the exit poll, AP VoteCast, the Current Population Survey, and the Congressional Cooperative Election Studies. Finally, we hope to evaluate how this data can be used to mitigate pre-election polling error.

Assessing Polling Errors in the 2020 U.S. Presidential Election: An Examination of State-Level Panel Data

Josh Pasek, *University of Michigan*

Kenneth Winneg, *University of Pennsylvania*

Kathleen Hall Jamieson, *University of Pennsylvania*

R. Lance Holbert, *Temple University*

Daniel Romer, *University of Pennsylvania*

Bruce Hardy, *Temple University*

Matthew Levendusky, *University of Pennsylvania*

Andrew Renninger, *University of Pennsylvania*

In 2020, a series of state-level pre-election polls indicated that Joe Biden would handily beat Donald Trump across many of the states expected to be pivotal in the election. Instead, election margins in these states were far closer than anticipated, leading to widespread perceptions of a polling failure. This study leverages probability-based panel survey data from Michigan, Wisconsin, Pennsylvania, and Florida to test a series of hypotheses about the source of these apparent polling errors. Specifically, we test the possibilities that errors (1) were attributable to the difficulty of recruiting Trump supporters, (2) were associated with a differential tendency to participate in surveys among eventual Trump voters, (3) could be attributed to variations in the success of Republican and Democratic campaigns in their voter mobilization efforts, (4) were associated with inaccurate expectations about which Americans would vote in the election, and (5) could be attributed to so-called "shy trump" voters who were unwilling to express their preferences in the survey. To examine these possibilities, we compare respondents reported election behaviors with their pre-election preferences, third party records on whether or not they voted, the distribution of observed votes in their precincts, and their tendencies to respond to pre-election survey waves. Based on the distinctions we observe, we discuss the extent to which the errors observed were consistent across states as well as whether any weighting or imputation strategy could be expected to properly adjust for the biases observed.

Evaluation of Polling Accuracy in 2020 Election Cycle Polling

Alex Von Hagen-Jamar, *Global Strategy Group*

We will conduct an in-depth analysis of our surveys across the country to evaluate the accuracy of polling in the 2020 cycle. Despite the broad adoption of weighting by education, polling in the 2020 election cycle appears to have systematically underestimated Republican support, much like it did in 2016. We focus on two potential sources of error: inaccurate projections of the 2020 general electorate due to unexpected turnout (coverage error) and changing response rates leading to underrepresentation of important groups of voters (non-response error). By examining response rates and comparing survey samples to actual turnout in the election, we will identify if and where unexpected turnout among specific groups of voters led to polling error in our surveys. Similarly, we will also investigate whether changing response rates led to different groups being underrepresented as we approached Election Day, particularly as response rates to phone surveys increased during the coronavirus pandemic. By understanding which groups were under-represented in polling and why, researchers can explore methods to reach these voters and ensure that weighting corrects for lower response rates.

2020 Election Polling: Deconstructing Our Error

Donald Levy, *Siena College*

After the November election facing criticism of ‘another polling debacle’ many AAPOR members were forced to immediately explain average errors of around five points virtually all in one direction. Now that the dust has settled, this paper looks at our polling, the NY Times/Siena Pre-election polling in an effort to determine how much error was in our polling, its causes and to propose potential remedies. Included in this analysis is an investigation of error in estimation by demographic group nationally and in key battleground states, evidence or lack thereof of the ‘shy’ voter theory – among those that responded, and the role played by differential non-response. Additionally, we explore potential effects of the coronavirus pandemic, early voting as well as potential mode effects. We also discuss the questions that many of us have faced as our 2020 polling was analyzed and debated both by those with and without experience in public opinion research. Did polling that as a whole over-estimated Biden’s performance suppress the Trump turnout? And, ‘How, given polling’s errors, can we believe you in the future?’

The analysis in this paper benefits from the input of not only those that worked on the NY Times/Siena polling but other pollsters, academics and a group of undergraduate students. We offer a series of potential actions that may be able to lead to more accurate estimates in future elections.

Don’t Let Polling Errors Ruin Your Day: Effective Methods To Adjust and Communicate Polling Estimates

Xinyue Zhang, *SurveyMonkey*

Tim Gravelle, *SurveyMonkey*

Laura Wronski, *SurveyMonkey*

Jack Chen, *SurveyMonkey*

It is known for decades that procedures like calibrating to demographics characteristics and administering Gallup's Likely Voter Model are widely used by pollsters to adjust polling estimates for election. However, the unique nature of the 2020 Election - conducted during the Coronavirus pandemic with massive early mail-in ballots - posed many challenges and uncertainties in identifying voter electorate. This paper will discuss two simple and practical approaches we employed here at SurveyMonkey that can reduce polling error by more than 30%: 1) anchor sample at initial weighting stage with vote 2016 calibration, 2) adjust sample at final stage to likely voters through probability-based models. We will compare the performance of the combination of these two approaches against traditional polling methods, based on average absolute polling error. In addition to providing insights into individual probability of voting, the probabilistic likely voter model is proven effective in understanding and communicating polling estimates, through model-assisted scenario buildings that fully explore how shifts in turnout among different demographics groups could swing the presidential approval.

Friday, May 14, 2021 | Day 4 of Conference

5/14/2021, 11:30 AM - 1:30 PM

Concurrent Session H: Prerecorded Sessions

Mini-Conference: Measuring Health-Related Behaviors During COVID-19

Predictors and Effects of Perceived Threat of COVID-19

Timothy Vercellotti, *Western New England University*

One of the challenges that public health officials have faced in the COVID-19 pandemic has been convincing individuals to take seriously the potential dangers that the virus poses. What are the factors that predict whether someone perceives a threat from COVID-19? Connecting a human face to the statistics on infection and mortality may make the difference. In this research I hypothesize that personally knowing someone who has been diagnosed with COVID-19, or who has died from COVID-19, increases an individual's perceived threat of also contracting the illness. I also hypothesize that perceived threat shapes support for government responses to COVID-19. I test these hypotheses using a statewide survey of Massachusetts residents in fall 2020, as the state entered its second wave of infections. Preliminary results indicate that personally knowing someone who has died from COVID-19 was related to increased perceptions of threat from the virus, and that women and Democratic voters also were more likely to express concern about the illness than men and Republicans. Feelings of threat also were related to whether an individual supported the state government's response to the pandemic, and whether someone would be willing to receive a vaccine to protect against a COVID-19 infection. In addition, perceived threat predicted views that engaging in a variety of post-lockdown activities, ranging from dining in a restaurant to attending worship services indoors, posed significant safety risks. Results of the study could provide guidance for ways to more effectively frame public health information related to COVID-19 and other communicable diseases.

Let's Spend the Night Together: A Challenge for Medically Optimal Coronavirus Social Distancing Policies

Mariah Evans, *University of Nevada, Reno*

Jonathan Kelley, *International Survey Center*

Sarah Kelley, *Child Trends*

During the coronavirus epidemic, much research attention has focused on social distancing during the day and at public venues, but the potentially dangerous practice of overnighting with people who are not part of one's household has been sorely neglected. Overnighting brings a person from outside the household into proximity indoors for an extended period, usually without any COVID-19 precautionary measures. The exposure is worse for romantic overnight guests, but even visiting family and friends (or roommates' romantic visitors) bring their germs with them. This research shows that overnighting has been common throughout the epidemic: Over half of all Americans had at least one overnight guest in the past month and about the same number (but not always the same people) spent at least one night in the past month at somebody else's home. The consequences are striking: People who overnight are more than twice as likely to report that they have had COVID-19. Who overnights? The practice is much more common among young adults (over 60% for those 25-34) than among their seniors (dipping below 30% for those age 65 and over). Moreover, risk seekers are hugely more likely to overnight than are their otherwise similar risk averse peers. Data: International Social Science Survey, Round 21, Cohorts 1-5, collected April – July 2020, N=4,815. Methods: Descriptive statistics and multiple regression.

Don't Stand So Close to Me: Examining Consistency Between COVID-19 Mitigation Attitudes and Behaviors Over Time

Thomas Brassell, *ICF*

John Boyle, *ICF*

James Dayton, *ICF*

Randy ZuWallack, *ICF*

The COVID-19 pandemic has had a significant health and economic impact throughout the United States, with the death toll exceeding 230,000 at the beginning of November 2020 and over 11 million people unemployed as of October 2020. Although recent data on a COVID-19 vaccine has been positive, it is expected that there will be limited quantities available initially – meaning controlling the virus will be reliant upon people continuing to engage in behaviors that mitigate the spread of the virus (e.g., wearing masks, socially distancing, etc.). However, as we reach the ninth month since COVID-19 was officially declared a pandemic in the U.S., there are concerns that the current spread of coronavirus is increasingly due to adherence to mitigation behaviors. Starting in March, we administered a national survey using a Census-balanced non-probability mobile panel to collect 1,000 completed surveys a month. The current study examines the overlap between respondent attitudes towards behaviors that limit the spread of COVID-19 and their reported related-behaviors. Specifically, respondents were asked about the importance of mitigation policies and behaviors (e.g., closing non-essential businesses, requiring face masks) and the number of days that they specifically engaged in certain behaviors and activities (e.g., wearing a face mask in public, eating at sit down restaurants, shopping for non-groceries). Our results find that the more important a respondent perceives a mitigation policy/behavior, the more likely their behavior will be consistent with that perspective. However, over time, behaviors such as eating at sit down restaurants, shopping for non-groceries, etc. have increased even among those that indicated that keeping non-essential business and staying 3-6 feet away from others were important. This suggests that adherence to mitigation behaviors requires more than just an understanding of the perceived importance. Implications for the current pandemic surge to be discussed.

Assessing COVID-19 Pandemic Concerns, Attitudes, and Behaviors Among Northern Colorado Adults

Maria Macpherson, *Health District of Northern Larimer County*

Suman Mathur, *Health District of Northern Larimer County*

Shaeyla Davis, *Health District of Northern Larimer County*

Sue Hewitt, *Health District of Northern Larimer County*

In Fall of 2019, the Health District of Northern Larimer County conducted a triennial Community Health Survey (CHS) of 2,532 adults in Larimer County, Colorado using address-based sampling and a mail-web mixed-mode approach. To assess the impact of the COVID-19 pandemic on the populations' overall physical, mental, behavioral, and economic health, the Health District invited 2,463 of the original respondent households to complete a follow-up survey in July of 2020. A total of 1,512 surveys were completed, with 1,239 of these surveys matched to cases in the 2019 CHS dataset. Both the 2019 and 2020 CHS included an open-ended question that asked respondents what they thought was the biggest health concern in Larimer County. Using Dedoose qualitative analysis software, 13 themes emerged, including concerns about lack of compliance with COVID regulations, school re-opening, the economy, politics and societal division, information and the media, and the strictness of COVID-19 regulations. Five themes were identified in both the 2019 and 2020 survey: concerns about mental health, health-care, cost of living, people experiencing homelessness, and the environment. Adults in Larimer County reported high levels of cooperation with COVID-19 infection control measures, including washing hands, social distancing, mask wearing, and reducing time in non-essential public places. However, close to half the population (45%) reported visiting with family members not in the same household in the past 30 days. A high proportion of adults expressed concern about receiving a COVID-19 vaccine when/if it becomes available. 66% of adults said they would get vaccinated but open-ended responses showed that respondents were concerned about the validity of deeming the vaccine "safe and effective". Finally, comparisons were made to repeated measures between both surveys, including mental health, employment, concerns over affording food and housing, and need and use of social services.

What Are the Implications of Polling Errors for Public Health Survey Estimates? An Assessment Exploring Potential for Bias Using the COVID-19 Health Monitoring Survey

Deirdre Middleton, *ICF*

Thomas Brassell, *ICF*

Randy ZuWallack, *ICF*

Rachel Kinder, *ICF*

James Dayton, *ICF*

Survey data is used to inform policies and practices which protect Americans health and wellbeing. However, as response rates decline, we face concerns that nonresponse bias may cause mismeasurement of attitudes, beliefs, and behaviors. Both the 2016 and 2020 election results showed that political polls, many of which use similar methodologies to public health surveys, consistently underestimated the number of people voting for Republicans. Additionally, recent studies show partisan leanings are linked to health outcomes during the COVID 19 pandemic. If Republican-leaning Americans are 1) as underrepresented in public health surveys as they were in political polls, and 2) exhibit different health behaviors, attitudes, and beliefs, then this could cause measurement errors for estimates critical to assessing and protecting American health. To begin to examine this issue, we explore respondent characteristics for the ICF COVID19 Monitoring survey, a cross-sectional, national survey that collected monthly waves of data from approximately 1,000 mobile panel members since March 2020 and includes respondents' political affiliation. We assess partisan differences in both highly politicized health behaviors, such as mask use, as well as less partisan behaviors such as drinking. We also assess if partisan differences changed over time, while considering major events that may have impacted these differences. In light of these partisan differences in estimates, we consider the potential impacts of under coverage and nonresponse of conservative persons in public health surveys.

Tracking COVID-19 Understanding, Concerns, Health Risks, Behaviors, and Attitudes

Laura Wronski, *SurveyMonkey*

Jon Cohen, *SurveyMonkey*

Jack Chen, *SurveyMonkey*

Once the coronavirus pandemic made waves in the U.S. in February 2020, SurveyMonkey began tracking the awareness, concerns, health risks, behaviors, and attitudes of people in the U.S., U.K., and Canada on a continuous basis, oftentimes collecting responses from close to 100,000 people per week. In this paper, we'll compile some of the key findings from this vast pool of data, sourced from our online river sample and weighted to be nationally representative. We'll keep a particular focus on the information that makes use of the scale of data collected to drill down to small geographic areas (by state, county, or zip) or to examine responses by subgroups and hard-to-reach populations (e.g. gender non-binary individuals, young black men). With the ability to show trends over time, we'll also pull in external data on case counts at the national and local level to see if attitudinal trends were correlated with the spread of the virus. Finally, because some data were collected in partnership with COVID Near You, a team of epidemiologists and public health researchers at the Harvard Medical School and Boston Children's Hospital, we'll tie in personal health information for our respondents to show how concerns and behaviors differed depending on one's underlying health condition.

The Impact of the COVID-19 Pandemic on the Social and Psychological Well-Being of Americans

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

With support from the National Science Foundation NORC at the University of Chicago conducted a three-wave, longitudinal survey of American adults using NORC's AmeriSpeak panel. By adopting baseline measures from the 1963 Kennedy Assassination Study, the 2001-02 National Tragedy Study, and the 1972-2018 General Social Surveys, it was possible to measure the social and psychological well-being both before the pandemic and after its emergence. Moreover, by interviewing respondents at three points from May to August, changes as the pandemic continued were monitored. Key findings included that 1) unhappiness, loneliness, and other measures of negative psychological conditions notably increased and in some cases climbed to record levels, 2) these measures uniformly remained stable at their elevated levels as the pandemic progressed, neither improving nor further worsening, 3) the greater

distress was related both directly to the pandemic's health impact and indirectly to its negative economic consequences, and 4) psychosomatic symptoms and stress levels varied notably across socio-demographics groups.

Conducting a Follow-Up Community Health Survey to Improve Local Public Health Response

Suman Mathur, *Health District of Northern Larimer County*

Sue Hewitt, *Health District of Northern Larimer County*

Maria Macpherson, *Health District of Northern Larimer County*

David Brown, *Colorado State University*

During Fall of 2019, the Health District of Northern Larimer County conducted a triennial Community Health Survey (CHS) of 2,532 adults in Larimer County, Colorado using address-based sampling and a mail-web mixed-mode approach. Given the Health District's involvement in community health planning and COVID-19 response, it was necessary to have updated data that accurately reflected the impacts of COVID-19 at the local level. Beginning in July 2020, the Health District invited 2,463 of the original respondent households to complete a follow-up survey to assess the impact of the pandemic on the population's overall physical, mental, social, and economic health. Push-to-web letters were sent to 1,841 addresses where the original survey was completed online. Printed surveys were mailed to 633 addresses where the original survey was completed on paper. Envelopes were hand-addressed and used adhesive stamps rather than metering to encourage opening and reduce cost. A \$1 pre-incentive was included in the initial mailing. Non-responders to both groups were sent a reminder postcard 10 days later. In late August, a final printed survey with online instructions was sent to 629 non-responding households where the original respondent was aged 60 or younger. In total, 1,512 surveys were completed: 1,104 surveys online and 408 by mail. Forty-nine survey responses were removed as duplicates, yielding 1,463 unduplicated surveys. After removing those who identified as a new respondent and those with mismatched demographic characteristics, 1,239 surveys were matched to cases in the 2019 CHS dataset. These responses were re-weighted to be generalizable to the adult population of Larimer County, using previously collected demographics of non-responders to correct for response attrition. Responses indicated changes in mental health status, drinking behavior, health insurance status, as well as a number of socioeconomic challenges.

Perceptions of the COVID-19 Pandemic in Virginia

Farrah Stone, *Virginia Commonwealth University*

The public policy response to the coronavirus (COVID-19) pandemic has impacted most aspects of daily life for all of us, from our ability to work and go to school to our ability to gather freely with friends and family. Acceptance of policies with such significant individual impact requires that the public have a level of trust and see the legitimacy in the measures and the ones making these decisions. This paper will examine state level polling data of adult residents in Virginia from April to October of 2020 asking respondents their views on how the president and Virginia governor handled the response, mask mandates and efficacy, potential vaccines and in-person schooling. There were significant regional differences in opinion and gender and race also played a role in attitudes toward policies and leaders. Since COVID-19 cases continue to surge, continued response will be necessary. These findings help to characterize the public's position towards and acceptance of current strategies and gives policymakers a view of where trust and legitimacy need to be increased if they hope to have the public's behavior follow.

Developments in Text Message and Mobile Phone Surveys

Replacing Online Non-Probability Panels With Short Message Service (SMS)

James Medick, *Precision Opinion*

On April 20, 2020 three identical surveys were launched using two sample bases. One sample was a random pull using one of the largest Online Non-Probability Panels companies. The second was a sample pull using a random selection of mobile telephone numbers. Both samples were drawn for three markets: national, the state of Ohio, and Jefferson County in Ohio. The responses were allowed to fall naturally with no quotas for gender nor age required. The panel house's sample was sent using a router. The cell phone sample invites were sent using Precision SMS Peer-to-Peer TCPA Compliant technology. The study was designed to stay in the field no longer than 3-days (April 20 through April 23, 2020). A key discovery was PrecisionSMS Text-to-Web methodology reaches an untapped market of respondents not found in Online Non-Probability Panels. Only 6% of SMS respondents to the national study were members of a panel. Non-probability panels for small markets (in this study 30,000 households) were ineffective at best. Using multiple Non-Probability Panels over a 3-day period 21 completes were generated while SMS Text-to-Web produced 116 completed interviews with 91 obtained in the first 6 hours. Key findings were: 1) SMS methodology delivered a higher participation rate than non-probability panels; 2) SMS Text-to-Web effectively reaches small markets that are underserved by panels; 3) SMS Text-to-Web delivers an accurately representative sample in all markets; 4) In the core market of 25-54 SMS Text-to-Web reflected the US Census in both the age and gender categories.

Accuracy of RDD-Sampling SMS-Invitation Web-Push Survey

Sebastian Kocar, *Australian National University*

Survey data collection underpins a large proportion of social science research across multiple disciplines, but is increasingly difficult: response rates are decreasing, and methods to sample national populations are growing more expensive and complex. In this presentation, I will present results of a fairly new approach to online data collection which benefits such as simplicity, time and cost effectiveness, but possibly on the expense of accuracy due to undercoverage of people without smartphones or Internet access and differential nonresponse. Traditionally, cross-sectional general population surveys use many other recruitment approaches: mail outs, telephone calls, or face-to-face contacts. Our approach uses random-digit selection (RDD) of mobile phone numbers combined with SMS invitations, with respondents asked to complete the survey online and a URL directing them to an online questionnaire. The data collection took place in Australia which remains one of the countries where no consent from a person is required before sending automatic SMS survey invitations. The online questionnaire consisted of primary and secondary socio-demographic questions, as well as questions on the use of internet, technology, health and wellbeing, life satisfaction, and personality traits. The vast majority of items had corresponding nationally representative benchmarks from large-scale surveys in Australia, and they were used to estimate accuracy of our approach to online data collection. Although text messaging had been, generally speaking, predominantly used only as an additional communication channel and for sending reminders, it showed to be an interesting alternative to other more complex and expensive survey designs. While the response rate was to no surprise considerably lower than in face-to-face, postal, telephone or probability-based online panel surveys, it did not translate into a major loss in accuracy relative to the nationally representative benchmarks and in comparison to probability-based surveys from a different benchmarking study.

Measuring Everyday Behavior With Smartphone Sensors: Potential Bias Due to Device Usage

Florian Keusch, *University of Mannheim*

Alexander Wenz, *University of Mannheim*

Frederick Conrad, *University of Michigan*

Smartphones have become central to the daily life of many people and are often present in the same contexts as their users. Researchers have started to take advantage of this phenomenon by inferring certain behaviors from the raw data of smartphone sensors, such as mobility, physical activity, social interaction, and sleep. For example, that a person is sleeping might be inferred from the fact that their phone is idle and that it does not detect sound or light. The success of the inferential step from raw smartphone data to behavioral outcomes depends on how smartphone owners use their device, including whether the phone is in close proximity to the owner throughout the day, for how long it is turned on, and whether it is shared with others. Against this background, we surveyed two samples of German smartphone owners – one from a probability based online panel (n=3,956) and one from a nonprobability panel (n=2,525) – on how they use their smartphones in everyday life, with a focus on three behaviors that are regularly inferred from smartphone sensor data: physical activity, mobility, and sleep. We identify groups of smartphone owners who report using their device in a way that – at least in theory – would allow researchers to measure these behaviors and those who do not. Preliminary results suggest that smartphone usage such as the amount of time the device is turned on and the way the device is carried around varies systematically by age, gender, and self-reported smartphone skills. In addition, we find that smartphone use correlates with the level of self-reported physical activity, indicating potential for bias if measured with smartphone sensors.

Text-Only Survey Notification? Lessons Learned From the U.S. Census Bureau

Elizabeth Nichols, *U.S. Census Bureau*

Shelley Feuer, *U.S. Census Bureau*

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

Rachel Gliozzi, *U.S. Census Bureau*

To measure user satisfaction with the 2020 Census online questionnaire, the U.S. Census Bureau re-contacted a sample of census respondents via text message and asked them to complete a short online survey. The representative nationwide sample was drawn from online responses to the actual census where a valid cell phone number was provided. In the hope of capturing a variety of experiences, we sampled early responders who reported in March through late responders in October; responders who completed the census in one sitting and those who broke off before completing the census; responders who used a PC or Mac to complete the census, and those who used mobile devices; and finally those who had their mailing material and Census ID and those who did not. Each sampled phone number received up to 3 text messages throughout an 11 day field period. We embedded a fully crossed experiment within the sample to determine whether the time of day the text was sent influenced overall response or the satisfaction results, testing a 12 noon text against a text sent at 6 pm. The time of day was consistent across the three texts. Early indication is that the texts sent 12 noon generate more response than those sent at 6 pm, with no difference in satisfaction between the two groups of responders. This presentation shares final response rate results by experimental group, whether those rates are affected by the sampling characteristics, whether there were any substantive differences in satisfaction answers between the two groups, and other lessons learned from a text-only survey notification contact method.

The Smartphone Usage Divide: Differences in People's Smartphone Behavior and Implications for Mobile Data Collection

Alexander Wenz, *University of Mannheim*

Florian Keusch, *University of Mannheim*

Smartphones have become an integral part of many people's daily lives, with ownership rates steadily increasing over the last decade. Survey practitioners have started to use these devices as new measurement tools, not only to implement mobile web questionnaires and diaries but also to capture new forms of data from the in-built sensors, such as GPS positioning or acceleration. Existing research on coverage error in these types of studies has distinguished between smartphone owners and non-owners. With increasing smartphone use, however, the digital divide of the "haves" and "have-nots" has shifted towards a second-level digital divide that also considers how people are able to use smartphone technology. Since smartphone-based studies require participants to engage with their device, such as downloading an app or activating location tracking, researchers not only need to understand which subgroups of the population have access to such technology but also how they use it. In this paper, we examine people's smartphone usage pattern and its implications for mobile data collection. We collected survey data from six samples of smartphone owners in Germany and Austria between 2016 and 2020 (three probability samples: $n_1=3,956$; $n_2=2,186$; $n_3=632$; three nonprobability samples: $n_4=2,623$; $n_5=2,525$; $n_6=1,214$) on frequency of smartphone use, self-reported smartphone skills, and activities carried out on a smartphone. To identify different types of smartphone users, we conduct a latent class analysis (LCA), which classifies individuals based on their similarity in smartphone usage patterns. We address the following research questions:

- Which types of smartphone users can we identify?
- Do the different types of smartphone users vary systematically by socio-demographic characteristics, privacy concerns towards research activities on a smartphone, and key survey variables?
- How does the composition of the smartphone usage types change over time?

Implementing Probability-Based Text Message Survey

Chintan Turakhia, *SSRS*

Jennifer Su, *SSRS*

Text message surveys have a unique advantage in conducting short, quick turn-around surveys in a cost-effective manner. Text message surveys can also be quite effective in reaching harder-to-reach populations. To-date, the use of this methodology has been very limited due to concerns of low participation rates and representativeness of text message-based surveys. SSRS has launched the first TCPA-compliant nationally representative probability-based text message-based panel. We conducted two separate text message surveys during the 2020 general election season, one running in parallel with a national RDD telephone survey and one running in parallel with a post-presidential debate reaction survey. In this brief, we provide demographic and substantive comparison of phone- and text message-based methodology. We also provide methodological guidance for planning and implementation of text message-based surveys.

Survey Redesigns and Other Methods Experiments

Impact of Survey Design Changes on the Quality of Health Care Utilization Estimates. Evidence From the Medical Expenditure Panel Survey and the National Health Interview Survey

Frances Chevarley, *Survey Research Methods Consultant, Retired from AHRQ*

Adam I. Biener, *Lafayette College*

Sadeq Chowdhury, *Agency for Healthcare Research and Quality*

Evaluation and development of public policies such as the Affordable Care Act and new proposed reforms rely on data on the utilization and costs of health care. Actionable data often come from household surveys, where interview practices and survey design are critical factors to ensuring reliable and consistent estimates of national trends. This paper analyzes health care utilization for the Medical Expenditure Panel Survey (MEPS) respondents matched with their previous year responses in the National Health Interview Survey (NHIS), and NHIS estimates for MEPS non-respondents. These surveys are essential for understanding national trends in health care utilization, and have recently undergone major improvements to their survey methodology with the intent of maintaining the quality of responses while mitigating declining response rates observed in many federal surveys.

MEPS is a nationally representative panel survey studying health care use, access, expenditures, source of payment, insurance coverage, and quality of care sponsored by the Agency for Healthcare Research and Quality. Because the MEPS sample is drawn from the previous year's NHIS sample (NHIS is sponsored by the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC)) it is possible to compare the current year estimates of MEPS respondents with estimates for the same respondents in NHIS in the previous year and also with NHIS estimates of MEPS non-respondents.

A number of initiatives designed to improve the reporting of health care utilization were implemented in MEPS starting in 2013. This paper studies the efficacy of these survey improvements by analyzing trends in ambulatory care visits in the MEPS and NHIS prior to and after the improvements to determine if their trends have returned to their previous trend and become more closely aligned as they had been prior to 2010.

Piloting the Transition From RDD Telephone to ABS Multi-mode Data Capture in the New York City Community Health Survey

David Roe, *Abt Associates*

Michael Witt, *Abt Associates*

Michael Sanderson, *New York City Department of Health and Mental Hygiene*

Steven Fernandez, *New York City Department of Health and Mental Hygiene*

Nicholas Ruther, *Abt Associates*

Ashley Bradbury, *Abt Associates*

John Sokolowski, *Abt Associates*

Amber Levanon Seligson, *New York City Department of Health and Mental Hygiene*

Since 2002, the New York City Community Health Survey (CHS) has been integral to the New York City Department of Health and Mental Hygiene's (DOHMH) mission to protect and promote the health of New Yorkers. The CHS has experienced declines in response rates using an RDD landline and cell phone sampling frame and methodology, leading to decreased efficiency and increased potential bias. In 2021, the CHS is moving to a multi-mode data collection using Address Based Sampling (ABS) and self-administered web and mail surveys, along with an inbound telephone survey. To develop and test new methods to obtain representative samples of adult New Yorkers using ABS and self-administered surveys, DOHMH contracted Abt Associates to design and conduct a pilot study as an add-on to the 2020 CHS. The pilot included qualitative and quantitative tests to develop and refine contact materials and surveys, examine sample design, evaluate survey participation at each stage of respondent contact and identify and mitigate mode effects on survey measures. The pilot also targeted English, Spanish, and Chinese respondents using sample frame and auxiliary information and tailored study materials to test their impact on survey response rates. This presentation will summarize many lessons learned related to the design, operation and results of the pilot that can inform other jurisdictions' efforts to transition from

RDD to ABS sampling frames and their related modes. We present findings related to sampling accuracy, targeting subgroups of interest with auxiliary frame data, participation rates at different stages of contact and nonresponse follow-up, and mode effects when comparing self-administered results with RDD results from the 2020 CHS. We also discuss incentive choice, mode preference, and resulting adjustments to the 2021 CHS design.

Adapting a Longitudinal Face-to-Face Survey for CATI and CAWI Administration

Daniel Lawrence, *NORC at the University of Chicago*

Ipek Bilgen, *NORC at the University of Chicago*

Kelly Pudelek, *NORC at the University of Chicago*

Lauren Sedlak, *NORC at the University of Chicago*

Anna Wiencrot, *NORC at the University of Chicago*

Face-to-face surveys have faced significant disruptions and delays due to the ongoing COVID-19 pandemic. The National Social Life, Health, and Aging Project (NSHAP) is a longitudinal, population-based study of adults born between 1920 and 1965 and their romantic partners that has been conducted face-to-face since its first wave in 2005-2006. For the current wave, NSHAP was already developing ways to collect complex questionnaire and biomeasure data remotely as a means to reduce costs; the need for remote data collection was further heightened by the pandemic. In late 2019, NORC conducted a pretest of the NSHAP questionnaire adapted for remote administration by phone, as well as respondent self-collection of biomeasures. As the pandemic forced the delay of the fourth face-to-face wave of NSHAP data collection, NORC also adapted the NSHAP questionnaire for web administration and conducted a large-scale pretest of the feasibility and data quality of respondent self-report of the most complex questionnaire sections, including a complicated social network roster, an expanded battery of cognitive measures, and a medication log. This paper describes the methodological and operational challenges in adapting a lengthy, complicated longitudinal questionnaire for CATI and CAWI modes. We describe in particular the most complex elements of the NSHAP questionnaire and how we balanced methodological rigor while minimizing respondent burden. We compare data from these two data collection periods against NSHAP's face-to-face benchmarks to examine overall data quality and item-level non-response. The paper concludes with a discussion of lessons learned about asking a longitudinal sample of older adults to try new data collection protocols as well as methods for ensuring data quality.

Understanding and Communicating the Impact of a Methodological Redesign on Key Health Survey Indicators

Brian Wells, *UCLA Center for Health Policy Research*

Todd Hughes, *UCLA Center for Health Policy Research*

Ninez Ponce, *UCLA Center for Health Policy Research*

The decline of telephone surveys due to low response rates and cultural shifts in telephone use has motivated many to consider implementing major methodological changes to help with response and cost (Olson et al., 2019). Following an extensive evaluation and experimentation process, the California Health Interview Survey (CHIS) fully transitioned to a mixed-mode survey design using an address-based sampling frame in 2019. A concern for many data users of CHIS was the impact the redesign, in particular the introduction of web data collection, would have on trending for key outcomes. To help navigate these changes, CHIS underwent an evaluation to try to determine whether changes observed were due to actual changes in the population over time or whether it was related to changes in the sampling frame and survey mode. Of particular concern were changes related to sample composition differences due to the sampling frame and mode, and measurement changes related to mode, in particular including previously unread responses, decreased satisficing responses, and serial position effects. We use a statewide pilot conducted in 2018 (Hughes et al., 2019) as a reference point to help guide interpretations. This paper will examine multiple key indicators (e.g., self-rated health, smoking status, hypertension, health insurance, gender identity) from CHIS and observe subgroup differences (age group, race/ethnicity, poverty level) for evidence of where breaks in trend are warranted and where trends should be interpreted with caution. We also discuss how we communicated methodological changes and their impacts on trends to data users.

Think About Your Household: Does Reminding Respondents About Household Size Improve Within-Household Selection in General Population Self-Administered Surveys?

Kristen Olson, *University of Nebraska-Lincoln*

Jolene D. Smyth, *University of Nebraska-Lincoln*

Probability samples of the general population require accurate within-household selection using probability or quasi-probability methods. Accurate within-household selection is difficult in self-administered surveys where the selection task is in the hands of the sampled householder. Although improvements have been made to quasi-probability within-household selection methods (e.g. next birthday method with confirmation question, Olson and Smyth 2017), roughly one-quarter of households with 2+ adults misselect a respondent. One reason may be that the household composition is not evaluated during this process. In this paper, we test whether reminding the sample householder to think about household composition yields higher rates of selection accuracy. Data are from the Nebraska 2020 Survey, a general population mixed-mode web-mail survey about COVID-19 and race relations of $n=10000$ Nebraska households selected from an address-based frame collected September-December 2020. We randomly assigned households to receive a cover letter containing an extra sentence to think about all of the adults in their household before the next birthday selection instruction versus the standard letter without this composition sentence. We also randomly assigned households to receive a questionnaire with two questions about household composition immediately preceding the next birthday confirmation question at the start of the questionnaire versus the standard placement of these questions in the demographics section toward the end. As of November 2020, response rates to the letter mentioning household size are one percentage point lower than the letter that does not mention household size (25.8% vs. 26.8%, $p=0.05$), and the number of adults in the household is not different across the two web questionnaire versions ($p=0.83$), indicating no deleterious effect of these two design decisions. We further evaluate response rates, composition of the respondent pools across experimental conditions, selection accuracy, and important survey estimates across the experimental conditions. We conclude with implications for survey practice.

Improving the Cost Effectiveness of a Targeted Mailed Survey to Determine Household Eligibility in a Face-to-Face Survey

James Wagner, *University of Michigan*

Wen Chang, *University of Michigan*

Heidi Guyer, *RTI International*

Many surveys have a target population that is a subset of the general population. One common tactic for such surveys is to “screen” households in order to identify households that contain eligible persons. This process can be expensive for a face-to-face survey, but generally produces high response rates. Starting with a mailed screening survey before introducing face-to-face attempts would be less expensive, but might lead to lower response rates. We examine the use of a mailed screening survey for the National Survey of Family Growth. In order to be eligible, households need to contain at least one person between the ages of 15 and 49. Previously, we reported the results of a first set of experiments that demonstrated the viability of this approach. Here, we report the results of a new experiment aimed at evaluating the cost effectiveness of alternative formats for the mailed component. In the previous experimentation, we employed an expensive mailer that obtained relatively high response rates. In the new experiment, we evaluate the response rates, costs, and observed eligibility rates for a less expensive mailing option. Preliminary analyses suggest that the low-cost mailer produces somewhat lower response rates than an expensive mailer but is a more cost-effective option.

The Less the Merrier? Investigating the Number of Household Members Asked at Rostering

Tzu-Jou Wan, *American Institutes for Research*

Jiashan Cui, *American Institutes for Research*

Danielle Battle, *American Institutes for Research*

The household rostering stage of the survey administration determines the participant's eligibility and survey weighting; however, the optimal number of people a survey should ask about is not agreed upon. Battle et. al (2014) showed that in a mailed survey, asking information about all 10 household members during rostering stage didn't make much difference in unit response, compared to asking information about 5 children only. The study evaluates this topic on a web survey to what extent the number of people asked at screener stage, and whether asking about children versus both children and adults in a household make a difference in sample composition of children. This study uses data from the 2016 and 2019 National Household Education Survey (NHES), an address-based sample of U.S. households. At the screening phase, sampled addresses were asked to provide the name, sex, and age of persons residing at an address. In 2016, the NHES screener asked information of up to 10 household members at the address, including both children and adults. In 2019, the screening stage only asked information of up to 5 children at the address. We will examine the number of children reported between the two years, and whether children of a specific age range were more likely to be reported in one year than the other, in addition to the completeness of the screener items. All the results will be compared between web and paper mode. Furthermore, there was a series of household member items at the topical phase of the NHES about the number and relations of persons living in the household. These reports at topical stage will be compared to the number and type of persons reported at the screening stage. The results of the study will inform the screener design of household surveys targeted at children.

Politics, Information, and Misinformation

Rural-Urban Divides, Partisanship, and Misinformation in Science: How Rural Resentment Moderates the Effect of Partisanship on Misinformation in Scientific Issues

Tomoko Okada, *University of Wisconsin-Madison*

In recent years, the concept of rural-urban divides has become increasingly prevalent and important to understand the current U.S. political landscape. The rurality in the place of residence is shown as a significant predictor of voting behaviors. Cramer (2016) explained a cause of rural-urban divides that people have developed place-based identity based on their place of residence. Among those who reside in rural areas, such a place-based identity has often been closely associated with “rural resentment,” perceived unfairness and anger against the central government and residents in metropolitan areas. Given that anger can increase the susceptibility to misinformation (Weeks, 2015), we have to consider the possibility that “rural resentment” can make people more susceptible to misinformation. If this is the case, individuals’ vulnerabilities to disinformation can differ based on residential patterns, resulting in inequalities between rural and urban residents in the process of democracy. This study examined such a possibility by relying on the 2019 American National Election Study survey data. In particular, the associations between partisanship, rural resentment, and the level of being misinformed were investigated with control variables using the three topics of misinformation: 1) vaccination as a cause of autism, 2) safety of genetically modified organisms (GMO), and 3) global warming. Preliminary analysis demonstrated the effects of these independent variables vary across the three topics. That is, partisanship, rural resentment, and their interaction effect were not significant in predicting the level of being misinformed about the link of vaccination and autism. In predicting the level of misinformation on GMOs, only the main effect of rural resentment has a significant and positive correlation. Also, the increased level of misinformation on global warming was significantly associated with partisanship, rural resentment, and their interaction, implying that rural resentment can increase the susceptibility to misinformation on global warming among Republicans.

Misinformed or Uninformed? The Prevalence and Consequences of Certainty in Political Misperceptions

Brian Guay, *Duke University*

There is widespread concern today that much of the public is misinformed, holding factually inaccurate beliefs that they believe to be correct. However, few studies to date have measured the certainty with which political beliefs are held. Those that do use methods that cannot meaningfully distinguish actual beliefs from uncertain guesses and have led to differing conclusions about the prevalence of certainty in inaccurate beliefs. In this paper I introduce a new measure of certainty that provides meaningful context to existing self-report measures of certainty and adjusts for differential item functioning (i.e., differences in how respondents use response scales). I show that past work exaggerates the degree to which the public is misinformed and that inaccurate beliefs are four times more likely to represent uncertain guesses than actual beliefs. I then provide the first empirical test of the widespread expectation that certain beliefs are more likely to resist correction and influence attitudes than uncertain ones. The findings presented here have implications for how researchers interpret findings from the growing body of research that documents and attempts to correct misperceptions.

Civic Engagement: The 'Bobi Wine' Social Media Effect

Juma Kasadha, *City University of Hong Kong*

Technological access and adaption in Uganda have shaped and increase the way social media users' access, share and discuss the different social-political activities. In Uganda, the musician [Bobi Wine] turned into a politician in 2017; and protested against the government's introduced Over the Top social media tax. The tax limited citizens' access to social media networks unless they paid the daily subscription fee of Ugandan shillings 200 (US\$0.054). In examining the Bobi Wine's social media effect, N=1151 respondents were randomly sampled. This study seeks to examine whether (i) Bobi wine's use of SM platforms fostered independent citizens' social political activism; and (ii) Bobi wine's use of social media impacted on Uganda's different political parties to use SM platforms to rally citizens' support and participation in social political activities in Uganda. Findings show that citizens' use of social media platforms and affiliation to respective political parties was statistically significant in predicting citizens' civic engagement [Social Activism]. The results of regression indicated predictors explained 6.8% of the variance at $R^2=.068$, $F(3,1147)=28.948$, $p=.001$. And statistically significant in predicting citizens' civic engagement [political Activism]; the results of regression indicated predictors explained 7.1% of the variance at $R^2=.071$, $F(3,1147)=30.384$, $p=.001$. Therefore, based on findings, we conclude that Bobi Wine's use of social media platforms was statistically significant in impacting on citizens' social political activism [civic engagement].

The "Echo Chamber" and Its Impact on Political Knowledge

John Huffman, *Growth Focused Insights and Research*

This paper will analyze the extent to which people consume media of similar types and how that correlates with political knowledge. An analysis of two surveys (n=2,000) conducted among US adults 18+ in Oct/Nov 2019 and Oct/Nov 2020 will be used to measure what media are consumed or not consumed, then assess political knowledge by asking factual questions. We will then analyze these findings using other variables such as political party affiliation, social media use, Presidential preference, attitudes towards the Impeachment of President Trump and standard demographic variables. This research will also create 'echo chamber types', for ease of analysis and as a nice heuristic to measure impact on other phenomenon such as attitudes towards COVID and how the media have reported on COVID. This paper will also discuss the potential impact of these 'echo chambers' for explaining the deep partisan divisions among the population.

Effectiveness of Russian Propaganda/Disinformation in the Near Abroad: The Case of Belarusian Presidential Election-2020

Natalie Rice, *University of Tennessee*

Oleg Manaev, *University of Tennessee*

Catherine Luther, *University of Tennessee*

Suzie Allard, *University of Tennessee*

Michael R Fitzgerald, *University of Tennessee*

In August 2020, the Belarusian presidential election and subsequent mass protests evoked vast international interest. We examined how Russian and Belarusian news media outlets and public officials framed the actions of the leadership of both countries in connection with these events, and explored whether the types of frames used appeared to be designed to affect public opinion. We conducted a qualitative framing analysis of news stories that were aired or published between May and September 2020 in major Russian and Belarusian state-run news media outlets.

We identified unique narratives in both Russian and Belarusian media news stories published before the election and after the election. The results of our analysis revealed a complete change of tone in news coverage from both countries pre and post-election. While the pre-election coverage in both Belarusian and Russian sources was mostly negative towards the other country and its leadership, after the August election Russian media sources began expressing open support for re-elected President Lukashenko, while Belarusian media outlets praised the close positive relationship between two countries.

Our analysis included use of the News Coverage Classification System which has been developed by our team. The tool utilizes two axes to gauge two key concepts in news coverage – the character of the information (evidence-based or false) and the character of the reporting (objective or biased). This system classifies narrative frames and provides evidence to classify narratives as examples of journalism, misinformation, propaganda or disinformation.

Comparing results of framing analysis with the results of public opinion surveys conducted in both countries allowed us to interpret our results in a broader framework of the effectiveness of Russian propaganda/disinformation in the broader political and cultural context of both countries.

Public Attention to Information in the 2020 U.S. Presidential Election Campaign

Josh Pasek, *University of Michigan*

Lisa Singh, *Georgetown University*

Stuart Soroka, *University of Michigan*

Jennifer Agiesta, *CNN*

Michael W. Traugott, *University of Michigan*

Grace Sparks, *CNN*

Ceren Budak, *University of Michigan*

Jonathan M. Ladd, *Georgetown University*

The 2020 U.S. Presidential Election Campaign was unique in a variety of aspects. Americans were making their choices amidst the backdrop of a global coronavirus pandemic, historic racial protests, and a host of inaccurate claims about the major party candidates. To understand what information reached ordinary Americans about this election, a team from CNN, Georgetown, and the University of Michigan conducted a nationally representative RDD telephone study for 12 weeks asking Americans an open-ended question about what they had seen, read, or heard about the major party candidates in the last few days. We used textual analysis tools to examine the words Americans associated with each candidate and topic modeling to understand how the things Americans mentioned about the candidates and the campaign shifted over time. The results of this research echoed a theme of the campaign wherein the things that were salient about Joe Biden were fairly consistent from week to week whereas there was far more variability in the terms and topics used in conjunction with Donald Trump. Despite this variability, however, recalled information about Donald Trump overwhelmingly focused on a single issue throughout the election cycle — a large portion of Americans were focused on his handling of the coronavirus pandemic. In this paper, we illustrate both the stability and variability in Americans' recalled information about the 2020 candidates and provide an assessment of how the use of various terms corresponded with sentiment toward each candidate.

Testing Response Formats

A Methodological Comparison of the Use of MaxDiff Tasks to Rating Scales in Survey Research

Christina Tworek, *HarrisX*

Anna Gurevich, *HarrisX*

Marina Taraukhina, *HarrisX*

Dritan Nesho, *HarrisX*

Rating scales are widely used for measuring importance attribution across academic and market research. Unfortunately, it is often difficult to interpret results from rating questions due to issues like satisficing, scale-meaning bias, and lack of discrimination between items. Such issues are especially problematic in online surveys and cross-cultural research, which have become increasingly popular. Maximum Difference Scaling (MaxDiff) is a trade-off analysis technique that uses multiple pairwise comparisons to measure relative differences between survey items. In a MaxDiff task, respondents are asked to indicate the most and least important items from a list, resulting in a relative ranking for each item. This technique provides strong discriminatory power, eliminates scaling bias, and reduces respondent fatigue. Despite the strengths of MaxDiff, rating scales continue to dominate survey research. Here, we provide a methodological comparison of rating scales to MaxDiff tasks. In two studies, we compared MaxDiff to rating scales in their ability to measure people's attitudes towards online data privacy and security. In both studies, respondents reported the importance of keeping different types of data (e.g., medical records, emails) private and secure. In Study 1 (N = 8,000), respondents rated the importance of each data type on a 10-point Likert scale (1 = Not at all Important, 10 = Extremely Important). To reduce fatigue, each respondent rated one-third of the items. In Study 2 (N = 5,018), participants evaluated the same data types in a MaxDiff task, but all items were shown. Our MaxDiff task resulted in better data integrity (i.e., relative rankings on all items for each respondent) and better discrimination between items. Our study is the first to directly compare traditional rating scales to MaxDiff using the same sample population and questionnaire items. We discuss various applications of MaxDiff and implications for market research.

How Does the Scale Affect Subjective Health Ratings? Insights From Web Probing

Katharina Meitinger, *Utrecht University*

Cornelia Neuert, *GESIS Leibniz Institute for the Social Sciences*

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

The self-rated health item is a crucial and frequently used question that is implemented in most large-scale surveys. However, the large scale surveys use different scales for this item. For example, the European Social Survey (ESS) uses a balanced 5-point scale running from "very good" to "very bad" with "fair" as mid-point of the scale whereas the GESIS Panel opted for "moderately" as mid-point. The Survey of Health, Ageing and Retirement in Europe (SHARE) and the International Survey Program (ISSP) use an unbalanced scale in which positive categories predominate and with "excellent" and "bad" as end points. The question remains how the different response scales influence the answer behavior of respondents and whether different associations are triggered for each scale point of each respective scale. This presentation discusses findings of a split split-ballot experiment that was implemented in a web survey with a sample of 1,600 German panelists of an online access panel. The study was fielded in July 2019. The experiment compared three different scale versions of the self-rated health question. Each scale version was followed by a closed-ended probe that asked respondents which factors they considered when deciding on an answer category. The aim of the closed-ended probe was to assess whether associations differed by scale point. The closed-ended probes were developed based on insights from previously conducted cognitive interviews in the U.S. and in Germany. In addition to a content perspective, we will assess whether the different scale versions affect response quality (e.g., item non-response) and response behavior (e.g., response time). Findings and implications are discussed.

An Eye-Tracking Study of Scale Direction Effect

Ting Yan, *Westat*

Scale direction effect refers to the impact of the direction in which a scale is presented to respondents on the resultant answers. A response scale can start from a positive to a negative end. It can also begin with a negative end progressing to the positive end. Holding other scale features constant, scale direction is found to affect response distributions by yielding more selections of scale points closer to the beginning of the scale. Although this phenomenon has been empirically demonstrated in different modes of data collection for respondents with different demographic characteristics, what remains understudied is the mechanism underlying the scale direction effect. In addition, it is not clear which scale direction is easier for respondents to use. To address these questions, I employed eye-tracking because it provides a direct window into how respondents process survey questions. In an eye-tracking study, respondents' eye movement is tracked while they are reading and answering survey questions. In this paper, I will examine two eye-tracking measures. Fixation counts and fixation duration are used to study respondents' attention to and cognitive processing of response scales. Pupil dilations are used to understand the extent of cognitive difficulty of using response scales. I will compare fixation counts, fixation duration, and pupil dilations by scale direction and by respondents' actual answers. Findings of this paper will shed light on how respondents process response scales and will have practical implications for questionnaire design.

An Even or Odd Number of Response Category: Investigating the Effect of Middle Alternative in GSS Abortion Attitude Measures

Wen-Juo Lo, *University of Arkansas*

Ronna C. Turner, *University of Arkansas*

Brandon L. Crawford, *Indiana University*

Kristen N. Jozkowski, *Indiana University*

The Likert scale has been applied to many questionnaires and surveys after Likert introduced his 5-point summative rating scale to measure attitude in 1932. However, the question remains regarding the middle response category endorsement in attitude measures. Krosnick and Presser (2010) summarized the potential satisficing effect when a survey included the middle point response category (i.e., an odd number of response format). They also pointed out offering a midpoint may be useful in increasing reliability and validity if respondents' attitude toward the question is neutral or unclear. Therefore, the purpose of this study is to evaluate and compare four different number response formats (i.e., 4, 5, 6, and 7 categories) in terms of latent classes and constructs using General Social Survey (GSS) abortion items. A total of 1,338 participants were randomly assigned to one of four sets of questions. These sets only differed in the number of response categories and this study only analyzed 6 GSS items related to abortion circumstances. The factor mixture model (FMM) that combined latent class and factor analysis (Muthen, 2008) was used. In addition, we applied both Bayesian model parameter estimations and the maximum likelihood estimation with standard errors and mean-adjusted chi-square test which we can compare results from different estimators. The findings regarding the differences in terms of the number of latent classes revealed and similarities of latent constructs are discussed.

What's Your Number?: Numeric Response Format in Measuring Political Attitudes

Jasmine Stephens, *Ipsos Public Affairs*

Randall K. Thomas, *Ipsos Public Affairs*

Megan Hendrich, *Ipsos Public Affairs*

Fully-anchored, semantically labeled scales have somewhat higher levels of validity than end-anchored scales (Krosnick, 1999). However, most smartphone respondents take online surveys in the portrait orientation, which means the horizontal presentation of five semantically labeled categories can extend off the screen. We developed an alternative for smartphone screens—replacing all semantic labels with numeric labels as clickable buttons to anchor the responses, with plus or minus indicators (e.g., How much do you like doing X? -2 -1 0 +1 +2). While many researchers provide semantic end labels for respondents (e.g., "Strongly agree" or "Strongly disagree"), we presented only numeric values. In our previous research, we found that fully numeric response formats functioned similarly to semantic formats, with equivalent means, standard deviations, and validity, while taking less time to complete. In addition,

we found that unipolar formats for both fully-labeled semantic and numeric formats had higher validity across a number of topics when compared to end-labeled or bipolar formats. While unipolar formats generally had higher validity across most topics, we have found one exception—bipolar formats may have higher validity for political issues. In this study, 2,347 respondents from the probability-based KnowledgePanel were randomly assigned one of three formats to evaluate a series of political issues, all bipolar and fully labeled: semantic, numeric, and numeric with a respondent instruction (“Use ‘-2’ if you strongly disagree to ‘+2’ if you strongly agree.”) to see if the respondent clarification improved numeric scale functioning. We found that numeric formats were again completed more quickly than semantic scales, and they showed equivalent means, standard deviations, and validity. Adding the respondent instruction did not improve item functioning. We discuss the future applications and some limitations we encountered in our exploration of the numeric format.

Doomed to Failure: Forgetting the Past in Item and Instrument Development

Todd Rockwood, *University of Minnesota*

Charles Cain, *University of Minnesota*

Sarah Putnam, *University of Minnesota*

Kyle Rudser, *University of Minnesota*

Melissa Constantine, *University of Minnesota*

Leslie Ricky, *Yale University*

Emily Lukacz, *University of San Diego*

Linda Brubaker, *Loyola*

Janice Miller, *University of Michigan*

This paper will present fundamental comparisons to illustrate the dangers associated with a two trends in survey item and instrument development. First, is a sustained effort to rely on a few or even a single item for measurement. Second, is the dependence on what is often and incorrectly labelled a Likert scale; which is a fundamental confusion of the difference between a Likert scale and a form of bi-polar response options. To illustrate these trends two different illustrative examples one drawing on the principles of Thurstone and another on Guttman measurement models will be presented. The Thurstone example relies on publicly available polling data from (pollingreport.com) regarding abortion. In the first example data will illustrate the progression from a single item with dichotomous response options (52% choice, 43% life), a single item with ‘Likert’ response options (curve with with short tails) and finally four questions illustrating a Thurstonian approach (near normal curve). The second comparison (epanel ballot experiment n=654) 2 illustrates a comparison between a single item built on the presumed equivalence of a Likert response scale and a Likert scale as opposed to an item built on Guttman scaling. The Likert item demonstrates a distribution consistent with a single item that represents a partial picture of global distribution as opposed to the item based in Guttman scaling with a global rating. This paper will provide a detailed rationale based on the underlying measurement theories employed, why the distributions are consistent with the principles of each theory and the risks associated with failing to

Origin and Identity: Analyzing a New 2020 Census Response Option

Anna Brown, *Pew Research Center*

D'Vera Cohn, *Pew Research Center*

Respondents who self-identified as White or Black in the 2020 Census were asked to write in more detail about their origins, the first time this option was included on the decennial count. (Other racial groups and Hispanics already could do so.) The Census Bureau asserts that this new feature responds to the changing and complex ways in which people view their self-identity, and will allow publication of more detailed data for specific population groups. How likely are people who provide a detailed origin to describe that origin as salient to them? This paper analyzes origin responses from a survey of about 3,500 U.S. adults who were asked to answer questions identical in wording to the 2020 Census race and ethnicity questions and to evaluate how well those questions describe them. We also asked respondents about their knowledge of their origins, their origins’ centrality to their identity and the strength of their connection to their origins, enabling us to test saliency of origin for different respondent groups.

Open-Ended vs. Closed: Evaluating Different Probing Formats in Web Probing

Cornelia Neuert, *GESIS - Leibniz Institute for the Social Sciences*

Katharina Meitinger, *Utrecht University*

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

The method of web probing integrates cognitive interviewing techniques into online surveys and is increasingly used to evaluate questions by collecting data on respondents' answer processes. Typically, web probes are administered directly after the question to be tested as open-ended questions with text fields. While the use of open-ended probes in web surveys has shown promising insights in discovering problematic survey items and inequivalence across countries, it is generally acknowledged that open-ended questions are more burdensome to answer for respondents and face more item-nonresponse.

A second possibility of administering probing questions is in a closed format. The implementation of closed probes drastically reduces the costs and burden involved in the data processing and analysis stages compared to open-ended questions, because it omits the development of coding schemes, the coding of the responses and reduces the response burden of answering. However, up to now, the open-ended and the closed web probing approach have not yet been systematically compared.

In this paper, we address the following two research questions: 1) Are open-ended probes and closed probes comparable with regard to the substantive themes provided? 2) Are open-ended probes and closed probes comparable with regard to data quality?

In particular, we compare the amount of insights in terms of validity (variety and prevalence of themes) as well as indicators of response quality (e.g. item nonresponse, nonresponse bias, response times).

Based on a sample of 1,600 German panelists of an online access panel, we conducted a web experiment comparing the responses of closed vs open-ended probing questions on three questions under consideration. The study was fielded in July 2019. Findings and implementations are discussed.

Survey Weighting and Estimation Strategies

Adding Measures of Political Engagement and Public Trust to the Weighting Toolbox- Is It Worth It?

Parker Quinn, *Polco*

Nick Mastronardi, *Polco*

Erin Caldwell, *Polco*

A variety of weighting variables have been used to correct the results of both probability and non-probability samples but the variables making up this secret sauce are not generally agreed upon or known. While most rely on a set of core demographic variables, we know calibrations by demographics alone often fall short. In an effort to understand the key variables for survey weighting, National Research Center at Polco examined its National Community Survey™ database of more than 100 matched sets of probability and non-probability community livability surveys to assess the contribution of adding measures of political engagement and public trust into the weighting models. The National Community Survey™ (The NCS) is a templated, benchmark used by local governments across the nation to measure quality of life, community livability, local government performance and civic participation. Probability-based surveys have been aggregated in this database since 1997 and non-probability surveys (a second survey completed alongside the probability surveys) since 2013. The database includes the data for more than 850 communities (probability samples) and 100+ matched pairs of probability and non-probability samples. The calibrated ratings from the surveys were compared at the national and local level to identify the importance of political engagement and public trust in sample correction.

Rethinking the Traditional Survey Weighting Paradigm - Are Design Weights Necessary?

Mansour Fahimi, *Marketing Systems Group*

Meng Li, *Ipsos Public Affairs*

Elahe Tashakor, *Ipsos Public Affairs*

For decades, research statisticians have relied on a fairly standard protocol for weighting survey data using a three-step process. In the first step, design weights are computed to reflect selection probabilities imposed by the study design. In the second step, although often skipped, design weights are adjusted for differential nonresponse patterns detected by comparing respondents and nonrespondents based on the available frame data. In the final step, nonresponse-adjusted weights are raked to a series of population benchmarks to compensate for biases that may result from lingering nonresponse or sampling undercoverage. The process of computing design weights is a nontrivial exercise that is becoming exceedingly inexact and often dependent upon ad-hoc assumptions and extrapolations. Working with the BRFSS survey data, we have examined the relevance of design weights when producing weighted estimates. Specifically, we have replicated the BRFSS weighting methodology while starting with various power transformations of the design weights – ranging from power of one to zero. That is, using a progression of design weights that starts with what BRFSS is currently using (power transformation of one) to an extreme scenario when all design weights are set to one (power transformation of zero). The resulting final weights under each scenario were then used to produce and compare estimates for a long list of key outcome measures. Our research shows that the impact of design weights on survey estimates is often negligible, yet they tend to increase the variability of the weights and hence reduce precision. This is particularly true as escalating nonresponse rates demand exceedingly more aggressive raking adjustments, completely overpowering any influence design weights may have. This research provides empirical evidence that the application of traditional weighting methodologies that incorporate design weights into the weighting process is an expensive investment that can pay little, if not negative, dividends.

The Importance of Telephone Usage as a Calibration Variable in Dual-Frame RDD Samples

Luis Tipan, *SSRS*

Jonathan Best, *SSRS*

Survey researchers regularly employ various weighting techniques to adjust for differential non-response and reduce bias. One common weighting method is the calibration of sample demographic distributions to population benchmarks through post-stratification or raking. Typically, these calibration variables include demographics such as sex, age, education, race and region. One calibration variable commonly used in dual-frame RDD landline and cell surveys is telephone usage. Specifically, the proportion of sample respondents that are cell phone-only, landline-only and dual-users are matched to benchmarks derived from the National Health Interview Survey Early Release Reports on Wireless Substitution. AAPOR Virtual Conference (Best & Tipan, 2020), assesses whether phone use is necessary as a calibration variable given that other variables typically included in calibration are highly correlated with phone use. We aim to answer two questions. First, does excluding a phone use calibration variable introduce bias in survey estimates? And second, does excluding the phone use calibration variable increase the sample's statistical power by lowering the unequal weighting effect. Using data from the SSRS Omnibus, we will run weightings both including and excluding phone use calibration, assessing potential bias by comparing results to question for which reliable benchmarks are available, including measures from BRFSS on smoking prevalence, high blood pressure and asthma.

Developing and Assessing Weighting Methods for the Redesigned National Health Interview Survey

James Dahlhamer, *National Center for Health Statistics*

Yangyang Deng, *ICF*

Ronaldo Iachan, *ICF*

Stephen Blumberg, *National Center for Health Statistics*

Matthew Bramlett, *National Center for Health Statistics*

Aaron Maitland, *National Center for Health Statistics*

Jonaki Bose, *National Center for Health Statistics*

Matt Jans, *ICF*

Adam Lee, Robynne Locke, *ICF*

The 2019 redesign of the National Health Interview Survey (NHIS) questionnaire provided the opportunity to re-evaluate long-running survey processes, such as nonresponse adjustments and other weighting steps. This paper details work done between the National Center for Health Statistics (NCHS) and ICF to evaluate different approaches for adjusting for nonresponse bias in the development of NHIS sample weights. Nonresponse models built with rich auxiliary data (e.g., interviewer observations of sample units, tract-level sociodemographics from the Census Planning Database) were estimated using machine learning methods (e.g., lasso and random forest), as well as more traditional logistic regression (single- and multi-level), to find best-fitting models predicting response propensity. With models in place, the impact of the following permutations on key NHIS health estimates was assessed: 1) the imposition of caps of various limits on the nonresponse adjustment factors to limit variance increases, and 2) different raking routines to extend the simple demographic post-stratification previously used with the NHIS. In addition to age, sex, and race/ethnicity, additional raking variables such as education, employment status, urbanicity, and location of residence (Census division) were explored. The resulting weights were then evaluated in terms of reductions in estimated nonresponse bias, variances for key health indicators, and mean squared error (MSE).

Assessing and Implementing Weighting Methods for the Redesigned National Health Interview Survey

Yangyang Deng, *ICF*

James Dahlhamer, *National Center for Health Statistics*

Ronaldo Iachan, *ICF*

Stephen Blumberg, *National Center for Health Statistics*

Matthew Bramlett, *National Center for Health Statistics*

Aaron Maitland, *National Center for Health Statistics*

Jonaki Bose, *National Center for Health Statistics*

Matt Jans, *ICF*

Adam Lee, *ICF*

The 2019 redesign of the National Health Interview Survey (NHIS) provided the opportunity to re-evaluate long-running survey processes, such as nonresponse adjustments and other weighting steps. The revised weighting methods and the evaluation of nonresponse bias considered the new survey design, the availability of rich auxiliary data, and the application of new modeling techniques. This paper is a sequel to another paper which evaluated non-response adjustment weighting methods for use with the redesigned NHIS, which was part of the work done by the National Center for Health Statistics (NCHS) and ICF to refine the NHIS nonresponse bias adjustments and the weighting process in general. We present results of an evaluation of different weighting methods which were considered for the 2019 NHIS including nonresponse bias adjustments at the household, adult, and child levels. Both machine learning methods and multilevel logistic regression models were evaluated in developing weight adjustment classes based on propensity quintiles. The paper then describes the process of retrofitting the weighting processes for the 2016, 2017 and 2018 NHIS using the methods newly developed for the 2019 NHIS. We compare the results of the new methods side by side with the previously published NHIS estimates and variances.

The Cracking of Monoliths: Improving Representation of Groups Through Subgroups

Seth Messinger, *Ipsos*

Randall K. Thomas, *Ipsos Public Affairs*

Omar Pedraza, *Ipsos Public Affairs*

Frances M. Barlas, *Ipsos Public Affairs*

Regarding demographics, every sample we obtain in research has some amount of divergence from the population demographics. Many of these demographic differences can be corrected through weighting. Often, for racial or ethnic groups, the entire group is treated as one without regard to subgroups (e.g. Hispanics overall rather than Cuban Hispanic, Mexican Hispanic, Puerto Rican Hispanic, or other Hispanic). We examined two separate studies comparing probability-based KnowledgePanel with opt-in non-probability sample in terms of bias regarding overall samples and by specific race-ethnic groups. We found that Hispanics had higher rates of error for both sample types when measured as a function of divergence from benchmarks established by federal benchmarks. We sought to explore how the use of weighting targets for specific subgroups might be able to reduce error for these groups. For Hispanics, we focused on using weighting that incorporated Hispanic subgroup (Cuban, Mexican, Puerto Rican, Other) as well as using age groupings, or age-gender groupings (targets were established for these subgroups within Hispanics). For Black respondents, we looked at regional distribution, also with age groupings and age-gender groupings. We compared the use of nested weighting targets for subgroups on the effects of bias for the race-ethnicity group and quantify each improvement and develop recommendations that can improve the representativeness of each group. We also discuss the limitations of such an approach, which include increasing weighting variance and becoming increasingly impractical with sub-groups that are relatively small.

Measuring COVID-19 Antibody Prevalence in Canada

Ilya Hekimi, *Statistics Canada*

There are many studies that look at COVID-19 prevalence in their populations, but they are often not based on probability samples. Statistics Canada has put into place a new survey to estimate antibody prevalence in the Canadian population, the Canadian COVID-19 Antibody and Health Survey. It comprises of a short questionnaire and collection of dried blood spots to be analyzed for the presence of COVID-19 antibodies. The target population for this survey are persons living in Canada aged 1 and up. This survey presents many challenges, both operationally and from an estimation standpoint. Both point and variance estimation need to integrate the false-positive and false-negative rates of the blood test. Additionally, there are opportunities to integrate non-probability data from blood donation organizations to our probabilistic sample.

5/14/2021, 12:00 PM - 1:30 PM
Concurrent Session H: Live Sessions

Mini-Conference: Results of the AAPOR/WAPOR Task Force on Quality in Comparative Surveys

Results of the AAPOR/WAPOR Task Force on Quality in Comparative Surveys

Lars Lyberg, *Demoskop (in memoriam)*
Timothy Johnson, *University of Illinois, Chicago*
Michael Robbins, *Princeton University*
Elizabeth Zechmeister, *Vanderbilt University, LAPOP*
Mandy Sha, *www.mandysha.com*
Zeina Mneimneh, *University of Michigan*
Kristen Cibelli Hibben, *National Center for Health Statistics*
Julie A. de Jong, *University of Michigan*

Multinational, multiregional, and multicultural (3MC) surveys, also called comparative surveys, are deliberately designed with the goal of achieving comparability across more than one population. In recent decades, 3MC surveys have increased substantially in number and geographic spread. As most large societies have cultural and linguistic minorities, many of which continue to grow in their diversity, 3MC issues are also increasingly relevant to single-country multicultural and multiregional survey research. The potential impact on decision-making and knowledge of the data collected in 3MC surveys is perhaps more significant than ever. However, this growth does not come without challenges. 3MC surveys are much more complicated than most single-population surveys and the problems associated with their planning and implementation tend to overshadow quality management and quality assessment activities. Furthermore, comparative surveys are conducted in different disciplines with different research traditions, hindering opportunities for advancement in improvements to data quality. In this context, a Task Force on Quality in Comparative Surveys was approved by both AAPOR and WAPOR in 2018 to promote dialogue and collaboration among people from different organizations involved in 3MC survey research with the goal of identifying the most pressing problems concerning data quality and recommending priorities for future research and collaboration. This panel covers essential background on 3MC survey research including conceptual challenges of defining and assessing comparability, specific error sources, and prevailing challenges associated with these surveys and reviews the process of the Task Force. Most important, the panel will present the overall Task Force findings and its recommendations, among which is the call for 3MC survey research to be established as a discipline of its own. Ample time will be allotted for discussion with and to ask questions of the Task Force members in attendance.

Mini-Conference: Mode Changes Involving Web Collection During the Pandemic

Who Will Participate During a Pandemic? The 2020 General Social Survey (GSS) Mode Shift and Resulting Participation Outcomes

Jodie Smylie, *NORC at the University of Chicago*

Rene Bautista, *NORC at the University of Chicago*

Benjamin Schapiro, *NORC at the University of Chicago*

Jaesok Son, *NORC at the University of Chicago*

Deanna Christianson, *NORC at the University of Chicago*

The General Social Survey (GSS) is a nationally representative survey conducted face-to-face every two years with the objective of collecting data on the attitudes and opinions of the general public of the United States. In 2020, the GSS initiated a series of major revisions to mode, methodology, and operations upon recognition of the obstacles presented by the COVID-19 pandemic to conducting face-to-face interviews safely on a national scale. The GSS was adapted from a face-to-face CAPI to two separate studies, a panel and a cross-section, which both employed a mixed-mode web and telephone survey. The GSS 2020 panel study sought to re-interview respondents who had completed the GSS in 2016 and 2018 and provides a suitable environment for exploring participation patterns and mode shift preferences for respondents who had first participated in the study via face-to-face interview. We will provide an overview of the methodological and operational adaptations made to shift the GSS survey mode from face-to-face to mixed-mode web and telephone, such as the adaptation of the questionnaire to web self-administration as well as the redesign of respondent materials and mailings to be based fully on mail, email, and phone outreach. We will then discuss participation outcomes from this mode shift. We will explore the characteristics of respondents who completed via web without prompting, respondents who completed via web with prompting, and respondents who completed via phone interview. We will use data points collected in 2020 and from previous round interviews, including age, race and ethnicity, educational attainment, income, and Internet access. This presentation will provide insight on participation patterns by the characteristics of respondents most likely to complete a study that has shifted mode from face-to-face administration, and may inform future interventions for appealing to respondents less likely to shift to the web or phone mode from face-to-face.

Response Mode in a Pandemic: Shift to Web in the 2020 National Survey of Children's Health

Scott Albrecht, *U.S. Census Bureau*

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

Jessica Jones, *Maternal and Child Health Bureau, Health Resources and Services Administration*

Carolyn Pickering, *U.S. Census Bureau*

Leah Meyer, *U.S. Census Bureau*

The National Survey of Children's Health (NSCH) is an address-based survey of households with children. Sampled addresses are mailed invitations to participate in the survey. All invitations include instructions to complete the survey using the web-based instrument. Select initial and first-follow-up mailings, and all second and third follow-up mailings, include a paper instrument with a postage-paid return envelope. Between 2016 and 2019, two-thirds of all responding households ($n = 330,000$) chose to respond via web. In 2020, 77% of respondents used the web-based instrument. Beginning in March 2020, many Americans moved to full-time remote work and distance learning as they complied with stay-at-home orders. We theorize that these arrangements encouraged a shift to web response as 1) some households added personal computers and internet access and 2) respondents are spent more time at home on their computers. We hypothesize that the shift to web was greater for households with children and working-age adults in telework-ready industries, and that there was an increase in Windows PC and Chromebook use. We use multinomial logistic regression on a pooled sample of addresses to evaluate the association between respondent characteristics and response mode choice in 2019 and 2020. Additionally, we explore options to disentangle the effect of the pandemic from other notable developments in 2020, including the Census "decennial effect". Finally, we discuss the long-term implications of the pandemic-induced shift to web. We anticipate that web response will regress as we return to shared public spaces. But the new normal may include more telework flexibilities and expanded internet access; we argue that web response will not regress to pre-pandemic levels. The U.S. Census

Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. CBDRB-FY20- POP001-0025.

Data Collection Changes Due to COVID-19 in a National Survey of Businesses: A Look at the Medical Expenditure Panel Survey - Insurance Component (MEPS-IC)

David Kashihara, *Agency for Healthcare Research & Quality*

Federal agencies have considerable infrastructure in place to conduct large-scale surveys. These surveys require months of preparation weeks of staff training to be ready to collect, process, and publish the data. In mid-March 2020, when the Federal Government declared a national emergency because of the coronavirus pandemic, the Medical Expenditure Panel Survey - Insurance Component (MEPS-IC) was in final preparations to field the survey. Data collection was expected to begin during the Spring and continue through the Fall, with follow-up activities continuing into the Winter. However, the onset of the health pandemic necessitated immediate major changes to the data collection activities. The MEPS-IC is an annual survey of over 42,000 U.S. business establishments and more than 3,500 governments at the state and local levels. The survey produces national and state-level estimates on topics including the percentage of employers offering health insurance, the number of enrolled employees, and average health insurance premiums, copays and deductibles. Before the pandemic, MEPS-IC participants were able to respond via the web, mail, or telephone. However, with the onset of the pandemic, challenges were presented as businesses closed and key survey operations shut down. This presentation describes the data collection of the 2020 MEPS-IC during the pandemic and explains the rationale behind the decisions to alter or not change certain processes. Response rates will be shown for each data collection operation and will be compared to prior years by select business characteristics such as industry and size. In addition, preparations for possible future survey interruptions will be discussed. The presentation will expand the limited field of knowledge about how business surveys pivoted on short notice to obstacles presented by the health pandemic and will provide insight to the decisions that shaped the data collection process along with the level of success achieved at each step.

Surveying by Mail During a Pandemic: Adaptation of Data Collection Due to COVID-19

Jessica Williams, *RTI International*

Brenna Muldavin, *RTI International*

Joli Brown, *RTI International*

Amie Lynch, *RTI International*

Carrie Darden, *RTI International*

Todd Prince, *RTI International*

Laura DiGrande, *RTI International*

The World Trade Center Health Registry is a NIOSH-funded longitudinal cohort designed to track the health effects of the terrorist attacks of September 11, 2001. Data on exposures and health status are collected by the NYC Department of Health and Mental Hygiene from registrants via a multi-mode survey that includes mail (TeleForm) and web. Four waves were conducted between 2003 and 2016. As the data collection agent, RTI International developed well-tested protocols for data capture and quality control. These protocols were expected to be applied when Wave 5 was launched in April 2020 to approximately 40,000 registrants. The purpose of this presentation is to share how the coronavirus (COVID-19) pandemic affected Wave 5 data collection, which ended in February 2021. Data will be presented by mode on the submitted surveys and data processing over the collection period. COVID-19 did not influence the web survey launch or processing, but the paper survey required reconsideration of previously tested methods. A web-only survey was not a viable alternative since a large portion of registrants requested paper surveys and/or did not supply email addresses. Standard operations at RTI International's Research Operations Center were therefore adapted to handle the large volume of returned paper surveys during COVID-19 work restrictions set by North Carolina's state government as well as RTI's Infectious Disease Prevention Team. WE will share these mitigation strategies and demonstrate how data processing continued. We will share the quality control measures put in place to maintain superior data quality comparable to past waves. The discussion will include operational and logistic challenges, identify trends in data error types and overall data quality, and include presentation of the financial impact of the mitigation strategies.

Transitioning From Mail to Electronic Contact: Acquiring Email From Establishments During the COVID-19 Pandemic

Joshua Langeland, *U.S. Bureau of Labor Statistics*

Robin Kaplan, *U.S. Bureau of Labor Statistics*

As federal agencies moved to remote operations during the COVID-19 pandemic, new methodologies had to be quickly developed and adopted for survey operations. The Bureau of Labor Statistics Current Employment Statistics (CES) program collects data from establishments to produce detailed estimates of employment, hours, and earnings of workers, and is used to produce the monthly jobs report. CES faced the challenge of no longer having the ability to print and use postal mail to send survey materials. Under normal operations, a survey packet is sent via postal mail to respondents prior to an interviewer attempting telephone contact as a way to increase legitimacy and respondents' awareness of the survey. The packet contains an introduction to the survey, a survey form, and in some cases, a copy of the most recent jobs report to explain the importance of the survey and increase motivation to respond. Without the ability to send the survey packet, interviewers now have to rely on collecting Email addresses from respondents in order to send them electronic forms. This presentation will summarize the results of an exploratory analysis to assess the effectiveness of having interviewers collect Email addresses from establishment respondents by phone during enrollment. The analysis will examine various establishment characteristics such as size of the business and industry to determine if certain subgroups have higher propensities to use electronic communications. The presentation will also include excerpts from a focus group conducted with CES interviewers to get their feedback on what techniques have worked well for successfully collecting Email addresses, and what challenges have arisen due to the lack of ability to send paper forms. Implications and recommendations for establishment data collection when transitioning to remote environments and electronic survey operations will be discussed.

Mini-Conference: Response Rates During a Pandemic: Operational Issues, Panel Surveys

The Impact of COVID-19 on Large-Scale Phone Survey Productivity and Response Rates

Matt Jans, *ICF*
Jamie Dayton, *ICF*
Randy ZuWallack, *ICF*
Don Allen, *ICF*
Josh Duell, *ICF*
Andy Dyer, *ICF*
Thomas Brassell, *ICF*
Sam Collins, *ICF*
Traci Creller, *ICF*

COVID-19 has had a complex impact on survey productivity and response rates. Anecdotal reports from a range of organizations and surveys suggest that both productivity and final response rates are up. At least two things may be responsible for this. First, more people are at home. Due to stay-at-home orders and lay-offs, more people are at home full-time than ever in recent history, making households easier to contact and sampled respondents (potentially) more available to be interviewed. Second, more interviewers are making calls from home. Individual interviewer productivity may improve when calling from home instead of from a centralized call center. Certainly, shift staffing, call-back scheduling, and other logistics are much simpler with home-based interviewing. This presentation compares outcome rates (e.g., contact, refusal, cooperation and response rates) from several ICF phone surveys conducted before and during COVID-19. Surveys were selected to minimize differences in target populations, topics, and calling protocols. The presentation addresses the following three questions about the effect of COVID-19 on phone surveys: 1) how have survey outcome rates changed since the start of COVID compared to the months prior to COVID, and, where available, the same months the prior year, 2) For interviewers who worked both from centralized call centers and from home, is their personal productivity similar or different in both contexts, and 3) depending on the return to centralized call centers, can we observe further or maintained productivity improvements, or performance regressions among interviewers who return to centralized calling? Findings are discussed in the context of COVID-19's impact on data collection, interviewer staffing and productivity, and the general difficulty disentangling interviewer effects from effects of the broader data collection context.

Conducting a Recurring National Health Survey During the COVID-19 Pandemic: Response Patterns and Lessons Learned

Debra Wright, *Insight Policy Research*
Meg Trucano, *Insight Policy Research*
Kelsey Walter, *IMPAQ International*
Katherine Allen, *IMPAQ International*
Paul Kallaur, *Center for the Study of Services*
Alok Shrestha, *Center for the Study of Services*
Matthew McCoy, *Precision Opinion*

The Fee-for-Service Consumer Assessment of Healthcare Providers and Systems (FFS CAHPS) survey is an annual survey of 275,000 Medicare beneficiaries enrolled in the Fee-for-Service program and covers all 50 States, the District of Columbia, and Puerto Rico. Conducted using a sequential mixed mode design (mail with telephone follow-up), the survey provides a nationally representative dataset that enables CMS to track beneficiaries' perceptions of the care and services they receive. The COVID-19 pandemic began to spread just prior to the 2020 data collection. In this presentation we describe adaptations made to continue data collection safely and patterns of response to mail and telephone compared to 2019. Contrary to expectations, response to the 2020 FFS CAHPS survey was substantially higher than in 2019. This increase was due to a higher rate of mail returns—7 percentage points higher in 2020 than 2019. Mail return rates surpassed those achieved in 2019 in all States but increased most in States located in the middle and southern regions of the United States. However, we found no overall

correlation between the local COVID-19 infection rate and change in mail response by geographic area. We also examined response across subgroups (including by sex, race, income, Spanish language preference, and disability status) using CMS administrative data. Mail return rates were higher across all subgroups in 2020. The number of completes obtained via the computer-assisted telephone interviewing follow-up was much lower in 2020 compared with 2019. Though fewer calls were made to reach mail nonrespondents, weekly contact rates increased slightly in 2020. We describe lessons learned, including our experience developing contingency plans, shifting interviewers to remote status, and implementing procedures to protect the safety of workers on site.

Impact of COVID-19 on Survey Response Rates in a Longitudinal Study of Young Adults

William Young, *Rutgers Center for Tobacco Studies*

Jonathan Wivagg, *Westat*

Ting Yan, *Westat*

Charles Carusi, *Westat*

Cristine D. Delnevo, *Rutgers Center for Tobacco Studies*

Daniel A. Gundersen, *Dana-Farber Cancer Institute*

The COVID-19 pandemic has affected the behaviors of Americans in many ways and this paper explores the impact of COVID-19 on survey response rates. The data we analyze come from the National Young Adult Health Study sponsored by Rutgers University. The study is a longitudinal panel of adults aged 18-34 (N=831) initially recruited in 2019 through RDD sampling of cell phone numbers nationwide. Overall, the study includes 3 waves conducted approximately annually with each wave including a Main survey and a shorter Follow Up survey. The data collection for the Main study of wave 2 was underway in spring 2020 when Americans saw the appearance of COVID and the initial rapid growth of the COVID infection rate. As COVID infections are continuing to rise as of the writing of this abstract, understanding any trends in survey response that might be attributable to the infection will inform the expectations of upcoming research efforts. To answer this research question, the research team mapped cell phone numbers to the geographic rate center associated with phone numbers of as of April 1, 2020 and also appended COVID infection rates for the rate center area as of this date. Additionally, the team appended state-level data on cell phone mobility from the University of Washington—a measure of the extent to which residents of a particular state heeded advice to stay home—for the same period. Our analysis will explore whether COVID infection rates and cell phone mobility significantly predict survey response. In addition to the overall relationship between infection rates, cell phone mobility, and response, the analysis will determine whether demographic and behavioral characteristics of survey respondents (known from the initial panel recruitment interview) interact with COVID infection rates as predictors of survey response.

Differential Nonresponse in a High Responding COVID-19 Surveillance Survey

Tracy Keirns, *University of New Hampshire*

Judy Rees, *Geisel School of Medicine at Dartmouth College*

Early in the pandemic, the University of New Hampshire Survey Center and the Geisel School of Medicine at Dartmouth began a statewide COVID-19 symptom surveillance survey, that also tracked attitudes and behaviors concerning COVID-19 as the pandemic progressed. A probability sample of 1,105 New Hampshire adults was recruited into a Dartmouth COVID-19 web panel. Panelists were told the study would last no more than eight weeks, however the pandemic and its effects on public opinion and behavior have continued and evolved over the past seven months and the panel has continued to be interviewed. One thousand, one hundred five panel members completed the initial baseline survey online out of 3,273 who were recruited, yielding an initial response rate of 34%. Of those who joined the COVID-19 panel, an average of 72% responded to daily diaries (later converted to weekly) and monthly attitude and behavior surveys. Over the seven months of the study, attrition has remained remarkably low, only reaching 10% after six months of the study. However, as the governmental response to the pandemic became increasingly entangled with partisan electoral politics, we became concerned that panelists were also becoming more partisan in their willingness to respond. This paper examines the attitudinal and demographic characteristics of those opting in and out of the study. Early analysis shows support for differential attrition based on party registration and education. As we move to collect more data on vaccine hesitancy, understanding the patterns of attrition are critical. The impact of partisan differential nonresponse to public opinion research generally, and COVID-19 or public health studies specifically, should be of great concern to the industry.

5/14/2021, 1:30 PM - 3:30 PM
Concurrent Session I: Prerecorded Sessions

Abortion Attitudes: Language, Culture, and Context

Latinx and Non-Latinx Attitudes Toward Abortion Morality and Legality and its Influence in Their Support Toward Women Seeking an Abortion

Alejandra Kaplan, *Indiana University*
Maria Montenegro, *Indiana University*
James Weese, *University of Arkansas*
Ronna C. Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*
Brandon L. Crawford, *Indiana University*
Kristen N. Jozkowski, *Indiana University*

Abortion continues to permeate US socio-political discourse. However, there is little consensus among what is known about abortion attitudes among the Latinx community—the fastest-growing US population, and how their beliefs compare to non-Latinx community. This presentation compares Latinx and non-Latinx people's support for their family and friends' decisions to have an abortion or continue an unwanted pregnancy -- as related to demographic and abortion attitude-related variables. We conducted two sets of hierarchical regression analyses to assess whether support for a friend or family member seeking an abortion or continuing a pregnancy is predicted by participants' attitudes toward abortion morality and legality while controlling for demographic variables (e.g., Latinx identity, religious affiliation, abortion identity). Abortion identity was significant when predicting support for abortion. Those identifying as pro-life were significantly less likely to support women seeking an abortion than participants with other abortion identities. When adding morality and legality attitudes into the model, abortion identity-maintained significance and held the same trend. Participants with higher abortion morality and legality scores (indicating abortion is moral and should be legal) had higher support for women seeking an abortion. Conversely, identifying as pro-life and Christian was associated with higher support for continuing a pregnancy. There were no significant differences between Latin and non-Latinx regarding support for abortion or continuing a pregnancy. Studies suggest that the US Latinx community has more pro-life leaning tendencies than non-Latinx members. Our findings suggest that support for a family member or friend considering abortion or continuing an unwanted pregnancy are similar for Latinx and non-Latinx adults, after controlling for key demographic and attitudinal variables. Thus, although abortion attitude differences may occur at the aggregate level for these groups, differences may be accounted for by factors similar to those used in this study.

Using Propensity Score Matching to Examine Differences in Abortion Attitudes by Language of Survey

Megan Solon, *Indiana University*
Brandon L. Crawford, *Indiana University*
Ronna Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*
Kristen N. Jozkowski, *Indiana University*

Previous research suggests that the language in which a survey is administered can affect survey responses. Among a Latino sample, for instance, Lee and Pérez (2014) found that survey language was significantly related to various political opinions even after controlling for differences in demographic and background variables. When considering multilanguage surveys and multilingual participants, it is important to know if and how participants' responses may have differed had they taken the survey in a different language. Although randomization is the ideal way to assess these potential differences, randomization of survey language is not always possible. Some disciplines have used propensity score analyses (PSA) as a way to assess "treatment effects" in the absence of randomization. We propose the novel approach of utilizing PSA to assess the potential impact of survey language. Specifically, we use PSA to investigate the influence of survey language on US Latinx responses to web-based survey

questions related to abortion attitudes. By viewing survey language as a treatment or intervention, PSA allows researchers to estimate how a participant's answers to a survey may have differed had they completed the survey in a different language (the counterfactual). Additionally, because PSA uses both a selection and an outcome model, PSA allows researchers to use separate sets of covariates related to treatment selection (in this case, survey language) and the outcome of interest (abortion attitudes). Our findings are based on results from two surveys (n=3,000; n=1,500) where approximately 1/3 of the sample completed the survey in Spanish. We will discuss multiple PSA approaches and the benefits of each approach. To conclude, we will highlight how PSA can help estimate the impact of survey language on response as well as help identify problems with translations.

Abortion Attitudes in the Context of the Coronavirus Pandemic

Kristen N. Jozkowski, *Indiana University*

Kathryn LaRoche, *Indiana University*

Ronna Turner, *University of Arkansas*

Brandon L. Crawford, *Indiana University*

Wen-Juo Lo, *University of Arkansas*

Background: At the onset of the coronavirus pandemic, many states placed temporary bans on medical services deemed “nonessential;” several states initially deemed abortion non-essential. This, coupled with the fact that seventeen states ban use of telemedicine for abortion, re-ignited debates about abortion accessibility in the U.S. Although, federal judges eventually blocked orders categorizing abortion as nonessential, such events demonstrate the political saliency of abortion. We examined whether and why people thought abortion should be (1) considered an essential medical service and (2) permitted via telemedicine during the coronavirus pandemic.

Method: Two waves of data were collected from Ipsos' Knowledge Panel (N = 919, Wave 1; N = 711 Wave 2), a probability sample of U.S. adults. Participants were asked whether they thought abortion should be considered an essential medical service and whether telemedicine should be permitted for medication abortion during the pandemic across 11 different circumstances (e.g., life endangerment, financial reasons). Participants were also asked open-ended questions to explain their close-ended responses.

Results: When asked generally, 44.9% indicated abortion should be essential and 46.7% indicated abortion should be available via telemedicine. However, there was greater support (60% to 81%) for abortion as essential for medical-related circumstances (e.g., life endangerment, baby not expected to survive) and less support (54% to 67%) for social circumstances (e.g., financial difficulties, not wanting more children). Participants' reasons for thinking abortion is/is not essential were primarily circumstantial, although some participants mentioned the importance of timing

Discussion: Relying on general items might be misleading regarding abortion attitudes in complex situations—as essential services or in the context of telemedicine--as support varied substantially. Open-ended findings suggest people's perception of abortion as essential may depend on circumstances people consider “essential.” People also recognized “time is of the essence” regarding abortion. Implications for abortion attitude measurement will be discussed.

Triangulating Quantitative and Qualitative Data to Unpack Complexity and Dissonance in Abortion Attitudes

Brandon L. Crawford, *Indiana University*

Kristen N Jozkowski, *Indiana University*

Ronna C. Turner, *University of Arkansas*

Wen-Juo Lo, *University of Arkansas*

Michael Traugott, *University of Michigan*

Abortion is a complex social issue. Thus, assessing abortion attitudes via survey questions can be challenging, particularly given that people's attitudes may be multi-dimensional. How people respond to survey questions may vary based on underlying factors such as if the question frames abortion as a legal

or moral issue, the circumstances surrounding the abortion decision (e.g. pregnancy as a result of rape, parents are not ready), and the gestational age of the pregnancy (e.g., 6 weeks) . Further, the complexity of respondents' abortion attitudes may manifest in seemingly dissonant responses when asking questions that span different contexts (e.g., abortion due to rape; abortion at 6 weeks gestation) and dimensions (e.g., legality, morality). To unpack the complexity and dissonance of survey responses, we triangulated findings from a quantitative pilot survey with findings from qualitative pilot follow-up interviews.

Specifically, we conducted an online survey using GfK's opt-in panel (English N = 1,093; Spanish N = 493) with follow-up in-depth interviews from survey respondents (English N = 16; Spanish N = 8). The survey comprised items assessing multiple dimensions of abortion attitudes; interviews focused on assessing people's salient beliefs regarding abortion as well as points of potential dissonance, allowing participants to explain their complex conceptualizations of abortion. After outlining our process for the mixed-method pilot study, we will discuss how themes from participants' in-depth interviews map onto their survey responses, providing context for potentially complex and dissonant survey responses. We will then share lessons learned and changes made to the protocol for the final data collection that we will administer via NORC's AmeriSpeak panel. We will conclude with a discussion of how the triangulation of these data can be used in the creation of new abortion attitude measures that may reduce dissonance or provide insight into the seemingly dissonant responses.

Should it be Legal and is it Moral? A Multi-Dimensional Assessment of Abortion Attitudes in the U.S.

Ronna Turner, *University of Arkansas*

Brandon Crawford, *Indiana University*

Wen-Juo Lo, *University of Arkansas*

Kristen N Jozkowski, *Indiana University*

People's attitudes toward abortion are complex, potentially due to the multi-dimensional conceptualizations some people have regarding abortion, creating challenges for measurement. To examine the multi-dimensionality of abortion attitudes, we examined people's attitudes through frameworks of two dimensions--morality and legality. Using a nationally representative sample (N = 919) collected via Ipsos' Knowledge Panel, we asked whether abortion should be legal and is moral in seventeen circumstances, based on items from the General Social Survey. At the aggregate, abortion was rated lower on the Morality scale (M = 3.158, SD = 1.241; 1 to 5 range) than Legality scale (M = 3.477, SD = 1.276) indicating people are more likely to believe abortion should be legal than is moral. When disaggregated by type of circumstance, the difference between morality and legality scores was smallest for circumstances related to woman's physical and mental health and rape--for these circumstances agreement that abortion is moral and should be legal was highest. Endorsement for both morality and legality was lower for abortion in circumstances such as financial reasons, contraception failure, and gender selection. The largest discrepancy in endorsement of abortion being moral and legal occurred for social reasons (e.g., financial, marital status of the couple). Across political affiliation, the smallest discrepancy between endorsement of morality and legality occurred for Republicans (.191 to .262) and those with no political affiliation (.091 to .292)—groups that most often oppose abortion. Independents and Democrats supported abortion legality at a higher rate than abortion being moral, with Democrats having the largest discrepancy (.109 to .517). People's attitudes toward abortion vary depending on the framework used to measure attitudes and this variability may be exasperated across certain subgroups. Researchers should consider other dimensions that may underlie abortion attitudes. Additional implications for abortion attitude measurement development will be discussed.

Applying Data Science Methods in Survey Research

Mitigating Biases in AI With Survey Science

Ruben Bach, *University of Mannheim*

Frederic Gerdon, *University of Mannheim*

Christoph Kern, *Ludwig Maximilian University of Munich*

Frauke Kreuter, *University of Maryland, Joint Program in Survey Methodology*

High-stake decisions in various contexts are increasingly guided by algorithms from artificial intelligence and machine learning. However, if the data that are used to train the algorithms contain biases from, for example, unequal participation of societal groups in the data collection, then the algorithms' predictions will likely be biased. Decisions derived from these algorithms may then result in unfair and potentially discriminatory decisions. Computer science research on fair machine learning developed a wide range of auditing techniques for evaluating the results of such processes in recent years and has started to develop methods for correcting biased algorithms. Survey science, however, has a long history of studying, understanding and correcting for biases in data collection contexts. This includes methods for improving inference from non-probability samples, utilizing information from "found" digital trace data and data integration techniques that aim at combining heterogeneous data sources. Thus, survey science can contribute to the advancement of the fair machine learning literature by evaluating data sources and correcting for potential biases before algorithms are trained. This paper reviews advances and challenges of research on data biases and algorithmic decision-making from a survey research perspective. It will focus on the question how survey scientists can contribute to research on unbiased algorithms and fair machine learning by helping detect, quantify and correct sample biases and selective participation in various data collection processes.

Scale Effects in Data Collection: When Does "Big" Mean Good?

John Eltinge, *United States Census Bureau*

In recent years, the survey community has encountered important challenges arising from increasing costs and declining quality in data collection. These challenges have led to in-depth development of data sources, methodologies and technologies intended to use "scale effects" to produce substantial improvements in quality at relatively small incremental cost; or to produce substantial reductions in cost with limited impact on quality. This paper explores some ways in which the concepts of "efficiencies of scale" and "diminishing returns" can be important in practical evaluation of the abovementioned developments. Five issues receive principal emphasis. (1) Review of "scale effects" related to measures of data quality; fixed and variable components of cost for data collection; design features that can be re-scaled to affect quality and cost; and evaluation of associated growth rates, network effects and development trajectories. (2) Application of these concepts to several areas of customary sample survey practice, including addition of questions or modules to omnibus survey instruments; shared use of collection platforms and other software; screening interviews for rare subpopulations; multiple-frame, multiple-mode surveys; address-based sampling; responsive and adaptive designs; and record linkage. (3) Extension of these concepts to the integration of multiple data sources, e.g., surveys, administrative records, web-scraped information and other forms of "big data" or "organic data." (4) Realistic methods for transparent empirical assessment of current or prospective efficiencies of scale in practical data collection work; and caveats on the reproducibility and replicability of some of these assessments. (5) Norms and standards for practical communication with stakeholders regarding quality, cost and related scale effects.

Concerns About Privacy When Sharing Data Collected Using Smartphones

Bella Struminskaya, *Utrecht University*
Florian Keusch, *University of Mannheim*

Smartphone data are increasingly used by researchers to study social and behavioral phenomena. Passively collected data from smartphone sensors (e.g., accelerometer, GPS) and log files (e.g., calls and SMS, app usage) have the potential to provide accurate information about users' behaviors and might thus replace survey questions. In addition, smartphone users are often asked to perform active research tasks (e.g., take pictures, scan receipts) to facilitate and augment self-reports. However, providing such data can pose certain risks to the participants and users might not agree to collect or share smartphone data due to security, privacy, and confidentiality concerns. If research participants who are concerned about their privacy and confidentiality refuse to share smartphone data differ from those who comply with such requests in attributes measured by the smartphone, the conclusions reached by researchers can be biased. In this presentation, we provide an overview of several of our recent research projects about the effect of privacy concerns on willingness to share smartphone-collected data. We show that users have differential concerns when asked to share different types of smartphone-collected data. These concerns vary across subgroups of smartphone users with different use habits indicating that conclusions drawn from the data might be biased. We investigate the role of trust in the data collecting organization, whether emphasizing privacy and confidentiality in the request to share data has an effect on willingness to share, and whether privacy concerns diminish with increasing smartphone use skills. Furthermore, we discuss the optimal ways of addressing privacy concerns in smartphone data collection and data sharing requests.

Crowdsourcing During a Pandemic: The Statistics Canada Experience

Martin Renaud, *Statistics Canada*
Wesley Benjamin, *Statistics Canada*

The COVID-19 pandemic created an urgent need for data on the impact it had on the Canadian population. Drawing on its previous experiences with crowdsourcing, Statistics Canada decided to use this cost-effective method to collect timely data, judging the conditions were favorable for using this non-probabilistic collection approach. From early April 2020 to the end of August, eight different crowdsourcing data collection cycles were completed, each on a different topic of interest affecting the lives of CanadianS in the context of the pandemic. This presentation will give more information on how the pandemic context led Statistics Canada to rely on crowdsourcing as a means to collect information on the impact of the COVID-19 pandemic on the Canadian population. The advantages and disadvantages of the method will be covered and emphasis will be put on the efforts that were implemented to ensure data quality and methodological rigor. Bias being a well-known problem of non-probabilistic methods like crowdsourcing, details will be presented on efforts that were implemented to mitigate such bias. An overview of research projects that were undertaken on methods to combine crowdsourcing data with data from probabilistic surveys will be given along with some of the initial results that have been observed. Finally, future crowdsourcing perspectives in Statistics Canada will be presented.

Machine Learning Predictions of COVID-19 Outcomes Using Survey Panels

Adam Eck, *Oberlin College*
Alex Dovichi, *Oberlin College*
Averly Sheltraw, *Oberlin College*
Becky McQuilken-Scholze, *Oberlin College*
Ishaan Thota, *Oberlin College*
Trent D. Buskirk, *Bowling Green State University*

The ongoing COVID-19 pandemic has had many short-term and anticipated long-term impacts on people and society. Not only have individuals suffered health consequences, but they have also experienced effects on their livelihoods (e.g., furloughs and loss of employment) and access to resources (e.g., housing, food, education). In a valiant effort to use their collective expertise to help society, many survey researchers have turned their attention to collecting information measuring the attitudes, behaviors, experiences, and opinions of the public in order to better understand how everyone's lives and shared

world are changing as a result of COVID-19. Survey panels, in particular, are especially valuable due to their longitudinal nature. These panels collect information that not only captures moments-in-time (recorded on each survey) but also how attitudes, behaviors, experiences, and opinions are changing over time (recorded across surveys). In this study, we investigate how information collected through nationally representative survey panels responding to the COVID-19 pandemic, such as the Understanding America Survey administered by USC and the Household Pulse Survey by the U.S. Census Bureau, can be analyzed using supervised machine learning methods (e.g., decision trees and forests, neural networks) to create insights into the impacts on respondent's lives. In particular, we study whether it is possible to predict future outcomes (e.g., COVID-19 testing and infection, loss of employment, decrease in food security) based on responses collected in earlier surveys. We also investigate what information captured by survey questions is most predictive of these outcomes. Although biomedical testing is clearly the most appropriate measure for determining whether someone is medically affected by COVID-19, insights gathered through this study could help inform more targeted testing protocols (e.g., screening who might need testing soon to best allocate limited testing resources), as well as guide public policy for mitigating other determinantal effects on people's lives.

Reproducibility Assessments of Analytic Findings Derived Through National Survey Data Integration Efforts.

Steven Cohen, *RTI International*

The quality and content of national population-based surveys are enhanced through integrated designs that link additional medical, behavioral, environmental, socio-economic and financial content from multiple sectors. A recent effort by the Committee on National Statistics of the National Academy of Sciences is serving as a catalyst to advance future national data integration efforts, as indicated in their recent report on Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps. These integrated data platforms include content drawn from administrative sources and research studies to enhance analytic capacity. The integration of national survey data with content derived from multiple data sources has the capacity to provide greater insights than possible from any of the component sources. Based upon the level of granularity characterizing the data profiles of the component data sets, the data integration effort can result in many-to-many linkages. While it is preferential to restrict these data enhancement efforts to situations that permit one-to-one linkages, this is often not possible. In this presentation, we focus on a data integration effort that is impacted by many-to-many linkages and provide a model for assessing the reproducibility of analytic findings attributable to alternative linkage applications. Examples are provided using data from the Medical Expenditure Panel Survey (MEPS) and cancer patient-level phase III clinical datasets. This data integration effort permits studies assessing the influence of health-related and socioeconomic factors, access to and use of health care services, health behaviors and preferences in concert with clinical trial treatment effects on cancer patient outcomes, heretofore not possible. In this study, we assess the reproducibility and stability of the analytic findings identifying factors influencing patient outcomes as the linkages are modified. Building similar evaluations into data integration efforts may serve to provide additional evidence in support of the integrity resultant findings.

Evaluating Methods and Modes in Election Polling

Election Forecasts Can Be Improved Using Wisdom-of-Crowds Methods

Henrik Olsson, *Santa Fe Institute*

Mirta Galesic, *Santa Fe Institute*

Wändi Bruine de Bruin, *University of Southern California*

Drazen Prelec, *Massachusetts Institute of Technology*

We show that election forecasts in 2020 and 2018 US elections are improved by adding wisdom-of-crowds questions to election polls. Wisdom-of-crowds questions ask poll participants about the percentage of their social contacts (social-circle question) or people in their state (state-winner question) who might vote for different candidates. These questions can be used to forecast elections on their own. They can also be integrated with responses to traditional poll questions that ask participants who they themselves will vote for (own-intention question). We explore two different methods for information integration: the Bayesian Truth Serum that overweighs participants providing more informed and careful answers, and full information forecast that optimally weights private and external signals. We compare different methods on national probabilistic samples (N=4000+) within the USC Understanding America Study panel. We preregistered our 2020 forecasts at <https://osf.io/x8jfk/>. The best predictions of the current 2020 election results were made by the social-circle question and most information integration methods. These forecasts outperformed forecasts based on average national polls. They were also better than forecasts based on the own-intention and state-winner questions in the USC poll. Similarly, in 2018, the social-circle and information integration methods outperformed own-intention and state-winner forecasts of national results. While our polls were not designed to provide state-level forecasts, we can compare the relative predictive accuracy of different questions within the same USC poll. In both 2020 and 2018, both wisdom-of-crowds questions and most information integration methods predicted state-level results better than the own-intention question. We investigate three reasons why wisdom-of-crowds questions might be more accurate: they might implicitly provide a more diverse sample, decrease social-desirability biases, and anticipate social influences on how people will vote. Taken together, our results suggest that election polling can produce accurate results when traditional questions are augmented by wisdom-of-crowds questions.

Is There a Shy Trump Voter: The Results of Two Experiments

John Huffman, *Growth Focused Insights and Research*

There are some suggestions that many political polls were inaccurate because of the "Shy Trump Voter". That is, supporters of President Trump are reluctant to reveal their vote intentions in surveys. This paper analyzes the data from two different surveys and two different experiments. One experiment randomly assigned half of the respondents to two different questions, one of which asked directly who they planned to vote for President, the other asked who their neighbors would vote for. The second experiment asked the same respondent both questions: who they voted for and who their neighbors voted for. The vote intentions are analyzed in terms of several variables including psychographic and attitudinal characteristics, impact of COVID, policy preferences, political party affiliation, media consumption, social media use, drivers of presidential preference and standard demographic variables. The data is gathered from 2,200 responses to two online surveys of a national panel of US adults 18+.

Georgia on My Mind: Examination of Data Validity Across Samples and Modes

Meng Li, *Ipsos*

Randall K. Thomas, *Ipsos Public Affairs*

John Lien, *Ipsos Public Affairs*

Frances M. Barlas, *Ipsos Public Affairs*

With online samples increasingly being used for polling, there have been questions about the efficacy and accuracy of such samples. Though online samples are often used for general population studies, interest has been growing in their use for smaller groups or geographies. We conducted a political poll in Georgia between August 13th 2020 and September 8th 2020 using three separate sample sources – 1) probability-based KnowledgePanel (KP), 2) an opt-in non-probability online sample, and 3) redirected inbound calling sample (RICS), which used IVR technology to present questions and receive answers. The number of participants we had for KP, opt-in, and RICS were 528, 1,331, and 1,185 respectively. We ensured that all items were formatted to be similar to those presented for the RICS sample to ensure comparability across modes and samples. Besides political opinion items, we also asked a series of questions that had benchmark values available for Georgia from Census, ACS, and NHIS (high quality federal surveys). All data were weighted to common targets for age-gender, education, income, and race-ethnicity. We found that results for KP and RICS were similar to each other for a number of political items but that opt-in values were comparable in other ways. In addition, we compared the sample sources for absolute differences between survey-based estimates and benchmark values. We discuss the implications of these different sample sources for use in surveys for targeted geographies.

Virginia 2020 Election Polling: A Comparison of Live Telephone Surveys Using Registered Voter Files and IVR Surveys via Redirected Inbound Call Sampling (RICS)

David Taylor, *Roanoke College*

Harry Wilson, *Roanoke College*

Scott Richards, *Reconnect Research*

During the week and half leading up to the 2020 presidential election, the Institute for Policy and Opinion Research conducted simultaneous polling of Virginia likely voters using two data collection methods. Live telephone interviews were conducted, sampling from a registered voter file provided by L2 Political, using a likely voter screen. During the same field period, interactive voice response data was collected using Redirected Inbound Call Sampling (RICS) whereby telephone calls placed by Virginians that were unable to reach their intended party were redirected to RICS and invited to take part in our survey. This presentation will compare using these methodologies, analyzing raw demographics, the effects of weighting, sources of potential bias and error, and, of course, the results of the horse race questions (president and U.S. senator) versus actual results for the Commonwealth of Virginia.

Comparing Phone and Web Pre-Election Polls in the 2020 New Hampshire Primary

Zachary Azem, *University of New Hampshire Survey Center*

Sean McKinley, *University of New Hampshire Survey Center*

Andrew Smith, *University of New Hampshire Survey Center*

Elections are always a vital event for survey researchers who conduct political polling as they provide an opportunity to compare survey results with an actual event – the election outcome. In addition to conducting its traditional New Hampshire primary RDD telephone poll for CNN, the University of New Hampshire Survey Center conducted a concurrent web-based poll using a probability-based web panel developed by the Survey Center, the Granite State Panel. While the telephone poll was accurate, the web panel was somewhat more accurate as it better picked up the late momentum of 3 place finisher Amy Klobuchar. This paper examines the differences between the two pre-election surveys, including mode effects, demographic biases, and field times. We present lessons learned from conducting polls using web panel polls during this election as well as later New Hampshire State Primary elections. As the industry moves further into using web-based polling, research like this is vital to understand the differences and ramifications of this shift.

A Comparative Study on the Presence and Statistical Modelling of Mode Effect in Election Polling for the 2016 and 2020 United States Presidential Elections

Heather H. Kitada Smalley, *Willamette University*

Christopher Wolf, *Oregon State University*

Public skepticism of polls has become common as inconsistent results are reported and unexpected election outcomes occur. While it has become commonplace to discuss the effect of mode on public opinion surveys, there has been little discussion regarding mode effects in election polls; however, this topic seems natural to investigate. The literature supports that social desirability bias is particularly prevalent for questions of a sensitive nature and in the presence of an interviewer. Therefore, we explore election data for the presence of mode effect using time series with generalized additive models (GAM). Mode effect is then estimated at state and at nationally aggregated levels to perform statistical mode adjustments, which are compared to observed election results. In this project, we build on our previous research (presented at AAPOR) modeling the “Shy Trump” phenomenon in the 2016 election polls and compare it to longitudinal statistical mode effect models in the 2020 election polls. The 2016 data contain information on 196 pollsters from November 2015 to November 2016. Using these data we were also able to show spatial trends and areas where mode effect was statistically significant. The 2020 data contain information on 223 pollsters from November 2018 to November 2020.

Differential Nonresponse by Party Identification in an Online Survey During the 2020 Presidential Election

Xiaolu Yan, *Ipsos*

Frances M. Barlas, *Ipsos Public Affairs*

One of the questions emerging from the assessment of the 2020 presidential election polling is whether or not polls experienced differential nonresponse from Republican or Trump-supporting Americans. A differential response rate by party identification could detrimentally affect the results of partisan questions like vote choice, candidate favorability ratings, and importance of issues facing the country, skewing the results to a more liberal or Democratic-leaning position. KnowledgePanel®, the first and largest online probability-based panel in the United States, is well positioned to conduct a study-level nonresponse analysis because party identification is part of our profile information on all panelists; we therefore can compare party identification among all those invited to take a survey to the identification distribution of those who completed the study. In the weeks leading up to the election, we fielded a number of general population studies of varying field periods, with from overnight polls to those with three, five and seven or more days in field. Overall, results showed minimal evidence of differential nonresponse by party id, though shorter field periods may have been more likely to see a slightly lower cooperation rate among Republican panelists. This suggests that a longer field period with online polls should be considered where possible and polls with a shorter field period should pay attention to the party ID distribution among respondents.

Combatting Partisanship in Measures of Issue Salience: A Novel Approach

Lucas Pyle, *Tufts University*

Ryan Shaffer, *Tufts University*

As the partisan divide has grown, issue ownership and voters’ perceptions of which political parties handle certain issues best has become more important and divided than before. Our study acknowledges the partisan divide and looks to break through partisan backing by testing a new method for measuring issue ownership. By diminishing the effect of partisanship in surveys, our study seeks to better understand which issues are most important to the electorate and which issues members of each party agree that one party handles best. We find that our new method of studying issue ownership decreases the impact of partisan backing and provides insight into which parties control which issues

Respondent Experiences, Burden, and Processing

Development of a Satisfaction Construct for Usability Evaluation of Questionnaires

Lin Wang, *U.S. Census Bureau*

Anthony Schulzetenberg, *U.S. Census Bureau*

Alda Rivas, *U.S. Census Bureau*

Elizabeth Nichols, *U.S. Census Bureau*

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

Shelley Feuer, *U.S. Census Bureau*

Usability evaluation is an integral part of survey questionnaire development. Questionnaire usability evaluation involves three aspects of respondents' survey-taking: success, efficiency, and satisfaction. Survey satisfaction refers to respondents' subjective feeling of being satisfied with the experience of taking a survey. Satisfaction is important because it influences respondents' willingness to answer survey questions, and consequently affects survey completion rates. As a latent variable, satisfaction is indirectly measured through a questionnaire that includes items deemed to be associated with the construct of satisfaction. Though a few instruments are available for measuring satisfaction of using software, an instrument measuring survey satisfaction is yet to be developed. We conducted a study of developing the survey satisfaction construct as the first step towards developing an instrument measuring survey satisfaction. Based on literature review, we started with a hypothetical two-tier model of satisfaction: General satisfaction (feeling satisfied with life in general) and survey-specific satisfaction; then empirically tested this model by conducting a six-question debriefing on 79 respondents who had just completed a survey. The debriefing collected respondents' opinions on satisfaction without a specific context (general satisfaction) as well as on satisfaction with completing a survey (survey satisfaction). Six hundred and seventy one phrases describing satisfaction were extracted verbatim. Of the 671 phrases, 222 unique phrases were identified. Three common themes emerged from the data: affection, cognition, and process. Each unique phrase was assigned to one theme. Multinomial logistic regression was performed to assess association between satisfaction type (survey vs. general) and themes. The findings support our hypothesis of a differential theme distribution between general and survey satisfaction. Survey satisfaction is more associated with experiences in the process of completing a survey, whereas general satisfaction is more associated with emotion induced by experiences in daily living. These results provide a basis for developing a survey satisfaction instrument.

User Satisfaction With the 2020 Census Online Questionnaire

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

The 2020 Census offered an online self-response option that over 53.5 percent of U.S. households used. The Census Bureau conducted a follow up user satisfaction survey with a sample of those respondents. The goal was to get an overall user satisfaction rating with the online form, as well as to gather feedback from respondents on what areas of the decennial census were difficult or caused confusion. Throughout the decade leading up to the 2020 census, researchers at the Census Bureau had conducted usability tests of the online instrument prior to the field tests. We were interested in comparing usability results from across the decade with the UX2020 survey results. While the overall satisfaction rating of the online decennial census questionnaire was high, there were variations across subgroups. For example, user satisfaction was higher for respondents whose address was prefilled on the form versus those who had to enter their addresses. Results also show that user satisfaction for early responders was higher than for late responders. Additionally, usability issues that were identified in user testing earlier in the decade, but had not been addressed, emerged as issues that respondents' mentioned as being problematic. For example, the Hispanic Origin question, separated from the race question, had caused confusion during usability testing and this was also reflected in respondents' feedback to the satisfaction survey. This talk will share these and other findings from the satisfaction survey of the decennial census, identifying issues we had suspected would cause confusion for respondents along with new results that we had not seen in user testing throughout the decade.

Bearing the Burden of Multitasking: The Impact of Respondent Multitasking on Perceived Difficulty and Burden

Erin Heiden, *University of Northern Iowa*

Eva Aizpurua, *Trinity College Dublin*

Ki H. Park, *University of Northern Iowa*

Mary E. Losch, *University of Northern Iowa*

Satisficing has been used to explain measurement errors in survey research. According to this theory, the probability of respondents providing suboptimal responses (i.e., engaging in satisficing behaviours) is the result of three factors: respondent motivation, respondent ability, and task difficulty. The latter will be affected by multiple aspects linked to the questionnaire design (e.g., how cognitively demanding items are), the mode of data collection (e.g. how familiar respondents are with the procedures and devices) and the context in which respondents answer the survey (e.g., environmental noise, multitasking). Previous studies have shown that respondent multitasking is linked to higher levels of distraction and lower levels of self-reported concentration. This study builds on previous research by examining the relationship between reported multitasking and perceived task difficulty, measured using self-reports of difficulty answering the questions and the perceived burden of the survey. We use data from a dual-frame random-digit-dial telephone survey of adults in a Midwestern state who were queried on topics related to awareness of and attitudes toward STEM education. Based on previous research, we anticipate that multitasking will be linked to greater perceived difficulty and a heavier burden. We also hypothesize that this will be particularly true for respondents with lower levels of education/knowledge (respondent ability) and interest (motivation). The implications of these findings and how they fit in with previous work will be discussed.

Using Dependent Interviewing to Ease Respondent Burden in a Self-Administered Survey

Jennifer Sinibaldi, *National Center for Science and Engineering Statistics*

Darby Steiger, *Westat*

The Survey of Doctorate Recipients (SDR), collected every two years, recently changed its design to include both cross-sectional and longitudinal samples. For many respondents, the main variables of interest (e.g., job title and employer) will not change biannually. Therefore, respondents would provide the same answers they reported last cycle. If the web survey was pre-filled with the responses from the last cycle and respondents were asked if the information is still accurate, it could potentially reduce respondent burden and the administration time of the survey, not to mention provide consistency of open-ended responses. This technique of confirming previously reported information, called dependent interviewing, is not often used for these benefits, and rarely in self-administered modes, making the application on SDR novel. To assess the feasibility of using dependent interviewing (DI) with the longitudinal sample in the 2021 SDR data collection, we implemented a pilot study using respondents (and their data) from the most recent SDR cycle. The pilot tested two versions of the SDR web questionnaire that differed in how the DI information was presented, alongside the traditional (not pre-filled) SDR web questionnaire to serve as a control. In addition, each SDR questionnaire was followed with new questions about the respondent's experience, addressing topics like their sensitivity to and enjoyment of DI. Research questions of interest included: real and perceived time to complete the survey, underreporting of job changes, usability issues, evidence of satisficing, and respondents' impressions of the experience. The two versions of the DI questionnaire were compared for differences in data quality and respondent preference. This presentation will share the design of the questionnaires and experiment, and report the preliminary results of this pilot to answer each of the research questions.

Using Electronic Paradata to Assess Predictors of Respondent Burden and Difficulty in a Youth Tobacco Survey

Matt Jans, *ICF*

Deirdre Middleton, *ICF*

Jennifer Greer, *ICF*

Sindhu Geda, *ICF*

Xiaoyi Deng, *ICF*

Ed Duckhorn, *ICF*

Akshay Sanganal, *ICF*

Andrea Gentzke, *Centers for Disease Control and Prevention*

Sean Hu, *Centers for Disease Control and Prevention*

Electronic survey paradata can help quantify respondent burden and difficulty. We explored how respondent demographics, survey mode, and tobacco use influence three burden and difficulty measures: (1) questionnaire duration, (2) question revisits and answer changes, and (3) item nonresponse. In 2019 and 2020, 33,549 middle and high school students completed the school-based National Youth Tobacco Survey (NYTS). Respondents completed an electronic questionnaire (with skip patterns) via tablet (98%) or web (2%) in person in classrooms; absent students could complete a make-up web-based questionnaire (5% of all completes). On average, respondents took 13 minutes to complete the survey (2019: 104 questions; 2020: 117 questions). The average ratio of revisits/answer changes to total questions asked was 0.16. Average item nonresponse was 2%. Between 11% and 26% of the variance in these three measures could be predicted; models were better at predicting revisits/changes and duration than item nonresponse. Including past-30-day cigarette and e-cigarette use as predictors did not improve model fit. Many predictors were significant due to large sample size, but most significant effects were small in magnitude (e.g., & 1 minute in duration, & 1 point in revisits/changes ratio, and 1 percentage point in item nonresponse). The strongest questionnaire duration predictors were being in middle school (v. high school) and Hispanic (v. White, non-Hispanic), both of which predicted longer durations (about 1.39 and 1.25 minutes, respectively). The strongest question revisits/changes predictor was mode, with web respondents (v. tablet) having a lower ratio of revisits/changes (about 0.15 less in both years). Race/ethnicity was the strongest item nonresponse predictors, with Hispanic students and Non-Hispanic Black students having about 1-2 percentage points higher item nonresponse (v. Non-Hispanic White). The results can inform question and procedural improvements for NYTS and similar youth surveys, including best practices for how paradata can measure respondent burden and difficulty.

Prompting Recall Inclusion: Wording, Time and Intensity; Results From the Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium

Sara Putnam, *University of Minnesota*

Melissa Constantine, *University of Minnesota*

Charles Cain, *University of Minnesota*

Kyle Rudser, *University of Minnesota*,

Emily Lukacz, *University of California San Diego*

Linda Brubaker, *University of California San Diego*

Leslie Rickey, *Yale University*

Todd Rockwood, *University of Minnesota*

For the Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium, *University of Minnesota*

Health research commonly asks people to recall past events which are affected by event salience. As salience decreases, individuals are likely to use heuristics to formulate responses. Many approaches have been used to enhance recall, including attaching recall to a salient event, prompting backward, forward or free recall, and deconstruction. Research conducted by the Prevention of Lower Urinary Tract Symptoms Research Consortium evaluated the impact of a simple prompt for a low salience item which is likely to be subject to exceptions. This experiment (n=654) included a basic variation in wording to test the effect across different initial prompts for recall.

- 1a, no prompt: Have you ever stopped doing things you enjoy, because of your bladder?
- 1a, prompt: Have you ever stopped doing things you enjoy, even if for just a short period of time because of your bladder?
- 1b, no prompt: Did you ever stop doing things you enjoy, because of your bladder?
- 1b, prompt: Did you ever stop doing things you enjoy, even if for just a short period of time because of your bladder?

Including the prompt promoted greater endorsement of occurrence for questions 1a (6% increase, p=ns) and 1b (9% increase, p=0.1, one tailed test). This increase was primarily the result of individuals shifting from the No, never stopped response category to the next category of Yes, one or two things. The effect is likely due to changing the heuristic used to formulate response as opposed to facilitating direct recall and enumeration. Given that the shift in responses was between two response categories, it is reasonable to use this type of prompt to identify the gross classification that something is likely to have occurred in the past. However, this type of prompt cannot be interpreted as directly enhancing the accuracy of the amount it occurred.

Associations Among Response Effects: Understanding Mechanisms

Allyson Holbrook, *University of Illinois Chicago*

Sowmya Anand, *University of Illinois*

Response effects occur when small features of questionnaires affect responses (e.g., acquiescence response bias or no opinion responding) or when groups of respondents show response patterns that seem to be unassociated with meaningful group differences in the construct being examined (e.g., extreme response style). A variety of theories and mechanisms have been proposed to explain response effects and when they occur, including satisficing theory and culture. Most research examining response effects has focused on one (or a limited set) of such effects, but much less work has examined the associations between different types of effects. Examining associations among response effects may provide insight into the mechanisms that underlie such effects, as effects with similar underlying causes or mechanisms should be highly related (and those with disparate mechanisms should not). We analyze data from a survey with a convenience sample of 900 respondents collected via Mturk in which we examined a large number of response effects in order to examine the associations among them to provide insight into shared mechanisms. The survey included measures of response effects including acquiescence response bias, no-opinion responding, nondifferentiation, effect of response numbering and response categories, extreme response style, response heaping or rounding, and effect of anchoring in slider questions. Our analyses focused on identifying whether response effects were associated with one another and whether such associations could be used to identify clusters of related effects with similar causes or mechanisms.

Using Census Data to Understand and Plan for Nonresponse

Exploring the Relationship Between Social Factors and Government Survey Participation

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

Luke J. Larsen, *U.S. Census Bureau*

Benjamin M. Reist, *NASA*

In this presentation, we explore the relationship between social factors, such as religious service attendance and volunteerism, and survey participation. In a previous study, we explored the social-economic-political construct of survey participation. However, we did not have good measures to fully study the social aspect of this construct. Recently, we had the opportunity to obtain better measures as part of the 2020 Census Tracking Survey, a national survey conducted by the U.S. Census Bureau that tracked public awareness of the 2020 Census. Respondents were asked about their plans to participate in the 2020 Census as well as a variety of other questions, including questions related to civic engagement and empathetic concern. We proposed these items as measures of one's social environment that might be related to one's attitude towards participating in a government survey. Using the Tracking Survey data, we modeled whether a respondent reported planning to participate or having participated in the 2020 Census against known and hypothesized factors of survey participation. We found that civic engagement and empathy were both important factors of reported participation in the 2020 Census.

Predicting Self-Response Rates for the 2020 Census

Steven Scheid, *U.S. Census Bureau*

Megan Parker, *U.S. Census Bureau*

A major part of the 2020 Census was the self-response operation, which attempted to motivate residents to respond by telephone, mail, or online. An accurate estimate of self-response at multiple levels of geography was required to plan for the 2020 Census and monitor data collection throughout the operation. To meet that need, the Census Bureau created the Predicted Self-Response Score (PSRS). This paper describes the creation of the PSRS and how it extends previous U.S. Census Bureau efforts to predict self-response rates. We will discuss the process of developing a new self-response regression model, including evaluating modeling approaches and selecting predictor variables. We will also discuss how fitting the model to American Community Survey response rates allows us to incorporate online self-response into the model, but required the predictions to be adjusted before being used in 2020. We will compare the predicted and actual self-response rates for the 2020 Census for the U.S. and for various lower levels of geography.

Do Census Audience Segments Predict Response Behavior To Mixed-Mode Household Surveys? Evidence From Address-Based Samples

Michael Jackson, *American Institutes for Research*

Recent research by the U.S. Census Bureau (Kulzick et al. 2019) clustered every census tract into one of eight “audience segments”, defined as “groups of census tracts with similar predicted self-response behavior...selected for their distinctive patterns of media consumption and distribution of mindsets.” Files identifying each tract’s audience segment are publicly available through the Bureau’s Response Outreach Area Mapper (ROAM) application. Although primarily designed to assist in tailoring public outreach efforts for the 2020 decennial census, these audience segments may also predict response behavior to other studies of households, particularly those that (like the census) initially request self-response by web or mail. If they do, they may be a useful resource to survey managers, insofar as they offer preexisting cohorts between which data collection strategies could be varied to achieve a more balanced distribution of responses. Using data from several cycles of the National Household Education Survey (NHES), a large-scale recurring cross-sectional study, this presentation will examine the utility of the Census audience segments for tailoring data collection strategies in mixed-mode, address-based household surveys. The discussion will focus on two research questions. First, do response rates, eligibility rates, and the distribution of responses by mode vary meaningfully between the audience segments? Second, based on randomized experiments incorporated into recent NHES cycles, is there evidence that the effectiveness of common data collection interventions—including cash incentives, advance mailings, and the offered response mode—varies between the audience segments? Time permitting, variation in substantive survey estimates will also be discussed. Results will provide insight into whether the Census Bureau’s tract-level audience segmentation research can be usefully applied in data collection contexts other than the decennial census.

Leveraging Census Planning Database Low Response Scores and Self-Response Rates to Increase Address-Based Sample Productivity

Jonathan Best, *SSRS*

Cameron McPhee, *SSRS*

Luis Tipan, *SSRS*

Arifah Hasanbasri, *SSRS*

Phil Benner, *SSRS*

Todd Hughes, *UCLA Center for Health Policy Research*

The Census Planning Database offers a wealth of demographic information down to the block group level that can be used to stratify address-based samples to target groups that are typically under-represented in ABS samples. For example, areas with a higher incidence of low-income households can be oversampled to help increase the number of low-income respondents in an ABS sample. However, one reason that these groups are typically under-represented is that they respond to surveys at lower rates. Thus, targeting these groups has the effect of decreasing response rates and increasing data collection costs. The Census Planning Database includes a low response score (LRS) and a self-response rate (SRR). This paper explores the feasibility of using these measures to create strata to target areas with higher response. The hope is that this will allow researchers to increase the incidence of harder-to-reach populations in address-based samples while minimizing the associated decrease in response rate. This paper will attempt to answer this question by analyzing the Census Planning Database on a national level, examining the distributions of the LRS and the SRR in areas with a high proportion of traditionally underrepresented groups (e.g., Hispanics, Low-income households, etc.). Additionally, we test the theory by applying it to the ABS sample used in the 2019 California Health Interview Survey. We will evaluate the effectiveness of stratifying the sample using the SRR by examining the response rates among subgroups identified in the different strata as well as the demographics and substantive survey responses to estimate the impact of such a design on sample representativeness.

Using Census Planning Database Tools to Predict Mail-Out Survey Response Rates: Comparison of the Low Response Score to the ACS Self Response Rate

Thomas Guterbock, *University of Virginia*

Kara S. Fitzgibbon, *University of Virginia*

Sean M. Johnson, *University of Virginia*

Laura A. White, *University of Virginia*

Edward Rincon, *Rincon & Associates LLC*

Mail-out surveys using Address-Based Sampling [ABS] generate variable non-response rates across geographic areas. Researchers can address this problem by oversampling areas expected to have low response, a method we call Response-Rate Adjusted Stratified Sampling Selection or RRASSS. A key source of information on expected response rates for tracts or block groups is the U.S. Census Planning Database [PDB]. Since 2014 the PDB has included a “Low Response Score” [LRS] based on a model that retrospectively estimated rates of mailed response to the 2010 Decennial Census (Erdman and Bates 2017). The 2020 PDB included a new variable at the tract level: the Self Response Rate [SRR]—the percentage that responded via mail or Internet to the 2014-18 ACS. We ask: Which is the better guide for RRASSS designs, LRS or SRR. We compare the predictors in four contexts. We show that LRS and SRR data for the Dallas region point to quite different areas as likely to have low response to the 2020 Decennial Census. Data suggest that the SRR would have been more predictive than LRS. We then examine three recent mixed-mode surveys using ABS samples: a survey of 95 counties across two states, covering the catchment area of a university cancer center; a survey of one partly urbanized county; and a survey in a densely populated suburban region. The first was stratified by tract, the others by block group. We compare the efficacy of LRS and SRR as alternative guides for RRASSS designs, looking first at how well they predict realized response rates. We then consider how well each tool, when used to guide differential sampling rates, would succeed in increasing socioeconomic and racial diversity in the realized sample. By comparing highly different survey contexts, we address the robustness of RRASSS designs based on LRS versus SRR.

The Robustness [Fragility] of Predictive Response Rate Analytic Models across Surveys and Survey Organizations

Colm O'Muircheartaigh, *NORC at the University of Chicago*

Chang Zhao, *NORC at the University of Chicago*

Ned English, *NORC at the University of Chicago*

We demonstrate how disposition histories inform a model that can be used to predict the number of completed interviews and response rate for a survey, and how that model may be used to predict the results of a different survey. In so doing we show that within a survey organization a model is relatively robust across surveys, despite challenges due to potentially incongruent instruments, incentive structures, questionnaire content and length. However, in transporting the model to a different field organization, major challenges may arise. Our research is useful for survey practitioners interested in developing and refining early prediction of survey completion and response. Since 2003 we have been developing predictive models to understand non-response in face-to-face surveys conducted by NORC, including the General Social Survey (GSS), The National Social Life, Health, and Aging Project (NSHAP), the National Longitudinal Survey of Youth (NLSY), and Making Connections. Such models serve as predictive tools that use data analytics to understand non-response in face-to-face surveys. In addition, we employ extant area-level information from the American Community survey and other sources to improve our model. Our approach has used past experience to predict a final number of completed interviews and derive a predicted final response rate earlier in data collection than would otherwise be available.

Strategies for Panel Management and Sample Representation

Designing an Overlapping Sample Refreshment for SIPP

Jason Fields, *U.S. Census Bureau*

Mahdi Sundukchi, *U.S. Census Bureau*

Tracy Mattingly, *U.S. Census Bureau*

In response to the 2018 wave 1 experience and realized (low) response rate, the Survey of Income and Program Participation (SIPP) management team initiated plans to revisit overlapping panels, last implemented in the early 1990's. These ensure that the cross-sectional analytic samples remain robust sources of intra-year economic and household dynamics, while preserving the ability to follow individuals and households over multiple years. The decision to initiate overlapping samples in 2019 was preceded by budget uncertainties and data collection challenges in 2018 and prior years. In preparation for the first wave of the 2018 SIPP Panel, the SIPP team at the Census Bureau revised the sampling plan to accommodate budgetary uncertainty related to funding through continuing resolutions. Instead of releasing the full sample and holding the interviewing period open for four-months, in collaboration with the Census Bureau's field management, we released sample for five discrete one-month collection periods. While this allowed the Census Bureau to reserve the latter two-fifths of sample in-case funding was not received. For several reasons this was not as successful as desired, incurring a significant amount of management burden and staffing uncertainty. Leading to 2019, we made the decision to return to a longer (4 to 5-month) interview period and adapt to low response by supplementing the fielded sample with new representative cases up to our funding threshold. In 2019 a new overlapping sample was added to the field workload. In 2020, another new overlapping sample was added. These were planned as part of the overlapping sample design, but have become more critical as the impact of the federal government furlough in 2019 and the COVID-19 pandemic was realized. This paper presents the design for the overlapping samples initiated in 2019, discusses longitudinal and cross-sectional data quality, implementation considerations, and product considerations.

Combining Data From Probability and Nonprobability Samples Using a Composite Estimator

Burton Levine, *RTI International*

Akhil Vaish, *RTI International*

Taylor Lewis, *RTI International*

Michelle Cavazos, *RTI International*

Rachel Harter, *RTI International*

We present methodology for combining probability and nonprobability samples that could improve the quality of survey estimates. Extending earlier approaches, we develop an optimal composite estimator that accommodates the scenario where both the probability and nonprobability estimates contain known bias in an estimate. Using our derived composite estimator, we examine the relationship between the bias characteristics of the samples and the weight attributed to each sample by the compositing factor. If the nonprobability sample estimate has much more bias than the probability sample, the composite estimator is mostly a function of the probability data. We present methodology to quantify an allocation of probability and nonprobability sample that minimizes the mean squared error of the estimate based on cost per completed interview and the bias and variance characteristics of estimates from the two samples. The addition of cost to the analysis will inform future allocations of probability and nonprobability respondents. We illustrate the application of these methods to combine a probability ABS (address based sample) survey and a nonprobability RICS (redirected inbound call sample) survey, an emerging nonprobability sampling methodology where misdialled calls to toll-free numbers are redirected to an interactive voice response system.

The Lifecycle of a Probability-Based Panelist

David Dutwin, *NORC at the University of Chicago*

Ipek Bilgen, *NORC*

J. Michael Dennis, *NORC*

Probability-based panels recruit their panelists from a single initial sampling frame and draw. Over time, additional panels are recruited with additional draws, while panelists who joined at an earlier time often begin to decline in their level of participation in the panel, or depart it entirely. This creates a complex myriad of a collection of samples, a collage of new and old panelists that have varying participation levels in subsequent surveys. This paper explores the lifecycle of these panelists within a probability-based panel. We will use NORC's AmeriSpeak, a probability-based panel largely recruited from a national frame of addresses via multiple mailed recruiting efforts and a substantial in-person nonresponse follow-up effort. Now entering its 7th year, like all panels it is a collage of a wide range panelists with regard to tenure and participation. In this presentation, we will explore the following questions about panelists' lifecycle: What is the pattern of participation for panelists as they age from novice to highly tenured? Which types of panelists exhibit significant decline in participation over time, and what is the net implication for the overall demographic cross-section of the panel? Do groups exhibit significantly different patterns of decline, for example, where some groups depart the panel more quickly than others? What can we say about the cost of certain types of panelists, given that groups exhibit differential costs in initial recruitment, and then later in their propensity to participate in an invited survey? Finally, can we predict who will become a lifelong "true believer" of the panel versus someone who departs from the panel shortly after recruitment? The presentation will conclude with an overall assessment of panel health and needs for panel maintenance, given the patterns we have seen from AmeriSpeak panelists over a six year period.

Over-Recruitment and Selective Retention: Examining a Strategy for Increasing Panel Selection Efficiency

Kenneth Winneg, *University of Pennsylvania*

Josh Pasek, *University of Michigan*

Chintan Turakhia, *SSRS*

Jennifer Su, *SSRS*

Kathleen Hall Jamieson, *University of Pennsylvania*

Panel survey quality depends on the composition of respondents recruited to a study as well as on the ability to maintain sampled respondents over time. Unfortunately for the composition of panel surveys, the population subgroups that are difficult to recruit are often particularly susceptible to attrition. To correct for this sort of error, scholars conducting panel surveys often oversample difficult-to-reach populations and upweight the respondents that remain. Both procedures serve to increase survey design effects for probability-based samples, thereby reducing the effective sample size. In this study, we present results from an alternative procedure for ensuring that typically underrepresented populations were included in a large probability-based panel study of the 2020 U.S. Presidential Election.

Due to better than expected recruitment rates among sampled households when many individuals were subject to stay-at-home orders in March, too many respondents were initially recruited to the study. This necessitated an additional subsampling procedure wherein some high propensity to respond individuals were randomly excluded from the panel. In this study, we examine the effects of over-recruitment and selective retention of respondents on the effective sample size of the study, the design effect of the study, and the overall costs of administration. We find that the exclusion of a subset of high propensity to respond individuals yielded large increases in the efficiency of the sample. We then discuss how variations in the number of panel waves required as well as the relative costs of recruitment and retention may make this a useful approach for panel selection under some circumstances.

Mixing Probability and Non-Probability Surveys: Testing the Blenders

Nick Mastronardi, *Polco*

Parker Quinn MS, *Polco*

Michelle Kobayashi, *National Research Center/Polco*

Many believe that the disruptions we are seeing in the field of survey research might best be solved through the blending of probability and non-probability surveys. The hope is that two methods combined will help to mitigate the individual biases of each method creating a hybrid that is both more accurate and cost effective. To better understand this blend, The National Research Center at Polco examined its national database of more than 100 matched sets of probability and non-probability samples from The National Community Survey™ (The NCS). The NCS is a templated, benchmark used by local governments across the nation to measure quality of life, community livability, local government performance and civic participation. Probability-based surveys have been aggregated in this database since 1997 and non-probability surveys (a second survey completed alongside the probability surveys) since 2013. The database includes surveys for more than 850 communities (probability samples) and 100+ matched pairs of probability and non-probability samples. The NCS database was used to identify the best weighting algorithms to create stronger hybrid data sets. In this analysis we compared raking, matching, propensity weighting and multiple regression with post stratification (MRP). Our analysis looked at not only which model is more effective to calibrate the data sets, but also which is the best fit for purpose given a number of the best models may be less feasible given their complexity and focus on single outcome fitting.

Does the Online Panel Choice Matter? Differences in Data Quality Among Online Sample Providers

Rositza Dobrikova, *AYTM*

Ivan Konanykhin, *AYTM*

Over the last decade, research data quality emerged as a major concern and a topic of heated discussions. Online sample providers differ in the ways they recruit respondents to join their panel and the ways they communicate with and reward their panelists. This exploratory study compares the data quality of US online panels, measuring both the number of respondents who failed data quality checks and the extent to which their inclusion in survey data affects the study results. Besides traditional metrics for assessing online data quality (speeding detection, open-ended responses review), we deploy a series of logic checks to evaluate respondents' data integrity. To determine the degree to which a panel-specific low-quality response may affect the outcome of a study, we perform statistical tests, controlling for false discovery rate. Additionally, correlations between various demographics (age, gender, ethnicity, education, etc.) and data quality indicators across panels is examined. The results of the data quality evaluation across the fourteen online sample sources are substantively different. We conclude that the choice of an online survey sample provider is critical, as it can impact data quality, data integrity, and ultimately, study results.

Non-Prob Online, Phone RDD and Phone Quota: Research Choices in the UK

Peyton Craighill, *U.S. Department of State*

Stephen Hornbeck, *US State Department*

Mary Sanderson de Flores, *US State Department*

The Office of Opinion Research has relied on robust dual-frame phone poll designs for research efforts in the United Kingdom. However, nearly all publicly available polling in the UK has moved to opt-in online panels. Subsequently, the market for telephone polling has narrowed considerably and prices have skyrocketed. In response to the COVID crisis, our office experimented with a non-probability Internet panel, to compare with a traditional RDD dual-frame design, as well as a quota based dual-frame omnibus phone poll. This paper will examine differences in overall data quality across a variety of metrics including attitudinal and demographic differences. Many of the comparisons will focus on attitudes and experiences with COVID. A key research question is to gauge the stability of trend data across these different methods, and whether stable political trends will be comparable. Additional analysis will include time of administration and cost.

Gauging Nonresponse Bias in Recruitment to a Probability-Based Online Panel

Scott Keeter, *Pew Research Center*

Nick Hatley, *Pew Research Center*

Andrew Mercer, *Pew Research Center*

Jesse Lopez, *Pew Research Center and Duke University*

As survey nonresponse has relentlessly driven up costs, many in the field have responded by turning to probability-based online panels. While expensive to recruit, such panels can be cost-effective in the long run and they have other benefits such as the ability to accumulate rich background data on participants. But nonresponse is not just a driver of costs. Given today's low response rates, nonresponse bias is an ever-present concern, even for probability-based panels. This paper will assess possible nonresponse bias in the recruitment of the 2020 cohort of panelists (N=1,636 enrollees of 14,383 sampled addresses) for Pew Research Center's American Trends Panel, a large panel now recruited using address-based sampling. Addresses sampled for the recruitment were matched to a commercially available voter file, which contains information about most of the individuals who live at each address. The files include voter registration status, past voting behavior (including both primary and general elections going back several cycles), party of registration (where relevant), a modeled political party score, age, race and location information such as voting precinct. Because the effort did not attempt to recruit a named individual at each address, the analysis will compare the characteristics of the collective group of people associated with addresses that yielded an enrolled panelist with those associated with addresses that did not yield a panelist. The results will speak to two important potential biases: the extent to which people who respond to surveys of this type are more politically engaged and different in terms of partisanship than those who don't respond. Additional analyses will examine the political characteristics of the places sampled, with a comparison of those that yielded a panelist and those that did not.

5/14/2021, 2:00 PM - 3:30 PM
Concurrent Session I: Live Sessions

Mini-Conference: Measuring and Tracking Attitudes and Behaviors Throughout the Pandemic

Longitudinal Data and Coronavirus: How Americans Lives and Views Have Changed (or Not) as a Result of the Pandemic

Ruth Igielnik, *Pew Research Center*

Since early 2020, Americans across the country have been experiencing a variety of disruptions and life changes as a result of the coronavirus pandemic. Leveraging the power of the more than 10,000 panelists in Pew Research Center's American Trends Panel we are able to look longitudinally at how people's lives have been impacted by the dual financial and public health crises. As millions of American families are stuck at home together, one area of exploration is how relationships have changed. Questions on relationship satisfaction among married or cohabiting adults originally asked in the summer of 2019 and repeated again in October of 2020 show little topline change in overall relationship satisfaction but looking at individual-level change tells a more nuanced story. This analysis will include a regression analysis to see whether certain aspects related to the pandemic -- such as job loss, kids going to school remotely, etc. -- may contribute to relationship satisfaction or assessments of how the relationship is going. In addition, a longitudinal exploration into job loss will examine Americans who have experienced long-term unemployment as a result of the coronavirus pandemic. Due to the robust sample sizes available through the American Trends Panel, this analysis will have the power to explore which groups have been differentially impacted by job loss and the lasting financial and personal effects. We have been able to field questions about job loss repeatedly over the course of the pandemic to help create a clear picture of the long-term unemployed. Deeper analysis reveals what factors are at play in how different groups of Americans experience different financial impacts of the crisis.

Spreading Knowledge, Not a Virus: Conducting Long-Term Public Opinion Research Around COVID-19

Mallory Newall, *Ipsos*

Chris Jackson, *Ipsos*

Frances M. Barlas, *Ipsos*

Linda McPetrie, *Ipsos*

The COVID-19 pandemic created unprecedented challenges and changes to our world in 2020. However, it also provided researchers with a unique opportunity to study Americans' behavior and attitudes against the backdrop of a once-in-a-generation health crisis. Since early March, Ipsos has partnered with Axios to track public opinion about COVID-19, through our Axios-Ipsos Coronavirus Index (with 30 waves, and counting, as of November 2020), run on Ipsos' probability-based KnowledgePanel with approximately 1,000 completes per wave. We compared our results with national benchmark results from CDC-released data and found clear correspondence between self-reported data and the CDC data demonstrating the external validity for our results. Though we analyze shifts and trends from week to week, one broader question remains: even after the pandemic ends, what are the lingering effects that Americans may face? This paper will tap into that large and robust dataset, and summarize both attitudinal and behavioral trends along with the mental, physical, and economic effects of the COVID-19 pandemic. Though the foundation of this paper is survey research among the general population, our research is also informed by a survey among people who have tested positive for the novel coronavirus. With 30+ waves of respondents, we are able to create a sub-panel to examine this niche and underreported audience. This paper will describe how attitudes among people who have actually experienced COVID-19 firsthand may differ from the broader public, and what lingering effects they may be facing as a result of contracting the virus.

Pandemic Response: The Challenges of Measuring “Stay-at-Home” Behavior

Michael Link, *Abt Associates*

Terms such as “quarantine,” “shelter-in-place” or “stay-at-home” are not new concepts, but have taken on new meaning in the context of a global pandemic. This can make measuring public behavior with regards to these important concepts a bit tricky. The meaning of “stay-at-home” during a pandemic differ greatly from a similar order during an acute event, such as a tornado warning. Moreover, public messaging around these concepts varies considerably across government and non-governmental agencies. Even within a single community, there may be competing messages and differences in understanding these concepts. Lack of clear and consistent messaging adds to the difficulty of operationalizing and measuring even seemingly simple concepts such as “stay at home”. Examining this issue, we leveraged a large database of COVID-19 related questions administered by leading survey firms and universities and made available publicly beginning February 15, 2020. Questions dealing with the concept of “stay at home” were selected and coded in terms of various features of the question such as specificity, notation of specific time frames or activities, etc. Building on a three-fold framework (e.g., “relative questions” – those with no objective point of reference; “qualifier questions” – those providing some type of “mental anchor” or object reference; and “multidimensional questions” – which delve into multiple, specific aspects of the concept), we examine both the nature and evolution of these questions over time as both the pandemic continued and the survey industry had more time to “compare notes” and circle in on more common ways of asking these questions. How researchers construct measurements of key concepts – such as “staying at home” - can have a profound impact on the data collected and our understanding of their implications for both behavior and policy.

Fielding a Nationally Representative Survey During COVID

Kerry Levin, *Westat*

Jocelyn Newsome, *Westat*

Jill DeMatteis, *Westat*

Maeve Gearing, *Westat*

Jonathan Wivagg, *Westat*

Once the COVID-19 pandemic hit the United States (U.S.), social science researchers began scrambling to field studies quickly. The goal was to gather time-sensitive information during an unprecedented time. Worldwide, about 1,800 social science articles about COVID-19 were published in the first half of 2020 (Aristovnik, Ravšelj, & Umek, 2020). Given the urgency, most U.S. studies opted to utilize panels and non-probability panels, resulting in findings not generalizable to the U.S. population. Westat collaborated with Stanford University’s School of Medicine to field one of the first nationally representative surveys on COVID-19. The development of the survey, which focused on the social, health, and economic impact of COVID-19, began in April and the survey was fielded in May 2020. Survey topics included: beliefs about the risk of contracting COVID-19 and probable health outcomes; how people’s behaviors had changed since the pandemic began; the influence of stay-at-home orders on those behaviors; and the economic impact of the pandemic. The study utilized address-based sampling and a single mode of data collection; the web survey collected both quantitative and qualitative data. This paper will discuss the process of deciding upon the methodology of the study, the experience of fielding the study, and selected findings. For example, we will discuss findings surrounding the financial impact of the pandemic, concerns voiced by respondents, and coping strategies.

Emergency Survey Response: Leveraging Partnerships for Global Representation

Frauke Kreuter, *University of Maryland, Joint Program in Survey Methodology*

Samantha Chiu, *University of Maryland, Joint Program in Survey Methodology*

Adrienne Bradford, *University of Maryland, Joint Program in Survey Methodology*

Xiaoyi Deng, *University of Maryland, Joint Program in Survey Methodology*

Alex Reinhart, *Carnegie Mellon University*

Curtiss Cobb, *Facebook*

Sarah LaRocca, *Facebook*

Katherine Morris, *Facebook*

Alyssa Bilinski, *Harvard Graduate School of Arts and Sciences, Interfaculty Initiative in Health Policy*

At the onset of the COVID-19 pandemic, analysts needed rapid, diverse, and accurate information to understand the spread and transmission of the disease. Organic, "big" data, such as GPS data to track community movements, provide detailed information about one topic (depth). However, surveys allow researchers to study a breadth of topics, including symptoms, social behaviors, financial changes, and mental health. Partnering with a global social media platform, Facebook, our survey was quickly disseminated to a representative sample across the world (averaging 200k respondents per day) in 56 different languages. The rapid response produced a high volume of daily statistics regarding the pandemic -- designed "big" data. The partnership with Facebook allows us to collect data from a probability-based sample. Facebook users selected via geographically-stratified random sample are directed to an online survey hosted by the University of Maryland or Carnegie Mellon University. Facebook also developed weights that adjusted the sample to represent each country's population. The sampling methodology avoids bias arising from convenience sampling processes often used when time is limited. Furthermore, the deliberate and precise measurements offer the opportunity to link these survey data with external sources of data. These surveys have produced real-time statistics to inform pandemic response.

Mini-Conference: Pivoting From Live to Virtual Call Centers and Training

Altered State: Performance Impacts of a Rapid Pivot From On-site to Virtual Call Center

Kurt Johnson, *RTI International*

Kelly Lynn, *RTI International*

Mandee Lancaster, *RTI International*

The coronavirus (Covid-19) pandemic has had a profound impact on data collection at a time of crucial importance for survey research. At RTI International, we faced the unexpected need to quickly pivot all telephone research from a centralized, large-scale on-site model to a fully Virtual Call Center (VCC) environment. The rapid transitioning of hundreds of interviewers and supervisors to maintain operational continuity and avoid project disruptions was complex, challenging and involved numerous stakeholders. Research and experience tell us that an onsite versus a home workspace can have a direct impact on productivity and the quality of work conducted. In this case, we assess quality and productivity metrics across several long-term survey projects comparing performance prior to and after shifting staff to remote interviewing. Preliminary analysis suggests that while the initial impact of moving to a virtual environment were immediate and indicate an initial decline in productivity, ongoing and responsive changes to the infrastructure, operational practices and processes, and support of the virtual environment corrected disparities. The presentation will discuss both the perceived causes and the adaptive solutions of addressing the shift in paradigms.

Creating and Utilizing a Remote Call Center During a Pandemic

Larry Danforth, *Center for Community Studies - Jefferson Community College*

Eager to ascertain the perceptions and behaviors of Northern New York residents in response to the COVID-19 pandemic, the local health departments looked to the Center for Community Studies at Jefferson Community College (Watertown, NY) to assist. The Center traditionally has employed current Jefferson Community College students to conduct interviews utilizing a local call center when telephone methodology is used for data collection. Unable to rely on their established telephone data collection methods, the staff at the Center were able to recruit and train a workforce of interviewers, work with the local health departments to design a survey instrument, and create a remote call center for data collection in a period of approximately two and a half weeks. This presentation will provide a timeline detailing the process the staff used to prepare and collect the data for this study.

Creation of Remote Call Center During COVID-19 – Moving From Centralized Calling to Remote Calling

James Dayton, *ICF*

Mary Penn, *ICF*

Aprille Hairston, *ICF*

Don Allen, *ICF*

ICF immediately considered the potential outbound call center impacts of COVID-19 following the first documented cases in the United States. As the pandemic gripped the nation, the probability increased that that centralized call center outbound dialing would be severely restricted or prohibited outright. Within weeks many jurisdictions and companies, including ICF, limited in-person operations to include only essential employees. Although it had been nearly a decade since ICF conducted even small-scale remote interview operations, our past experience permitted us to quickly develop robust remote interviewing contingency planning processes centered around work-from-home (WFH) interviewer requirements, quality assurance and technical implementation. We will share our experience and challenges moving from 100% on-site call center operations to 100% off-site remote operations over the course of a few weeks. We will cover the process to determine if needed call center technology can be safely and securely deployed in a home-based environment that is outside of company firewalls and in compliance with our internal data protection protocols and Internal Review Board (IRB) respondent protection requirements. Can the required technology be effectively deployed using consumer-level internet connectivity in interviewer homes where other household members may be attempting to work, attend on-line classes and have other broadband needs. Assessments establishing the suitability of interviewing

staff home-based work environment will also be discussed. Finally, we will explore the safe deployment of equipment, development of interviewer technical support helpdesks and the required updates to our interviewer supervision, quality assurance processes and other management procedures to assure the success of our home-based interviewing staff who no longer have the luxury to interact face-to-face in a centralized work environment.

Transitioning From In-Person Training to Virtual Training in Response to COVID-19

Brenda Rodriguez, *Abt Associates*

Strong, well-trained interviewers are a foundation of high quality data collection. Traditionally, in person training programs are an invaluable method of ensuring field interviewers' understanding of data collection procedures and protocols. As the National Evaluation of the 2nd Generation of Health Profession Opportunity Grants Program (HPOG 2.0) entered the planning stage for the second wave of survey data collection, COVID-19 and social distancing recommendations forced the project to rethink our approach to training. Transitioning our traditional in person training to a completely remote training approach meant developing methods to provide the same high-quality training to our interviewers without the benefits of in-person instruction. Our goal was to create training sessions that were flexible enough that they were both conducive to on-line learning and allowed project staff to evaluate field interviewers' project knowledge and comprehension of the data collection protocols. The evaluation team divided the training into two main components: 1. Self-guided Modules: Six self-guided training videos focused on study background, collecting reliable data, guidance for interviewing, gaining respondent cooperation, importance of locating, and data security; 2. Virtual Classroom Sessions: Two interactive half-day sessions via video conference focused on question-by-question specifications, hands on locating procedures, sample management, incentives and frequently asked questions. Forty experience field interviewers successfully completed the two part remote training session. This presentation will describe the techniques used to create the self-guided training videos and virtual training sessions. We also discuss which methods were most successful, challenges encountered, and important lessons learned.

Now You See It, Now You Don't: Adapting Visual Aids Designed for In-Person Use to Interviewing by Phone

Andrew Jannett, *Westat*

Catherine Billington, *Westat*

Colleen Keating, *Westat*

In-person data collection protocols often make use of visual aids, such as show cards, to increase the accuracy of responses and improve data quality. The goal of including these peripheral materials is to facilitate selection of the most accurate response by the respondent. Visual aids can be helpful when response options for a question are numerous, lengthy, or complex. They may also help the respondent respond more accurately to questions that are sensitive, confidential, or personal. Visual aids may display text, numbers, images, diagrams, or even videos. Traditionally, visual aids are either printed or displayed on a computer screen during the interview. With the onset of the COVID-19 pandemic in the US, most in-person research was temporarily halted. Some research projects were able to continue data collection by converting to interviewing over the phone. However, this presents a fundamental challenge for the use of visual aids. While research on the efficacy and use of in-person peripherals is sparse, information on how they can be transformed to support phone data collection, and the potential effect on data quality, is emergent at best. This presentation describes several options for adapting visual aid content to phone data collection. Approaches include posting visuals to the web, screen sharing, mailing paper copies, or re-formatting interview questions and response options. We review the challenges, concerns, and benefits of each approach. We present qualitative feedback from interviewers regarding use of visual aids, and discuss the need to identify whether the remote aids were actually used by respondents and on what devices, if applicable. More generally, we highlight the need to track the use of visual aids, and discuss a reliable metric to quantify the value of visual aids used in data collection in any mode.

Mini-Conference: Qualitative Data Collection Methods: COVID-19 Adaptations

Conducting Focus Groups During the Pandemic - What Have We Learned?

Cynthia Robins, *Westat*

Darby Steiger, *Westat*

Kerry Levin, *Westat*

Qualitative researchers across multiple disciplines have been faced with unanticipated challenges during the COVID-19 pandemic (e.g., Marques & Theiss, 2020; Dodds & Hess, 2020). With in-person focus groups not possible during a pandemic, studies have had to pivot to online options. But, behavioral norms are not the same in online focus groups as in person, and there are significantly more distractions in people's homes than at a focus group facility. Consider the following situations that could occur during an online focus group: A male participant is lying in bed not wearing a shirt; a 5-year-old is loudly demanding her mother's attention; a dog is barking in the background; a participant is speaking but did not unmute her line. Before the pandemic, the chance of encountering any – let alone all – of these scenarios during focus group sessions was zero. Focus groups were carried out in-person in a brick-and-mortar facility. Respondents were carefully prescreened. Children and dogs generally did not accompany the participants, and clothing was never optional. Moreover, facilities had a gatekeeper and could send anyone home whose demeanor or behavior were questionable. Now, however, qualitative researchers need to expect the unexpected. This presentation will focus on such challenges faced by our research team, and lessons learned as we approach another year of focus group data collection-from-a-distance. We will address: Tips for recruiting participants; Features to consider when selecting an online platform How to address technical support needs of participants, observers, and moderators both prior to and during the group; Topics for re-training moderators for online groups; How to ensure respondent privacy and confidentiality; How to implement a distress protocol, if needed and; How to manage client expectations for the online focus group experience.

The Pluses and Minuses of Virtual Focus Groups With Vulnerable Youth: A Systematic Assessment

Kathryn Henderson, *Westat*

Debra J. Rog, *Westat*

Bernadette Blanchfield, *Westat*

Noah Lipshie, *Westat*

Ellie Kerr, *Westat*

COVID-19, and its accompanying shelter-in-place policies, posed significant challenges for many data collection efforts, including the evaluation of the U.S. Department of Housing and Urban Development's Youth Homelessness Demonstration Program (YHDP). YHDP is an initiative that encourages diverse communities across the country to develop and implement coordinated community approaches to prevent and end homelessness for youth, ages 14-24. A critical component of the evaluation of YHDP is the inclusion of youth's perspectives about how well available services and supports are meeting their needs and their recommendations for change. Youth voice in our baseline data collection is included through in-person focus groups conducted across 13 case study communities. When shelter-in-place policies restricted researchers' ability to travel to sites to collect in-person data, the evaluation team adapted its data collection strategy by shifting to virtual focus groups. Shifting to a virtual format for conducting over 30 focus groups with a population of vulnerable youth required the adoption of new processes for recruiting and consenting participants and setting up and conducting the focus groups, including innovative strategies for eliciting rich conversation in virtual settings. In addition to the content data collected on the experiences and perspectives of the youth, data were also collected to assess the quality and efficiency of the virtual focus groups. Data were collected and analyzed on participant recruitment, technological issues, process adaptation issues, and variables related to the quality and amount of data collected on the content of the groups. In this presentation, we share the results of this analysis and lessons learned for maximizing the virtual format for data collection. In particular, we identify the situations most conducive for conducting virtual groups with this vulnerable group that may extend to other vulnerable populations.

Qualitative Research in the Time of Coronavirus: Evaluating the Quality of Cognitive Interviews Using Remote Communications Modes

Alfred Tuttle, *U.S. Census Bureau*

The Census Bureau routinely pretests survey instruments in order to maximize their validity and reliability and to minimize respondent burden. This pretesting work has traditionally been done by conducting cognitive and usability interviews in-person with respondents in our headquarters, or in other places convenient to respondents that offer privacy, such as public library study rooms. Due to the COVID-19 pandemic and social distancing protocols, researchers have halted in-person interviews and are relying completely on remote communications modes, primarily telephone and Skype for Business. While these alternatives to in-person interviews have so far proved adequate in terms of meeting the goals of pretesting, the switch to remote testing prompted an inter-divisional team of survey methods researchers to undertake a collaborative effort to systematically evaluate the quality of data obtained. Following Conrad and Blair (2004), we are developing a research agenda that operationalizes measures of quality to facilitate evaluations within, and comparisons between, in-person and remote testing modes. Examples of quality metrics include the number, accuracy, and presumed impact of questionnaire problems identified; the volume and quality of verbal reports; differences in interview management activities and their effects on respondent behavior and data quality; etc. Beyond the immediate needs of assessing the quality of our work under pandemic conditions, we see this as an opportunity to evaluate and improve our qualitative research generally and hope to maintain a regular research evaluation program going forward. We are looking for feedback on our direction for evaluating qualitative interviews and pretesting modes, and suggestions for other approaches we may not have considered yet.

Lessons Learned From Conducting Online One-On-One Qualitative Interviews During the Pandemic

Xin (Rosalyann) Yang, *Westat*

Ting Yan, *Westat*

Darby Steiger, *Westat*

Andrew Caporaso, *Westat*

Antonia Warren, *Westat*

Alexis Kokoska, *Westat*

COVID-19 has forced changes in data collection procedures. Survey researchers who conduct face-to-face qualitative interviews have needed to quickly transition to using online research methods. Previous research found online qualitative interviews to be a viable alternative to face-to-face interviews and that they increase the variety of samples (Deakin and Wakefield, 2013; Upadhyay and Lipkovich, 2020). However, some researchers suggest that interviews conducted online require more rapport building and generate fewer non-verbal cues (Shapka et al, 2016; Lo lacono et al, 2016). This paper draws upon our experience of conducting online one-on-one qualitative interviews (e.g. cognitive interviews, usability testing) with several different population segments during the pandemic, including youth. We will first present data on participants' receptivity to participating in an online interview by examining compliance to the request to turn on their camera and sharing their screen during the interviews as well as interviewers' observations of participants' comfort level and cooperation during the interview. We will review how we adapted protocols for interviewing youth and obtaining parental permissions in a virtual setting. We will then evaluate the effectiveness of virtual qualitative interviews in four different ways. First, we will discuss interviewers' experience on what went well and what went poorly. Second, we will systematically investigate the types of questions and issues that arise during online interviews through behavior coding of the recorded interviews. Third, we will evaluate the quality of information obtained during the interviews. Lastly, we will explore the characteristics of the people who responded to our recruitment ads to participate in an online interview. Given the challenges in data collection posed by COVID-19, these results will be of interest to researchers considering using online qualitative research methods.

Using Remote Cognitive Testing to Modify Questions About Education During the Coronavirus Pandemic

Kathleen Kephart, *U.S. Census Bureau*

Jonathan M Katz, *U.S. Census Bureau*

Jasmine Luck, *U.S. Census Bureau*

Jessica Holzberg, *U.S. Census Bureau*

COVID-19 has had a significant impact on K-12 education and therefore has presented unique challenges for the measurement of education during a time in which capturing these changes is extremely important. The response of schools to the ongoing coronavirus pandemic has varied widely over time and by region, further complicating measurement. In response to this challenge, we modified some of our typical question development and testing approaches and remotely cognitive tested and deployed questions on several education surveys. In this paper, we provide an overview of these efforts, which include quick remote cognitive testing with participants from geographically diverse areas. As part of a research project planned in fall 2019, we were conducting in-person cognitive testing of new content with school staff on virtual schooling for the National Teacher and Principal School Survey when the pandemic began. We quickly pivoted to remote cognitive testing to adjust probes and develop new questions that capture what schools did for distance learning. The questions are currently being fielded for the 2020-21 school year. We further adapted these questions to be asked of parents for a different survey, the Current Population Survey School Enrollment Supplement, and then tested them with the parents of K-12 students in late May/June for fielding in October 2020. In both of these projects, we used iterative, remote cognitive testing to provide maximum flexibility. We also leveraged the ability to recruit geographically diverse participants via remote testing. In this research, we will share our lessons learned about crafting and testing questions to measure the new and dynamic phenomenon of widespread distance learning during the coronavirus pandemic. We will also give an overview of school staff and parental experiences with distance learning in the spring of 2020.