AAPOR

AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH

77th Annual Conference

Come Together

Advancing Inclusion and Equity Through Data Collection, Measurement, and Community

Conference Program

May 11-13, 2022

Sheraton Grand Chicago • Chicago, IL

Contents

Assessing Question Validity Across Cultural Contexts

Responses Without Borders: A Large-Scale Qualitative Assessment of Response Scale Performance Across 11 Language Families
David Vannette, *Facebook*

Survey questions commonly use a scaled set of response alternatives that permit respondents to indicate varying strength and, in some cases, direction of their response to the question prompt. However, retaining measurement equivalence across languages and cultures is a substantial challenge in survey methodology given the simultaneous difficulties of, among other things, procuring comparable translations, maintaining the semantic distance between response scale points, and differential cultural response styles. This paper describes the results of a large qualitative effort to generate cross-culturally equivalent rating scales for 12 constructs (likelihood, concern, satisfaction, importance, amount, value, agreement, trust, when, ease/difficulty, usefulness, frequency). Our work proceeded in three phases: (1) translatability assessment by trained linguists; (2) 340 in-depth interviews in 11 diverse countries, including Japan, Côte d'Ivoire, India, Brazil, Egypt, Germany, Indonesia, Mexico, Thailand, United States, and Poland; (3) expert review across body of qualitative evidence. Our results indicate that linguistically and cross-culturally equivalent response scales are possible for the constructs we studied, despite some variation in the performance of those scales across different languages. In our presentation, we will present recommended response scales and share best practices for large-scale qualitative assessment

Measuring Tolerance for Ambiguity - A German-Language Adaption and Validation of the Tolerance for Ambiguity Scale (TAS)
Almuth Lietz, *German Center for Integration and Migration Research*

Diversity in lifestyles and opinions or in terms of social groups and religions shape life in Germany. Whether diversity poses a challenge to social cohesion or instead improves society's potential for innovation and creativity depends in particular on how a society values diversity. A central variable related to this is the Tolerance for Ambiguity (TA). Originally defined as a stable personality trait that is independent of situation and context, it is now more often assumed that TA can change over time and is partly understood as a more context-specific construct. Although it is an important indicator with regard to the acceptance of diversity and thus appears essential for the functioning of a pluralistic society, it is rarely asked in population surveys in Germany. Moreover, hardly any validated measurement instruments exist in the German language that are also suitable for population surveys, which are often restricted by time and monetary constraints. This paper proposes a measurement instrument in German language to address these concerns. The development is based on the English Tolerance for Ambiguity Scale (TAS), which is surveyed with 12 items covering four subdimensions: valuing diverse others, change, challenging perspectives, and unfamiliarity. The items were translated into German using the TRAPD approach and tested. In order to evaluate the psychometric quality of the translated TAS, reliability (Cronbach's alpha, test-retest stability) and validity (factorial and construct validity) are considered. The article makes at least two contributions. First, it proposes the German-language version, and second, it tests the scale using a confirmatory approach to factor analysis rather than the exploratory approaches used in previous studies

The Impact of Response Styles on Subjective Socioeconomic Status in Cross-national Research
Julie De Jong, *University of Michigan*
Caitlin Hayward, *University of Michigan*
Fred Conrad, *University of Michigan*
Chuzhu Zhong, *University of Michigan*
Junhui Qian, *University of Michigan*

Subjective socioeconomic status (SSS) has been established as a reliable predictor of objective socioeconomic status (SES) as well as overall health, well-being, and other population indicators. Measures of SSS are particularly important in studies of comparative global health and other cross-national research, where individual components of SES, such as

educational attainment and income, may not be directly comparable between countries and present a challenge to comparative analyses. At the same time, significant cross-cultural differences in how respondents answer survey questions have been demonstrated, including differences in acquiescent and extreme response styles, potentially compromising the value of comparative SSS across cultures. For example, researchers have found that collectivism, power distance, and other elements of Hofstede's framework of cultural dimensions can differentially affect the prevalence of response styles across cultural groups, highlighting the role that culture may play in the quality of survey data . On conceptual grounds, accurate measurement of SSS is vulnerable to bias from differential use of extreme response styles because of the scale on which it relies, but this has not been investigated empirically.   We address this issue in the current study by analyzing data from a web survey completed by learners (n=460,000) in University of Michigan MOOCs from over 200 countries. At the macro-level, we examine the extent to which various cultural dimensions predict response styles across individual countries and regions. At the micro-level, we compare the within-country predictive power of SSS with other demographic variables to investigate potential mechanisms associated with differences in response styles

The Moderating Role of Democratic Governance in the Association Between Personal Values and Political Ideologies
Zheng Wu, *Northwestern University*

Decades of research were dedicated to the dispositional psychological profiles of right-wing and left-wing political attitudes. Jost et al. (2003) offered a unifying motivated social cognition framework: people embrace right-wing ideology to reduce their uncertainty, and to justify inequality. People higher on right-wing orientation are higher on conservative values and resistant to change and new information, while people higher on left-wing orientation are higher on openness and self-transcendence values. I proposed that the government tradition of democracy moderates these relationships. Ideologies are socially constructed superstructures. In less democratic countries, people are less exposed to ideologies' discursive superstructures and less engaged with using political attitudes as venues for value expressions. It was thus hypothesized that lower democracy should attenuate values' associations with political orientation and cultural conservatism. Moreover, in countries with less democratic governments that operate on the tradition of having the government maximize individuals' welfare in a centralized manner, conservative values for resisting change are more compatible with attitudes for broad government interference in the economy. In these situations, conservative values should negatively predict economic conservatism. Using the World Value Survey 2011 to 2014 wave with 49 countries and both objective and subjective operationalizations of democracy, we found supporting evidence. Higher democracy is associated with more positive relationships between Schwartz conservative values (e.g., tradition and safety) and right-wing political attitudes; higher democracy is also associated with more positive relationships between liberal Schwartz values (e.g., benevolence and universalism) and left-wing political attitudes, controlling for competing moderators. Also consistent with our predictions, lower country democracy attenuates conservative values' associations with cultural conservatism and predicts more negative association between conservative values and economic conservatism.

Jewish Ethnicity and Democracy: A Conjoint Analysis
Hannah Ridge, *University of Chicago*

Israel exists in tension. It seeks to be both a democracy and a Jewish state. With a growing non-Jewish Arab population, these goals are difficult to maintain simultaneously. Politicians and citizens are faced with choices about Israel 's future identity. This project assesses the relative value Jewish Israelis place on potential elements of Israel 's future, especially democracy, Jewish identity, and peace through a traditional conjoint study on a Jewish Israeli panel based on Shamir and Shamir's (1995) examination of goals for a future Israeli state. Within the Jewish population of Israel , there are multiple ethnic groups who co-exist unequally. Ashkenazi Jews – Jews of European extraction – are culturally privileged in Israel and the US, relative to Sephardic and Mizrahi Jews – Jews of Middle East/North African extraction. This pattern is referred to as Ashkenormativity (Shohat 2003). Ethnic groups may have a different sense of linked fate with other communities in the state. Non-Ashkenazi Jews place greater value on the Jewish identity than the "secular Zionist project of modern Israel " (Lewin-Epstein and Cohen 2018, 2126). Recent research in race and ethnic politics suggests that groups show skin-color-based internal hierarchies, and intra-group complexion differences drive differences in political and racial attitudes (Ostfeld and Yadon 2021; Yadon and Ostfeld 2020). These less-privileged Jewish groups might then be more interested in preserving the Jewish majority and status position. The study finds that the relatively-

lower-status Jewish groups in Israel have different perspectives on Jewish nationalism and the resolution for the Israel - Palestine situation than Ashkenazim.

Social, Cultural, & Political Attitudes
Explaining the Gap between American Hispanics' and Non-Hispanic Whites' Attitudes and Beliefs on Global Warming

Jared McDonald, *Stanford University*
Jon Krosnick, *Stanford University*
Taylor Orth, *Stanford University*
Bo MacInnis, *Stanford University*

According to conventional wisdom, the only people who can afford to be concerned about the natural environment are those whose basic survival needs are satisfied. This presumption stands in contrast to decades of survey evidence showing that American Hispanics express "greener" beliefs about the seriousness and importance of climate change than do non-Hispanic whites, while the former individuals are less economically and socially privileged than the latter. This paper explores the causes of this "Hispanic green gap." Data from multiple surveys of representative national samples (N =4,808) conducted between 2006 and 2015 demonstrated that American Hispanics were more likely than non-Hispanic whites to view global warming as important to them personally and as a serious threat to the nation and the world. Even after controlling for party identification, Hispanics were shown to ascribe greater importance and seriousness to climate change due to heightened perceptions of personal vulnerability to the negative consequences of climate change. These perceptions of personal vulnerability have been caused by greater tangible risks that individuals face (such as proximity to the ocean, extreme heat) and less resiliency (such as lower income) to recover from the negative effects of a changing climate. Thus, consistent with theories related to environmental deprivation, the Hispanic green gap can be attributed to climate change having direct and local effects on American Hispanics, who then support policies intended to address climate change or mitigate its impact. Rather than being a "luxury" concern for people whose basic living needs are taken care of, green attitudes appear to be more prevalent among Hispanics precisely because they are more likely to bear the brunt of climate change.

Clean Air, Clear Scales: Experiment to Assess Scale Understanding Across Cultures

John Lee Holmes, *The Social and Economic Survey Research Institute (SESRI)*
Abdoulaye Diop, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Kien Trung Le, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Maryam Ali Al-Thani, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Rima Charbaji-Elkassem, *Qatar University Social and Economic Survey Research Institute (SESRI)*

This paper examines a split ballot experiment alternating a 10 point scale with a fully labeled 4 point scale to assess comprehension across ten languages and three socio-economic status groups surveyed by telephone in Qatar. Qatar University's Social and Economic Survey Research Institute (SESRI) conducts two omnibus phone surveys each year among its Qatari, high income ("white collar") and low income ("blue collar") migrant populations. One of the most difficult issues is assessing whether measurement of low income migrant workers is accurate when using a ten point scale with endpoint anchors. Frequently there is low variation in satisfaction scores between "blue collar" and "white collar" populations which presents a puzzle: to what extent are these results due to varying expectations between the two populations versus a simple measurement problem? On two recent surveys the issue of air quality was examined and low income migrants rated air quality higher than Qatari nationals or high income migrants. To assess the extent to which scales were understood similarly, SESRI researchers devised two experiments on two consecutive surveys. The surveys were conducted in ten languages with audio-visual monitoring for quality control on a CATI system. In this paper we review the results by socio-economic status (Qatari, white collar, blue collar) and by language family (Arabic, English/Tagalog, South Asian languages). End point anchored numeric scales are used in many cross-cultural surveys, such as the World Values Survey. There is reason to be concerned about whether different populations and socio-economic groups understand these scales the same way. Yet there is also a concern for concept equivalence across linguistic groups when translating fully labeled scales. We believe these experiments have value in verifying scales' robustness for cross-cultural survey measurement.

Use and Trust of Government Information Sources and Public Opinion About Climate Change in a Technocratic Society

Sonny Rosenthal, *Nanyang Technological University*
Pengya Ai, *Wee Kim Wee School of Communication and Information at Nanyang Technological University, Singapore*

Climate change is a pressing issue that governments are working to address from a top-down perspective. In contrast, citizens can take individual actions to help mitigate climate change, such as conserving energy. Even when efforts take such a bottom-up approach, government can play an important role by communicating with the public about the human cause of climate change, the challenges it presents for the future, and opportunities for individuals and society to enact solutions. But for government to be effective in that role, the public needs to trust it as an information source. The current study examines public opinion of climate change in relation to the use and trust of government information sources. Data come from a national survey of 1,118 Singapore heads of household using probability-based door-to-door sampling. Singapore is a unique place to study this topic for two reasons: (1) Singapore is regarded as a technocracy, using science as a basis of policy, and (2) the public generally trusts, or at least defers, to the government on many issues, including climate change. Separate linear regression models predicted belief in climate change and climate change skepticism. Belief in climate change was positively related to the use of government information sources (β = .10, p = .008) and trust of government information (β = .14, p < .001). Also, there was a positive interaction effect (β = .02, p = .019), where the greater the trust, the more positive the relationship between the use of government sources and climate change belief. The second model showed climate change skepticism was unrelated to the use of government information sources and negatively related to trust in government (β = -.13, p < .001). Although these effects were small, they parallel recent work on the effect of worldview in public understanding of climate change.

Shades of White: How White Identity Shapes Attitudes About Racial Politics

Geneva Cole, *University of Chicago*

Research on white identity has largely focused on a white racial consciousness associated with racial conservatism. However, I argue that there are multiple dimensions to white identity that reflect distinct racial narratives and are associated with different attitudes towards racial politics. Most existing work focuses on the first dimension, consciousness, which indicates political awareness of group position. I theorize a second dimension, called valence, which indicates how positively or negatively individuals view whiteness as a structure. Using an original measure—the white advantage scale—I find that awareness of the racial narrative of the social, political, and economic advantages that come with being white motivates support for both the Black Lives Matter movement and associated policies, like police reform and reparations. This paper uses a mixed-methods approach that combines a case study of the Minneapolis-St. Paul metropolitan area in Minnesota both before and after the 2020 uprising, and a nationally representative survey of white Americans. I find empirical support for the two dimensions of white identification and evidence that where individuals fall on these dimensions are associated with distinct patterns of attitudes about Black Lives Matter and racial justice policies.

Rethinking Rural Identity: Examining Associations Between Strength and Operationalization of Place-Based Identity and Rural Americans' Policy Preferences and Values

Alee Lockman, *Harvard University Faculty of Arts & Sciences*

A growing body of work has sought to understand what rural identity is, how to measue it, and how it influences rural Americans' role in the political process. Existing narratives in both popular media and academic literature suggest that "rural identity" is largely a reflection of racial and economic resentment and rural Americans' perception that urban elites look down on them; however this framing may underestimate the importance of strong place-based connectiveness and underlying values systems that may be linked to "rural identity." How "rural" is both defined by researchers and perceived by rural Americans themselves also influences how rural identity is studied. It is unclear how Americans' perception of their own place of residence influences the existence and effect of place-based identity on policy preferences and values. The lack of consistent and accurate measurement of what "rural" is further affects the way rural Americans' views are studied and how rural identity is understood by policymakers and researchers alike. More work is needed to better understand the existence of rural identity, how it intersects with other established group identities, and how it informs rural Americans' political behavior and views of major health and social policy issues.

Using public opinion data from the 2020 American National Election Survey, I investigate how the existence and strength of place-based identity is associated with Americans' values, beliefs, and opinions of health and social policy issues. I find evidence that rural Americans' views of numerous issues differ from those of non-rural Americans, even after accounting for other intersecting group identities. I also find that strength of place-based identity and how place of residence is perceived by rural Americans themselves differentially affects the relationships between place and policy views, suggesting the importance of incorporating subjective assessments of place in future work on rural identity.

Measuring Racism & Identifying Methods to Reduce It

30 Years Later: Impact of the 1992 Los Angeles Riots on Race Relations

Alejandra Alarcon, *Center for the Study of Los Angeles*
Brianne Gilbert, *Loyola Marymount University*

On April 29, 1992, a majority white jury acquitted four police officers of beating Black motorist Rodney King. A series of riots and civil disturbances, now known as the 1992 Los Angeles riots, broke out soon after the verdict was read. These events emerged during an intense and unstable political climate as Los Angeles residents dealt with the negative impacts of a national recession, a real estate downturn, a rise in crime, and increasingly tense race relations. In the days, weeks, and months that followed, the city and the nation aimed to understand the underlying causes that led to this degree of social upheaval. Questions about racial tensions demanded answers about Angelenos' way of thinking. Since 1997, the Center for the Study of Los Angeles (StudyLA) at Loyola Marymount University has marked the quinquennial anniversary of the riots with a cross-sectional study designed to produce trend data that captures residents' attitudes, concerns, and overall outlook on the city. The 30th anniversary study (2022) incorporates responses from 1,000 randomly and ethnically represented residents in the city of Los Angeles (margin of error ±3%). Like previous studies (1997, 2002, 2007, 2012, and 2017), the 30th anniversary study provides trend data on several themes such as direction of the city, trust in government, sense of community, race relations, and possible future riot activity. Results of uni- and bivariate analyses of the question categories described above investigate how response patterns on these topics have changed over time, demonstrate how issues of race relations and police and community relations have remained salient, and expound the legacy of the 1992 Los Angeles Riots.

Measuring Race Relations in Society Through Survey Research

Keith Neuman, *The Environics Institute for Survey Research*

Race and racial injustice are now front and centre issues confronting society – and not just in the USA. Canada is among the most racially diverse countries, but systemic racism continues to be an underlying reality, especially for those who are Black or Indigenous. Much of the prevailing narrative – driven by personal experience, anecdote, media reporting and political agendas – is divided between those who see a positive story of progress in civil rights over generations and those who see a negative story of enduring oppression of racialized people. There is no way to reconcile these competing perspectives, but social research can help bridge this divide by generating credible, empirically-based evidence that documents collective beliefs, attitudes and experiences across the population. The Race Relations in Canada research program is a groundbreaking example of how survey research can make an important contribution to this critical issue, by establishing benchmark indicators of race relations based on the perspectives and experiences of individuals across racial and ethnic groups, and measuring over time to identify how such indicators are evolving. National surveys were conducted in 2019 and 2021 with large national samples that included substantial representation from the country's largest racialized populations (Indigenous, South Asian, Chinese and Black) to capture the perspectives and experiences of each group (along with non-racialized Canadians). The research measured opinions about the state of race relations and the extent of discrimination against specific groups, perceptions about racism as systemic versus the result of individual prejudice, and direct experience with racial discrimination. The 2021 survey also explored how the COVID-19 pandemic intersects with race in terms of access to health care, vaccine take-up, and anti-Asian sentiment. This paper provides an overview of the research, presents key findings, and discusses lessons for building metrics around race and racism in society.

EPR: A Theory of Prejudice Reduction and Political Attitudes

Alauna Safarpour, *Harvard University*

This paper develops Engagement, Perspective-Taking, and Recalibration (EPR), a theory of prejudice reduction and its consequences on political attitudes. I argue that interventions using engagement to encourage perspective-taking reduce prejudice and recalibrate the subject's emotional orientation toward an out-group. Using EPR, I develop an intervention to reduce prejudice toward African Americans among a diverse sample of adults. The intervention encourages subjects to adopt the perspective of a Black man who experiences prejudice and make choices regarding how to respond. Using an online survey experiment in which 1,261 adults completed either the treatment or a placebo task, I find that the intervention significantly reduces prejudice, with the largest effects among those with the highest baseline animus. Reducing prejudice increases support for policies aimed at helping Black people. These results provide insight into the nature of prejudice and its impact on political attitudes, and offers a low-cost intervention to combat rising hate-related incidents in society.

Breaking the Bubble: Does Exposure Reduce Prejudice?
Martin Koenen, *Harvard Kennedy School*
Adrian Blattner, *Stanford University*

What are the effects of exposure to political out-groups on affective polarization? We analyze a field experiment in Germany that matches participants for a one-on-one conversation with a political opponent. First, we find discrimination in partner choice based on gender, immigration status, and age. Second, we document a taste for discrimination and significantly lower levels of trust towards supporters of other parties in incentivized dictator and trust games. Third, we provide suggestive evidence that both affective and ideological polarization is reduced among participants who have had a conversation with a political opponent. Together, the results provide suggestive evidence that even short exposure to political opponents can reduce polarization among individuals who seek contact with outgroup members. To test for the effect of contact in a population that did not seek contact, we propose an additional lab-in-the-field experiment.

'Truly American': A Multidimensional Experimental Approach to Unpacking the Symbolic Boundaries Around American Identity
Victoria Asbury-Kimmel, *Harvard University*

The criteria for membership in the national community is consequential for social attitudes, behaviors, and policy decisions. I engage innovative experimental and survey methods to provide a multidimensional portrait of who is considered truly American. I apply a paired choice and rating conjoint survey experiment on a nationally representative sample of 1,067 non-Hispanic White U.S. adults to establish the effect and relative weight of individual factors—including "ascriptive" and "acquired" characteristics—on perceived Americanness. I find that even after controlling for salient factors such as citizenship status, employment, criminality, welfare usage, and religion in the conjoint experiment, Asians, Blacks, and Latinos are perceived as less American than Whites. Additionally, the combination of being unemployed for 3 years, receiving welfare, and having a criminal history of assault, reduces one's perceived Americanness nearly as much as being a non-citizen immigrant. I also assess explicit moralized perceptions and Americanness evaluations of the four major ethnoracial groups in the United States in a racial attitudes survey and show, empirically, the interlaced nature of the moral and racial boundaries around American identity.*

A Focus on the Respondent Experience: Self-Administered & Interviewer-Administered Considerations

Redesigning and Assessing Online Survey Login Screens in Spanish and English
Erica Olmsted-Hawala, *U.S. Census Bureau*
Marcus Berger, *U.S. Census Bureau*
Jessica Holzberg, *U.S. Census Bureau*

In many online surveys, respondents are first asked to log into the survey instrument. A log-in screen that is cluttered or confusing may lead to respondents getting frustrated, taking a long time to log in, or abandoning the survey task entirely. Lack of clarity about how to switch the language of the survey when it is available in multiple languages can

also present challenges. Designing a usable login screen is critical for reducing respondent burden and increasing response rates. In this presentation, we discuss the redesign and usability testing of the login screen for the American Community Survey (ACS). The redesign used a more conventional layout to better conform to the standard of reading left to right, such as left aligning text and switching the location of the User ID login fields and the image of where to locate the User ID. The redesign also attempted to focus the user on the login task by removing extraneous text and highlighting the language toggle for users who would prefer to respond in Spanish. We conducted usability testing in English and Spanish to compare the current design and the newly developed design and obtain cognitive feedback on user expectations for this type of screen. Results indicate that the new design was more efficient for both English and Spanish speakers, but highlighting the Spanish language toggle in the new design did not have as noticeable an effect as anticipated. English speakers tended to favor the new version and Spanish speakers tended to favor the current production version. As a result of this testing, the team recommended moving forward with the new design with additional alterations to address concerns raised by the Spanish-speaking participants. We discuss the process of the redesign and usability testing and offer general suggestions for improving online survey login screens.

Developing a Self-Completion Version of the European Social Survey: Results From an Experiment in Austria

Tim Hanson, *City University London*
Professor Rory Fitzgerald, *City, University of London (ESS ERIC)*

Face-to-face surveys have become more expensive in many European countries alongside declining response rates and increased concerns about the influence of negative interviewer effects. As a result, the status of face to face surveys as the gold standard for survey research in Europe has been increasingly questioned with some now replaced by self-completion surveys instead. However, self-completion surveys face their own challenges such as lower response rates, difficulties with within household selection and concerns about administering long surveys. These debates about different data collection modes have been brought into greater focus due to the COVID-19 pandemic. The pandemic meant that face-to-face, in-person, data collection was not possible in the short-term. There are also questions of whether there may be a longer-term impact on the feasibility of face-to-face fieldwork. The pandemic, and its impact on face-to-face fieldwork, has presented particular challenges for surveys that usually reply on face-to-face fieldwork. This includes the European Social Survey (ESS), which has developed and tested a self-completion version of the survey for the first time. This paper addresses two specific questions related to fielding self-completion (web and paper) surveys drawing on a push to web experiment based on the ESS questionnaire, conducted in Austria in 2021. First, it looks at the feasibility of fielding longer surveys using self-completion instruments by testing a survey length normally reserved for face-to-face surveys and comparing it to a shorter version of the same survey. Second it addresses the question of whether a higher conditional incentive can increase participation and data quality. The results suggest that good response rates can be achieved with a long questionnaire (c. 50 minutes) in self-completion format. However, the results raise some questions over the relative exclusion of population sub-groups (versus the face-to-face method) and higher rates of non-response for certain items.

Data Collection During the COVID-19 Pandemic: Pivoting to Remote Methods in 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*
Nathaniel Poland, *NORC at the University of Chicago*
Andrea Malpica, NORC at the University of Chicago
Michael Kogan, HRSA / Maternal and Child Health Bureau

The Maternal and Child Health (MCH) Jurisdictional Survey collects critical data on the health and well-being of mothers, children, and families in the eight U.S. jurisdictions (Puerto Rico, U.S. Virgin Islands, Palau, Marshall Islands, Guam, American Samoa, Northern Mariana Islands, and Federated States of Micronesia). Between May 2019 and February 2020, data was collected face-to-face in each jurisdiction. The Maternal and Child Health Bureau (MCHB) and NORC at the University of Chicago had begun planning for the next round of data collection when, in March 2020, plans

for face-to-face data collection were largely put on hold around the globe. During the COVID-19 Pandemic, many face-to-face surveys transitioned to remote data collection. However, the MCH Jurisdictional Survey could not due to sampling constraints and we needed to rethink of plans and methodologies to allow enumerators and respondents to stay safe during face-to-face data collections. These changes included shifting field team trainings from in person classroom session to online sessions over Zoom, adjustments to quality control and field management procedures intended to limit contact between field team members and the public, and the adoption of CDC and WHO recommendations to mitigate the risk of disease transmission between field team members and respondents. Data collection successfully resumed in June 2021 in American Samoa and Palau with these new processes in place. While neither location had experienced community spread of COVID-19, fear of the virus among residents made safety protocols critical. Lessons learned from these first two data collections were incorporated into plans for additional data collections in Federated States of Micronesia, Commonwealth of Northern Mariana Islands, Marshall Islands, and Guam beginning in fall 2021. This paper presents lessons applicable to future data collections in the Pacific Basin and to international locations where shifting to phone or online methodologies is not possible

Crafting a Respondent Centered Informed Consent Process
Seth Messinger, *Ipsos - Seattle, WA*
Nina Seminara, *Ipsos US Public Affairs*

Research within conflict zones, low security areas, and among migrants moving through such zones is critical if, public opinion is to be measured, social conditions are to be evaluated and mitigation strategies for those caught up in these chronic states of emergency are to be implemented. Yet researchers face two main barriers to conducting their work, which often lays the foundation for aid programming. One barrier is finding ways to work with affected populations in a manner that is considered, and experienced as, safe for the research team and the participant population alike. The second barrier is one that is unanticipated: the bureaucratic and regulatory ethical framework that governs research activities and seeks to ensure that human research participants are treated humanely with justice, beneficence, and a concern for persons. Too often regulatory bottlenecks prevent granting approval for this kind of work and Institutional Ethical Review Committees err on the side of caution and prevent research activities from proceeding. Here we describe an approach that has received international agency approval and is proving effective with local research populations. In addition this approach builds capacity through convening and supporting local ethical review committees in order to be able to advance projects that takes the pillars of ethical conduct seriously and allows research to progress.

Ongoing & Updating: Adding a mode to ongoing & accepted survey designs
An Experimental Evaluation of Changing to a Mixed Mode Design in the National Study of Caregiving
Mengyao Hu, *University of Michigan*
Vicki A. Freedman, *Survey Research Center, Institute for Social Research, University of Michigan*
Justin Kamens, *Westat*
Shawn McCloskey, *Westat*
Mick P. Couper, *Survey Research Center, Institute for Social Research, University of Michigan*

Limited experimental research has explored in a national context how the transition from telephone to a mix of web and telephone modes affects response rates (RR), field effort, and potential bias in estimates. We extend this research by experimentally evaluating a change from a phone-only to mixed mode design for a national follow-back study of caregivers to older adults. The National Study of Caregiving (NSOC) interviews family and unpaid individuals assisting older adults in the National Health and Aging Trends Study (NHATS). Prior NSOC rounds (e.g., 2015, 2017) were by telephone and most NSOC interviews were completed within one month of NHATS. In 2021, NHATS embedded an experiment so that half of NSOC-eligible caregivers were randomized to the usual telephone-only design and the other half to a sequential mixed-mode design offering web and telephone options. Contact information has been obtained for approximately n=2,700 NSOC-eligible individuals through October 2021 (n=3,000 expected when fieldwork is completed by the end of 2021). Preliminary results show a higher predicted RR at completion for the mixed-mode design (65% RR; about half online) relative to phone-only in 2021 (59%) and prior rounds (60%). Among cases randomly assigned to sequential modes, RR and % by web varied by available contact information. Those with a mailing and

email address (30%) had the highest response rates and % by web (79% RR; 79% by web) and those with no mailing address (15%) had the lowest rates (41% RR; 9% by web). Average days since NHATS was only slightly longer for the mixed mode sample (20.7 days; 84.3% <=30 days) relative to the phone-only sample (19.1 days; 86.1% <=30 days). We will also explore effects of changing to a mixed mode design on key NSOC estimates (including caregiver hours, relationship quality, income and wellbeing) and discuss implications for trend analyses.

Feasibility and Effectiveness of the Addition of In-bound IVR to Online-only Surveys
Ashley Amaya, *Pew Research Center*

About 7% of U.S. adults do not use the internet, 16% are not digitally literate, and half cannot read above an eighth-grade level. These attributes make it difficult to participate in self-administered online surveys. Moreover, non-internet and less literate individuals are disproportionately likely to be older, have less formal education and live in rural areas than their counterparts, introducing the potential for bias. Meanwhile, a growing number of surveys can only be completed online. Some web-only surveys attempt to account for non-internet and less literate populations using weighting adjustments, while others provide internet access or an alternative mode of data collection to individuals without internet access. But these are imperfect solutions. Weighting assumptions may be faulty. Providing internet access does not address literacy challenges and may not successfully recruit individuals who consciously choose to be offline. And the introduction of some alternative modes may introduce interviewer effects or be infeasible due to timelines and budgets. To address these weaknesses, the Pew Research Center fielded a study in March 2021 to test the feasibility and effect of collecting data through inbound interactive voice response (IVR) in addition to the internet. An invitation was mailed to 10,000 addresses and gave individuals the choice of completing the survey online or dialing a toll-free number to respond via IVR. The study yielded 1,332 completed interviews, 1,250 via web and 82 via IVR. Ultimately, IVR respondents were different from web respondents and better represented groups typically underrepresented by online panels. Unfortunately, the proportion of IVR respondents was too small to meaningfully shift overall estimates. Moreover, while the introduction of IVR appears a potential mode to improve representation in online-only surveys, additional research is required to overcome logistical and data quality challenges and further increase response.

Not By Design: How Interviewers Administer In-Person Show Cards during Telephone Interviews
Alexis Kokoska, *Westat*
Danielle Mayclin, *Westat*

Like many in-person surveys, data collection procedures for the Household Component of the Medical Expenditure Panel Survey (MEPS) needed to be adjusted due to the COVID-19 pandemic. MEPS was designed to be administered in-person, with interviewers providing hard copy show cards to respondents at various items. MEPS primarily relies on show cards for items with long or technical response sets, which may be difficult for a respondent to remember without a visual aid, and time-consuming for an interviewer to read aloud. Show cards are also used to provide examples that might assist the respondent in selecting an accurate response. Starting in April 2020, MEPS introduced telephone interviewing, which meant that interviewers were no longer able to provide physical show cards to respondents in person. Instead, respondents could access the show cards online or, if they were unable or unwilling, the interviewer could read the show card text aloud to the respondent. Because show cards are not designed to be read aloud, such administration is likely difficult for interviewers and respondents. In this presentation, we analyze interviewers' behaviors in reading show cards aloud during telephone interviews. In particular, we assess if interviewers are more likely to read the text placed at the beginning as opposed to the end of the show card; how often they read the show card text verbatim; if their behaviors in administering the show card depend on amount of text; and if their administration behaviors change over time within an interview. Additionally, we consider whether interviewer administration of show cards impacts data quality. As MEPS continues to be a multi-mode data collection for the foreseeable future, this analysis may highlight opportunities for revising telephone procedures, enhancing training, or redesigning complex questions

From One, Many: Hatching a Multi-Mode, Multiple-Respondent Supplement via a Household Interview
Jessica Novik, *Westat*
Darby Steiger, *Westat*
Angie Kistler, *Westat*

Sandra Decker, *Agency for Healthcare Research and Quality (AHRQ)*

The Agency for Healthcare Research and Quality's (AHRQ's) Medical Expenditure Panel Survey (MEPS) provides nationally-representative data on health care expenditures, health care usage, and household characteristics. For the Household Component (HC), Westat interviews annual panels of about 10,000 households 5 times over 2.5 years. MEPS-HC data are collected (via CAPI or CATI) with one individual on behalf of the household, as well as a yearly SAQ for adult householders. Recognizing that social and economic conditions of a household are important predictors of health status and health care usage, AHRQ sought to add another self-adminsterered questionnaire (SAQ) to measure social and behavioral determinants of health (SDOH). In 2020-21, Westat designed and launched this SAQ as a multimode instrument (web and paper) that aimed to protect privacy due to the sensitive nature of some of the items. The multimode data collection for the MEPS SDOH SAQ occurred during the first half of 2021, followed by paper survey nonresponse follow up during the second half of 2021. This paper will review the policy goals behind the SDOH, the protocol that was designed to encourage web response, data collection challenges during the coronavirus pandemic, and our experiences with both modes of data collection, including overall and item non-response rates.

Results from the National Health Interview Survey Follow-Up Health Study: Innovation in Data Collection

Results From the National Health Interview Survey Follow-Up Health Study: Innovation in Data Collection

Grace E. Medley, *National Center for Health Statistics*
Benjamin Zablotsky, *National Center for Health Statistics*
Adena M. Galinsky, *National Center for Health Statistics*
Antonia J. Warren, *National Center for Health Statistics*
Duong T. Nguyen, *National Center for Health Statistics*

As part of CDC's Data Modernization Initiative, the National Center for Health Statistics (NCHS) has been exploring novel ways to supplement current data collection. This panel focuses on one such initiative, the National Health Interview Survey Follow-up Health Study. From June – October 2021, National Health Interview Survey (NHIS) and National Health and Nutrition Examination Survey (NHANES) conducted a pilot study to examine the feasibility of collecting respondents' height, weight, waist circumference, and blood pressure, administering a venous blood draw, and collecting a urine sample from adult NHIS respondents in their homes. Such data collection would enhance NCHS's health surveillance capabilities by increasing the utility of NHIS data.   This panel consists of five presentations, each covering a different aspect of the study. The first presentation will describe the study design and field operations in addition to examining cooperation and response rates by demographic subgroups.   The second presentation will compare results from the current study to results from a 2019 study where respondents were asked about their willingness to participate in a hypothetical biomeasures study. Differences in respondent characteristics (sociodemographic and health characteristics) associated with willingness to participate will be explored through logistic regression modeling.   The third presentation examines survey responses that may predict noncontacts to follow-up studies and may predict eventual participation among those who initially avoid speaking with follow-up study staff.   The fourth presentation will explore the top three reasons for refusal and the sociodemographic, health, health care access and service utilization, and paradata predictors of each of these three reasons using bivariate and multivariate analyses.   The final presentation will compare NHANES and NHIS Follow-up Health Study data collection, laboratory processing, and procedures for reporting results to respondents.

So You Want to Collect Biomeasures from Survey Respondents: NHIS Follow-Up Health Study Pilot

Grace Medley, *National Center for Health Statistics*
Antonia J. Warren, *National Center for Health Statistics*
Adena M. Galinsky, *National Center for Health Statistics*
Maria A. Villarroel, *National Center for Health Statistics*
Benjamin Zablotsky, *National Center for Health Statistics*
Aaron K. Maitland, *National Center for Health Statistics*
Duong T. Nguyen, *National Center for Health Statistics*

The National Health Interview Survey (NHIS) and National Health and Nutrition Examination Survey (NHANES) are both principal sources of information on the nation's health, however they collect different kinds of data with different survey designs. NHIS's large nationally representative sample can be used to calculate national and sub-national estimates of health conditions and health care access and utilization but is limited to data collected through an in-person survey. NHANES, in addition to an in-person survey, also includes a comprehensive physical examination that collects both physical measurements and biospecimens on a smaller sample. One way to enhance the National Center for Health Statistics' health surveillance capabilities is to increase the utility of NHIS data by adding physical measurements and biomeasure collection in a home setting.   From June – October 2021, NHIS and NHANES conducted a pilot study to examine the feasibility of collecting respondents' height, weight, waist circumference, blood pressure, and resting heart rate, obtained a venous blood sample, and collecting a urine sample from adult NHIS respondents in their homes. Adult respondents who completed the NHIS in English and resided in the selected sample areas were eligible to participate in the pilot study. Eligible Adults were introduced to the study at the end of their NHIS interview and those who agreed to be contacted could schedule the home health exam in the weeks following their interview. The home health exams were conducted by phlebotomists and participants received a $75 prepaid card and the results of all measures collected. This presentation will describe the study design and field operations in addition to examining cooperation and response rates at each stage by demographics. It will also offer insight into the challenges of collecting biomeasures during COVID-19, lessons learned, and implications for the future of physical measurements and biomeasure collection on the NHIS

When Expectation Meets Reality: A Comparison of Respondent Willingness to Participate in a Hypothetical Versus Pilot Follow-up Study to the National Health Interview Survey

Benjamin Zablotsky, *National Center for Health Statistics*

Adena M. Galinsky, *National Center for Health Statistics*

Aaron Maitland, *National Center for Health Statistics*

Grace E. Medley, *National Center for Health Statistics*

Maria A. Villarroel, *National Center for Health Statistics*

Antonia J. Warren, *National Center for Health Statistics*

The National Health Interview Survey (NHIS) has been monitoring the health of the United States for over 60 years on a broad range of health topics through personal household interviews. In recent years, NHIS has begun exploring additional ways health data may be captured beyond structured health interviews, including supplementing interview data with biomeasures, which have the potential to provide estimates of undiagnosed conditions and biomarkers of the population by demographic subgroups. This talk will discuss two studies conducted among NHIS adult respondents concerning the use of biometrics. In the first, survey respondents in 2019 were asked about their willingness to participate in a series of hypothetical studies, including providing body measurements (e.g., height, weight) and a blood sample in their home. The second, a pilot study launched in 2021, assessed the feasibility of collecting body measurements and biospecimens in the respondent's home by a trained phlebotomist, following the completed NHIS. There is notable value in comparing results from these two studies to determine if hypothetical scenarios are in fact predictive of participants' actual willingness to participate in a study. In 2019, approximately 35% were willing to participate in a hypothetical in-home examination. Among those willing to participate, approximately 70% would provide both body measurements and a blood biospecimen, while 25% would only provide body measurements. In 2021, preliminary results show that approximately 30% of respondents were willing to be contacted by a scheduler and about half completed the in-home examination. Nearly all participants provided both body measurements and blood samples.   This presentation will explore respondent characteristics associated with willingness to participate in each study through logistic regression modeling, accounting for the respondents' sociodemographic and health characteristics. The presentation will conclude with lessons learned with a focus on the value of asking respondents about hypothetical studies.

Predictors of Noncontact and Participation Following Agreement to Participate in a Follow-up Study

Adena Galinsky, *National Center for Health Statistics*

Grace E. Medley, *National Center for Health Statistics*

Maria A. Villarroel, *National Center for Health Statistics*

Antonia J. Warren, *National Center for Health Statistics*

Benjamin Zablotsky, *National Center for Health Statistics*
Duong T. Nguyen, *National Center for Health Statistics*
Aaron K. Maitland, *National Center for Health Statistics*

Noncontacts are increasingly becoming the primary challenge of not just survey research but all in-person and phone population research. Results from a recent study, conducted by the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Surveys (NHANES), offers some insights into what differentiates noncontacts from those who speak to study staff in order to refuse or agree to participate, what differentiates noncontacts who ultimately participate from those who do not, and what strategies work best to convert noncontacts. From June – October 2021, adult respondents who completed their NHIS interview were invited to participate in a follow-up study. They were told that if they agreed, a scheduler would contact them about making an appointment for a phlebotomist to visit their home to measure their height, weight, waist circumference and blood pressure; obtain a venous blood sample and collect a urine sample; and give them a prepaid card and their results. In the weeks that followed the NHIS interviews, study schedulers attempted to call those who agreed to schedule the appointment. A series of strategies designed to reach respondents who did not answer the phone included texts, emails, physical mailings, and calls from specially trained conversion specialists. This presentation examines survey responses that may predict noncontacts to follow-up studies and may predict eventual participation among those who initially avoid speaking with follow-up study staff. We also examine whether the mode and type of strategy that was implemented immediately before the contact that resulted in a scheduled appointment among noncontacts differed by respondent characteristics, and how the duration between the interview and the initial contact attempt relates to participation. The results from this analysis may be useful in the development of conversion efforts for follow-up studies.

Why are Survey Respondents Refusing a Free Health Exam at Home?

Antonia Warren, *National Center for Health Statistics*
Adena M. Galinsky, *National Center for Health Statistics*
Maria A. Villarroel, *National Center for Health Statistics*
Grace E. Medley, *National Center for Health Statistics*
Benjamin Zablotsky, *National Center for Health Statistics*
Aaron Maitland, *National Center for Health Statistics*
Duong T. Nguyen, *National Center for Health Statistics*

The National Health Interview Survey (NHIS) is an integral source of information on the health of the United States population. It is used to calculate quarterly and annual national health estimates for geographic, demographic, and socioeconomic subgroups. However, the ability of the NHIS to monitor overall health of the population could be enhanced by adding a health exam with physical measurements and biospecimen collection, thereby enabling validation of interview data and estimation of the prevalence of undiagnosed health conditions. To test the feasibility of such an addition, the National Center for Health Statistics conducted a pilot study between June and October 2021, in which a portion of adult respondents from households in selected sample areas who completed the NHIS Adult interview in English were offered a home health exam. One of the main study objectives was to learn about respondents' reasons for refusing to participate in the health exam. Gathering this information would lead to the development of better strategies to address respondent concerns and improve study recruitment. NHIS interviewers asked NHIS respondents the reasons for refusal using an open-ended question. Initial categorization of the responses revealed several themes. This presentation will explore the top three reasons for refusal: lack of interest (44.7%), lack of time (26%), and privacy concerns (16.5%) and the sociodemographic, health, health care access, service utilization, and paradata predictors of each of these three reasons using bivariate and multivariate analyses. Understanding which factors predict common reasons for refusal can inform the planning of similar studies to improve response rates, customize study materials to better address respondents' concerns, and in training interviewers in describing the study to respondents.

Biospecimen Collection in Nationally Representative Population Surveys: A Comparison of NHANES and NHIS Methods

Duong Nguyen, *National Center for Health Statistics*
Adena M. Galinsky, *National Center for Health Statistics*

Aaron K. Maitland, *National Center for Health Statistics*
Grace E. Medley, *National Center for Health Statistics*
Maria A. Villarroel, *National Center for Health Statistics*
Antonia J. Warren, *National Center for Health Statistics*
Benjamin Zablotsky, *National Center for Health Statistics*

The National Health and Nutrition Examination Surveys (NHANES) collects nationally representative health examination and biospecimen data in standardized mobile examination centers (MECs). NHANES collects data on approximately 5,000 participants annually but the sample size does not allow for the calculation of annual or sub-national estimates. In comparison, the National Health Interview Survey (NHIS), which interviews approximately 27,000 adults annually, can be used to calculate quarterly and annual national health estimates for geographic, demographic, and socioeconomic subgroups. However, NHIS does not currently collect examination or biospecimen data. NHANES collaborated with NHIS to explore the feasibility of collecting objective exam data in respondent's homes and providing their results as part of the NHIS. From June - October 2021, a sample of adult NHIS respondents were invited to join this feasibility study. Respondents who agreed to be contacted had the opportunity, after their NHIS interview, to schedule a home health visit. During that visit, a phlebotomist measured the participant's height, weight, waist circumference, blood pressure, and resting heart rate. The phlebotomist also obtained a venous blood sample and asked the participant to provide a urine sample. Final NHIS protocols for collecting physical measurements and biospecimens differed from those used by NHANES. Variation existed in terms of equipment used, staffing models, laboratory processing schedules, reporting respondent results, incentive levels, and other factors. A comparison of NHANES and NHIS data collection, laboratory processing, and report of findings procedures will be presented. Survey-specific restrictions on reporting, component-level cooperation rates, and participant concerns and feedback will be noted.

Experiments in Using Incentives to Boost Response Rates
Assessing the Impacts of Different Incentives and Use of Postal Mail on Response Rates
James Ellis, *Virginia Commonwealth University*
Julie Charbonnier, *Virginia Commonwealth University*
Cydney Lowenstein, *Virginia Commonwealth University*
Anne Buford, *Virginia Commonwealth University*
Daniel Davis, *Virginia Commonwealth University*

The Survey and Evaluation Research Laboratory at Virginia Commonwealth University (VCU-SERL) conducted a survey of recent college graduates about the impacts of undergraduate education. The sampling frame contained cases for which a commercial matching service returned postal mail addresses. For some of those cases, email addresses were returned. The survey was long and the scale would be large; therefore, VCU-SERL conducted a pilot survey (sample n = 3,648) to assess ways to maximize survey participation and optimize costs. All cases received up to four mailings. Those with email addresses received up to five emails. VCU-SERL tested up-front cash ($1 or $2), gift cards contingent on survey completion (no offer, $5, $10, and $20), a prize drawing contingent on survey completion, and various combinations of mailing materials. The pilot survey, conducted March to June 2020, was available via Qualtrics and paper. There were 1,019 respondents (28% response rate). The gift card was the strongest influence on response rates. Both the $20 and $10 gift cards generated statistically significant effects (37% and 31% response rates, respectively) compared to the $5 and no offer options (24% and 22%, respectively). These and other smaller effects informed the plan for the production survey. The production survey ran from December 2020 to May 2021 with a probability sample (n = 50,495) that received multiple postal mailings and emails, $1 up-front cash, $10 contingent gift card and a contingent prize drawing. There was also an email-only supplement without any postal mail communication (n = 52,173). There were 13,294 probability sample completions (26% response rate) and 2,054 email-only completions (4% response rate). The production survey response rate closely matched the rate predicted by the pilot survey. This work demonstrates the utility of pilot testing and the importance of postal mail recruitment in some research contexts.

Effect of Incentive and Length of Online Surveys on Drop-Off Rates: An Examination of Undergraduate College Students

Rebecca Oh, *EEOC*
Matin Mirramezani, *The Generation Lab*
Cyrus Beschloss, *The Generation Lab*

Numerous studies have suggested that survey drop-off rates rise as survey length increases. In particular, literature points to an ideal median of 10-minute polls in the digital space before drop-off becomes significantly prevalent. In this study, we expanded on the findings from the previous analyses to examine the relation between length and incentive in web-based surveys. Using the Generation Lab's comprehensive frame of college students in the United States, a random sample of 20,000 students was selected and randomly assigned into one of two survey groups prior to email contact. Group A contained 10,000 students who were asked to complete a 10-minute survey of 30 questions for a direct incentive of $4. Likewise, Group B contained 10,000 students who were asked to complete a 5-minute survey of 15 questions for a direct incentive of $2. The 15 questions assigned to each respondent in Group B were randomly generated from the pool of 30 questions given to Group A. To explore the combined survey length and incentive effects, we investigated overall response rates for each survey group and conducted a chi-square test for independence to determine any significant response rate differences across the two survey groups. We repeated this procedure specifically using proportions of respondents who have started their respected survey but never completed it. Furthermore, in order to uncover underlying factors of survey drop-off, we explored various demographic segmentation, such as college type and category of college degree. This study informs real-life decisions that survey professionals have to make regarding the trade-off between survey length and incentive.

Money Talks: The Effects of Monetary Incentives on Earnings Response in SIPP

Shalise Sarah Ayromloo, *U.S. Census Bureau*
Kelly R. Wilkin, *U.S. Census Bureau*

The Survey of Income and Program Participation (SIPP) has a history of using conditional and discretionary monetary incentives to induce survey responses. While incentives have been effective in increasing unit response, little is known about their effect on item response. This paper exploits a multi-wave random monetary incentive experiment for the SIPP 2014 panel to examine the effect of incentives on earnings non-response. We show that individuals in incentive-eligible households have a 1.3-percentage-point lower earnings non-response rates than those in incentive-ineligible households. This effect is robust to controls for observed and unobserved individual heterogeneity and non-random panel attrition in a correlated random effects specification. Further, we find the effect is driven by a $40 incentive assignment and not the $20 incentive. Consistent with theories linking unit and item non-response, we find that contemporaneous earnings non-response is associated with a higher probability of attrition in the following wave, but the $40 incentive mitigates this relationship.

For Whom the Buck Tolls: Impact of Incentive-motivated Respondents on Survey Estimates

Allison Zotti, *U.S. Census Bureau*
Maura Spiegelman, *National Center for Education Statistics*

Decreasing response rates and increasing data collection costs have forced survey organizations to consider new data collection strategies (Groves and Heeringa, 2006). Incentives have been found to be effective in increasing response rates for most surveys (Mercer et al. 2015). However, most analysis of incentives focuses on response rates, costs, or sample composition. For its 2017-18 data collection cycle, the National Teacher and Principal Survey (NTPS) explored the use of prepaid incentives in an effort to increase teacher response rates and overall sample balance. This experiment was designed to test the effectiveness of small cash incentives sent to teachers. Incentives increased response rates, reduced the time in which sample members returned their completed questionnaires, and increased the representativeness of respondents for characteristics that are part of the sampling frame and known for both respondents and nonrespondents. However, incentives may be particularly effective in circumstances that can't be disentangled from survey estimates. For example, the NTPS asks teachers about their base salary, additional jobs, and other sources of income. This presentation investigates what impact incentives have on teachers with different levels of income and financial supplements. If an incentive influences teachers with lower or higher incomes to respond at different rates, this could lead to bias in survey estimates. Other teacher characteristics, such as subject area or years of experience, will also be analyzed to determine if certain types of teachers more influenced by incentives than others.

Political Polarization, Ideology, & Impact on Presidential Elections

The Impact of Authoritarian Attitudes on Presidential Elections

Robert Oldendick, *University of South Carolina*
Chandler Case, *University of South Carolina*

The Presidency of Donald Trump has sparked a renewed interest into the role that authoritarian attitudes play in American politics. Using data from the 2012, 2016 and 2020 American National Elections Studies, this research investigates the extent to which this trait is evident in the American public during this period. This work also examines the correlates of authoritarianism, particularly describing the extent to which background characteristics such as age, race, gender, education, income, and region are associated with this trait. In addition, it examines the relationship that authoritarianism has with other attitudes, such as political efficacy, trust in government, political knowledge, questions of race, and social issues. It concludes with an examination of the effect of authoritarian attitudes on different types of political participation and ultimately its influence on Presidential vote in these elections. Results indicate that there has been a rise in authoritarian attitudes in the mass public during this period, that this increase has occurred differentially among various subgroups, and that such attitudes are significantly related to several political attitudes, including vote choice.

Commonality and Polarization: Results of a Conjoint Experiment

Robert Petrin, *Ipsos Public Affairs*
Nicolas Boyon, *Ipsos Public Affairs*
Justin Gest, *George Mason University*

No week goes by without a new poll highlighting the pervasiveness of political and ideological polarization in the United States. This raises the question of how much party preference now defines Americans' identity relative to other individual and group characteristics; that is, how much more or less likely are Democrats and Republicans to identify with someone based on their party affiliation than they are to identify with someone based with on their faith, their ethnic background, their nationality, their level of education, or the type of area where they live?   This paper presents results from a national conjoint experiment conducted among 3,500 US adults. Each respondent was shown a series of 10 comparison exercises.   Each exercise presented respondents two separate, randomly-chosen profiles constructed from six broad sets of demographic characteristics, and asked them to select which of the two profiles they thought they had more in common with ("Do you think you have more in common with Person A or Person B").   The data are analyzed using the methods of Hainmueller and Hopkins (2015) and Hainmueller et al. (2013) to estimate average marginal causal effects (AMCEs) and multidimensional composites of those individuals members of particular demographic groups feel to be the most "like them".   Overall, we find that political affiliation plays a strong role in individuals' affiliation with others; a role that is on par with the role religion played in previous eras.   We also find substantial variation in out-group affiliation by racial/ethnic and urbanicity.

Political Opinions Among a Polarized Public

Samara Klar, *University of Arizona*
James Druckman, *Northwestern University*
Yanna Krupnikov, *Stony Brook University*
Matthew Levendusky, *University of Pennsylvania*
John Barry Ryan, *Stony Brook University*

When the public is so politically polarized, partisan gaps seem to exist across a huge array of policy issues--even when the issues might not seem particularly political. To understand when and why we can expect partisan gaps in public opinion, we administered a panel survey of over 3,000 Americans covering a two-year span (beginning before the 2020 Presidential Election and the COVID-19 Pandemic, and ending in the spring of 2021). Our findings are three-fold: first, most Americans are not affectively polarized and do not display large partisan gaps in opinions and behavior. But about a third of Americans do express intense disdain for the out-party and it's among this group where we find sharp distinctions in attitudes regarding both politics, public health, and more. The survey data support our theory that public

opinion and political behavior will diverge along party lines among the most polarized only when the issue has become politicized by party elites.

Next-Door Neighbors and Affective Polarization
Jennifer Lin, *Northwestern University*

Republicans and Democrats alike prefer living in areas where co-partisans are more prominent than out-partisans. With limited exposure to out-partisans in daily life, and even less in online contexts, there is little opportunity for people to become exposed to opposing ideas and preferences. This research contributes to the work on partisan neighborhood sorting by looking at preferences for specific next-door neighbors, who are people with whom one has the greatest probability of contacting and interacting with each day. Moreover, this project contributes to research in affective polarization in the context of out-partisan interactions. I introduce two novel experiments, both designed to test preferences for neighbors and subsequent interactions. The first looks at renting apartments, and assumes that apartments are in urban areas and the rents are for short term commitments. Participants are shown different options of places for rent, which include information on the cost, amenities, community and neighbors of potential apartment homes. Participants are then asked to select one of two options. This tests the priorities that people give to finding a place to live, especially whether people prioritize community conditions or the people in the communities. To expand on this, I introduce a second experiment where participants are asked to imagine that they are buying a home. This design introduces long term commitments and options to live in urban, suburban or rural areas. Participants will be shown a house, along with the neighborhood qualities and close neighbor personalities, and will be asked to put the amount that they are willing to pay for it. I hypothesize that people with higher partisan animosity would put greater weight on neighborhood partisanship when considering where to live. Furthermore, people will prefer neighbors who are counter-stereotypical to the out-party if they cannot live next door to people of their own party.

A Meeting Place & More: A Panel on the History of AAPOR
A Meeting Place and More: A Panel on the History of AAPOR
Joerg Tiede, *American Association of University Professors*

Harry H. Field, the Father of AAPOR
Tom Smith, *NORC at the University Of Chicago*

**NOTE THIS PAPER IS BEING PRESENTED AS PART OF THE AAPOR HISTORY ROUNDTABLE AND SHOULD NOT BE SLOTTED INTO ANOTHER SESSION. Harry H. Field was the founder of the National Opinion Research Center (NORC) (1941) and father of AAPOR (1946). Having immigrated to America from England in 1919, he worked in publishing, journalism, advertising, and as an author before turning to public opinion research in the mid-1930s. He established Gallup's first international affiliate in Great Britain in 1936 and later assisted in adding Gallup affiliates in France and Australia. In 1941, he established NORC at the University of Denver as the world's first university-based. non-profit survey research organization. He organized the Central City Public Opinion Research Conference in 1946 that directly led to the founding of both AAPOR and WAPOR. While seeking to further expansion of public opinion research in Europe in the aftermath of World War II and secure UNESCO support for international polling, he died in an airplane crash in France in 1946

Are You For Real? Detecting Fraudulent & Poor-Quality Respondents
How Well Do Attention Checks Identify Fraudulent Online Responses, and Which Types of Attention Checks are Best?
Emily Geisen, *Qualtrics*
Stephen Poole, *Qualtrics*

Attention checks are a popular and easy method to identify and prevent fraudulent responses in online surveys, especially from online nonprobability panels. However, some research suggests that these checks do not identify all types of fraudsters (Kennedy et al., 2020). Other attention checks may inadvertently flag legitimate survey responses as fraudulent because they are too difficult. In addition, many researchers are concerned that even if valid, attention checks disrupt the flow of the survey and reduce participant trust. In this presentation we explore whether different types of attention checks are more effective at accurately flagging fraudulent responses.   To explore this, we surveyed

4,000 respondents from five online nonprobability panel providers. Respondents were randomly assigned to one of four different treatment groups (commitment, factual, speed flag, open-ended response) or the control group (which did not receive an attention check). We then created a composite quality index comprising more than a dozen other quality indicators including both digital fingerprinting technology (which uses IP address and online behavior to identify bad actors) as well as straightlining, inconsistent responses, speeding, and a review of open-ended responses.   For each attention check, we will compare the average composite quality score for respondents who failed the attention check with respondents who passed. We will also look at the rate of false positives (responses we deemed valid based on the composite score that were flagged by the attention check). Next, we will compare the composite quality score of respondents who passed the attention checks with the control group to assess the impact of having an attention check on overall quality.   We will conclude with a discussion about whether attention checks can be a useful tool in identifying fraud, whether they should be combined with other quality indicators, and which type of attention checks are most accurate.

How Fast is Too Fast? Identifying Speeders in Survey Data
D. Sunshine Hillygus, *Duke University*

Inattentive or careless respondents can jeopardize the reliability and validity of survey estimates. One common approach for identifying these low-quality respondents is to check survey response time for "speeders." Respondents who answer quickly are assumed to not have not given much thought to their answers. Although the concept of speeding is straightforward, the challenge is defining how fast is too fast in practice. The optimal response time depends on the reading skills of the respondent as well as the type and difficulty of the survey questions being asked. With these considerations in mind, researchers need to balance the risk of flagging too few inattentive speeders with the risk of flagging valid, but quick respondents. Speeding metrics vary widely in practice, with response times variously operationalized for a single question, sets of questions, or the entire questionnaire. In this analysis we evaluate a variety of speeding definitions and thresholds using a 2020 YouGov survey that allows for multiple validity checks: test-retest comparisons, evaluations of validated vote, criterion validity checks with well-established attitudinal measures, and predictive validity checks with other data quality metrics. The goal of the project is to provide evidence-based research on speeding metrics to help determine how best to identify low-quality respondents.

Satisficing in Web Surveys: Testing Multiple Speeding Interventions in the American National Election Study
Jill Carle, *Westat*
Rosalynn Yang, *Westat*
Matthew DeBell, *Stanford University*
D. Sunshine Hillygus, *Duke University*
Ted Brader, *University of Michigan*
Michelle Amsbary, *Westat*

Survey respondents often take mental shortcuts and provide acceptable, but not optimal answers, often referred to as "survey satisficing" (Krosnick, 1991). This phenomenon is particularly challenging in web surveys where respondents complete surveys without interviewers present (e.g., Heerwegh and Loosveldt, 2008). Previous research suggests that prompt interventions for speeding and satisficing are effective at reducing these undesirable response behaviors in web surveys. However, little work addresses the impact of multiple speeding interventions on speeding and satisficing behavior within the same survey. Furthermore, while initial research suggests that speeding intervention efficacy may differ by respondent age and education level, this critical link remains under evaluated. Through a 2x2x2+1 factorial design experiment in the American National Election Studies (ANES) 2020 Time Series Study, we will test the impact of three variations on speeding interventions, including varied response time lengths, forewarning approaches, and intervention types. We will evaluate the impact of each intervention type on speeding and satisficing outcomes in addition to the cumulative impact of receiving multiple interventions over the course of a survey.

Evaluation and Comparison of Digital Fingerprinting Technology to Identify and Prevent Fraudulent Responses in Online Panel Surveys
Emily Geisen, *Qualtrics*

As online nonprobability panels become more widely used for survey research, it has become increasingly important to identify and prevent fraudulent survey responses. For many years, researchers have used a variety of strategies such as attention checks, straightlining behavior, inconsistent responses, review of open-ended responses and other methods to ensure quality responses. Although these methods can be beneficial at identifying many types of fraudulent responses, they are imperfect. For example, they can be time consuming to program (straightlining or inconsistency checks), to implement (review of open-ended responses), or they can be ineffective on certain types of surveys (speeding checks on surveys with complex branching). Furthermore, as fraudulent actors (e.g., bots and click farms) become more sophisticated, these quality checks may become less effective. As a result, we explored the extent to which newer digital fingerprinting technology (which evaluates IP address, tampering of URLS and other online behavior) can improve the ability to accurately identify fraudulent survey responses over traditional quality checks alone.   To explore this, we surveyed 4,000 respondents from five online panel providers and compared over 15 traditional data quality indicators (e.g., attention checks, straightlining, speed) with three digital fingerprinting technologies. First, we will compare the three digital fingerprinting technologies to see where they agreed, where they disagreed and why. Then we will compare the digital fingerprinting technology with traditional methods to understand where these methods overlap and where they differ.   We will then present a conclusion on the utility of digital fingerprinting technology, whether it is necessary to use both of these strategies, and highlight the areas of quality that still need improvement, even with the presence of both methods.

Improving Satisficer Detection in Large-Scale Surveys through Experimental Questionnaire Design
Arismita Guharay, *Civis Analytics*

The goal of this research was to determine if iteration and experimentation on questionnaire design could significantly improve the detection of satisficing behaviors for survey respondents in a nonprobability, or opt-in, web-panel sample. Previous Civis research on detecting satisficing behaviors in survey respondents had already shown that a combination of click-timing paradata and "trap" survey question responses were quite effective for detecting satisficing behaviors in domestic and international survey settings. However, determining whether we could further improve on our methods to increase the sensitivity of our satisficing-detection module was crucial to our team. Our goal was to minimize this form of error, correct any resulting biases in the data, determine the optimal list (and format) of questions to utilize to reduce false negatives and positives in satisficer-detection. Ultimately, could we iterate on our original satisficer-detection module to reduce measurement error and get closer to assessing the truth in future survey research? In order to answer this, Civis experimented with the question order of the satisficer-detection module, question text formatting and content, quality terminates for the worst offenders during the prescreen, and the inclusion of warning messages for offenders of satisficing. We examined how the various treatments affected the response distribution of the survey data and click-timing paradata, respondent propensity for choosing the trap response options, and whether there were any changes in associations for particular variables and outcomes of interest. Ultimately, we studied the variation in rates of satisficing across all of the experimental arms of the study to identify the optimal satisficing-detection module for online survey research; as well as determining how best to treat the satisficers during the post-processing to obtain the most accurate estimates from our survey data.

I Strongly Agree to Disagree: Scale Design, Rankings, Ratings, & Response Formats
When Neither Rankings Nor Ratings Will Suffice: Assessing the Paired Comparison of Values Method
Eric Plutzer, *Penn State University*

Among constructs measured by questionnaires, values are especially challenging. Theories of values presume a hierarchy among well regarded goals. Everyone values safety, longevity, community harmony and meaningful relationships, differing in their relative preferences. Rating scales often show non-differentiation because all stimuli are popular, leading Rokeach, Inglehart, Schwartz and others to utilize ranking methods Ranking, however, is cognitively difficult. Moreover, ranking data is unsuitable to many common analysis strategies because the sum and mean of the ranks are fixed, and the rank of any item is dependent on the ranks of the others. In response to these challenges, Bilsky and his colleagues developed a hybrid method of relative ranking (2008, 2015).   In this approach, values are presented in pairs and the respondent not only indicates a preference, but also rates the strength of that preference, eliminating the linear dependence among the ranks. This task is more realistic – allowing respondents to value one goal slightly

more than another (or much more). Early assessments of paired comparison of values (or PCV), were promising but the technique was abandoned because of the large burden it placed on respondents. Evaluating every combination of 20 terminal values would translate into 190 tasks. I revisit the method of paired comparisons using a list of nine democratic values. In the treatment, a planned missing design presents each respondent with a random subset of fifteen of the 36 possible pairings. Controls rate the same statements using the ESS's rating instrument. RQs: Comparing PCV to traditional rating scales: Which method produces more reliable measures of each of the nine values? Which method shows more evidence of satisficing. Which method results in measures that best predict criterion variables. Which method can best approach the dimensional structure expected by theory? Substantively : What elements of liberal democracy do Americans value most?

Towards Improving Measurement of Discrimination: Replication of Response Format Effects

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

With recent social dislocations and profound polarization of many groups, attention to prejudices and discriminatory behaviors has significantly increased. Accurately measuring experiences of perceived discrimination is important to track change and identify groups being impacted and areas needing to be rectified. The measurement of discrimination has often varied based on the type of discrimination, rendering comparisons across types of discrimination difficult, if not impossible. It has become apparent that consistent measurement for self-reported discrimination across various types of discrimination (race-ethnicity, gender, disability, etc.) is necessary. In our first study, we reported results of an experiment on discrimination measures using a probability-based sample. For our second study, we used opt-in non-probability sample, and sought to replicate the results of the first study. Dichotomous response formats (DRF) using yes-no responses typically have been found to yield higher prevalence results than multiple response formats (MRF – "Select all") across a wide range of topics, from donations, health conditions, and measures of importance. With 5,829 respondents from an opt-in sample, we randomly assigned respondents to two different response formats (DRF or MRF) and various recall periods ("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.). If a person indicated they experienced any discrimination, we asked how negative of an impact it had on them (measuring salience). While magnitude of reports differed somewhat from one sample to the other, the general pattern of results were replicated: 1) reports of discrimination were higher for the DRF than the MRF, and 2) gender discrimination was reported most followed by race-ethnicity discrimination.

A Comparison of MaxDiff, Top Choice, and Rating Scale Ranking Methods

Sam Gutierrez, *Momentive*
Timothy B. Gravelle, *Momentive*

Forced top-choice ranking and rating scales are the most commonly used ranking questions in determining hierarchy within attribute lists, seen in both public opinion surveys (most important issue polls), and market research contexts (product feature prioritization). MaxDiff is another survey method that claims to provide a more accurate, clearer, and less burdensome method to establish rankings between attributes by using implied preferences derived from multiple iterations of attribute comparisons. For example, instead of answering a single list of 15 attributes, a respondent instead chooses the 'best' and 'worst' attributes within multiple subsets of 4-5 attributes. This research aims to determine if MaxDiff indeed provides insights different from traditional ranking questions. Do the various ranking methods differ from each other in terms of ranking outputs, and if so, are there also differences across varying lengths of attribute lists? We will conduct a comparison both across and within three ranking methods (top choice, rating scale matrix, and MaxDiff), with each method also employing three different lengths of attribute lists (5, 10, and 15 attributes). Attribute lists will be identical across ranking methods and attribute list lengths. Each ranking method will be followed by an additional survey inquiring about length, and ease of use and understanding. 9 parallel online surveys (1 for each of the 9 comparison groups) will be conducted on the Momentive survey platform.

Understanding Acquiescence Bias in Online Survey Panels
Amy Obendorf, *Civis Analytics*

Acquiescence bias is the tendency for respondents to select "agree" on agree-disagree scales in surveys, regardless of the content of the question. The concept of acquiescence bias traces back to the psychological concept social desirability, which theorizes that people want to appear likeable to others, and are therefore more likely to respond with positive response options, even if the response disguises the respondents' true feelings. Acquiescence bias presents a threat to data quality in survey research because it can create a response bias that favors positive response options, which can ultimately produce false conclusions about groups of respondents who are more likely to acquiesce. The purpose of our study is to understand the prevalence of acquiescence bias in different types of survey questions. To do so, we placed a series of six opinion statements or questions in a web survey of adults in the US, along with six questions that contradict the first set of questions. We assigned an acquiescence score to each respondent, where they received a point for every instance of agreeing to both contradictory statements. We tested acquiescence scores across four different question types: 1) a matrix with a five-point Likert scale; 2) a binary agree/disagree matrix; 3) a multi-select list of opinions; and 4) a single-select series of questions. Acquiescence scores were compared across the questions to determine which question formats lead to the highest incidences of acquiescence. We found that respondents were significantly more likely to agree to contradictory statements when the questions were asked in agree/disagree formats. We also found important subgroup differences, highlighting that not all respondents are equally prone to acquiescence bias, reinforcing findings from existing literature. Controlling for other types of satisficing behavior does not completely eliminate this effect. This research demonstrates the importance of good question wording in questionnaire design.

Agreeable Research: Comparing Response Patterns in Agreement and Item-Specific Scales
Megan Hendrich, *Ipsos Public Affairs*
Randall K. Thomas, *Ipsos Public Affairs*

Humans are often taught that to get along, they must play along. As such, one type of response bias identified in survey research is acquiescence bias. Acquiescence bias describes the tendency for interview respondents to choose an 'agreeable' response option rather than a response option that more accurately reflects their viewpoints. Acquiescence bias is believed to be more prominent with agreement scales and less likely to occur with item-specific scales (e.g., Krosnick & Presser, 2010). However, one issue that has confounded tests comparing agreement and item-specific response formats is scale polarity (generally bipolar for agreement vs. unipolar for item-specific formats; see Dykema et al., 2021). We report the results of two studies—each with around 11,000 respondents—comparing the use of bipolar agreement and bipolar item-specific formats. Respondents were randomly assigned one of six behaviors and asked to indicate how much they like doing the behavior (using a bipolar liking scale) and how much they agreed with a statement about the behavior ("I enjoy doing X"). In the first study, the responses were horizontally arrayed. However, to control for the possibility that respondents would simply endorse the response in the same left-right position, the second study presented vertically arrayed responses. Across all behaviors, we found that the response distributions for the agreement and item-specific response formats were indistinguishable. We then examined the criterion-related validity of the scales (i.e., frequency of doing the assigned behaviors), and we found equivalent validity for each format. We conclude that most research finding higher rates of acquiescence bias for agreement scales has mistakenly conflated the findings for differences in scale polarity. Agreement scales appear to provide equivalently valid data as item-specific scales.

Rare & Diverse Populations: Subgroup Estimation & Comparisons
Estimating Sexual Orientation and Gender Identity Prevalence
Kristie Healey, *ICF International*
Ronaldo Iachan, *ICF*
Yangyang, *Deng*

This paper updates the methods ICF presented in 2019 for using logistic regression and multi-level models to estimate Sexual Orientation and Gender Identity (SOGI) status where data are only available for a subset of states. SOGI

questions are included in the Behavioral Risk Factor Surveillance System (BRFSS) as an optional module. Participation has increased from 19 states when the module was first implemented in 2014 to 32 states in 2020. As of 2020 approximately 95% of the US adult population lives in a state that asked the SOGI module at least once in the past three years. We use data from the 2018-2020 BRFSS to refine the models for estimating prevalence of transgender, bisexual, and/or gay/lesbian identity in states that do not use the module, and imputing identity for respondents who are asked but do not know or refuse to answer these questions. We explore whether the availability of more state level data allows for more demographics or state indices to be considered as predictors. In phase 2 of the analysis, state level data will be combined to produce estimates at the national and Census region levels, and validated against direct estimates from publicly available data sources such as the National Health Interview Survey (NHIS). We explore differences in LGBTQ prevalence as they relate to variations in health behaviors and risk factors, and begin to explore the feasibility of using LGBTQ identity estimates from large scale federal surveys as population controls in weighting, including whether adding sexual orientation to poststratification may improve the accuracy of health outcome measurements

Using a Combination of Probability and Nonprobability Sampling Strategies for NHPI and LGBTQ Health Studies

Todd Hughes, *UCLA Center for Health Policy Research*
Sean Tan, *UCLA Center for Health Policy Research*
Ninez Ponce, *UCLA Center for Health Policy Research*
Margie Engle-Bauer, *SSRS*

Measuring health indicators for historically underrepresented and underserved populations like Native Hawaiians and Pacific Islanders (NHPI) and Lesbian, Gay, Bisexual, Transgender and Queer (LGBTQ) individuals is important in developing programs to improve the health outcomes among these groups. However, survey sampling from these small population groups can be quite challenging, and may require the combined usage of multiple sampling approaches to obtain sufficient samples to produce reliable and representative estimates. In 2019 and 2020, the UCLA Center for Health Policy Research conducted two studies focused on NHPI and LGBTQ individuals in California, to measure attitudes and behaviors regarding smoking and tobacco use as well as various tobacco use cessation strategies for the California Tobacco Control Program. These two studies, called the California Pacific Islander Health Attitudes and the Behavior Study and the California Health Attitudes and Behavior Study, each started by re-contacting LGBTQ and NHPI respondents of the California Health Interview Survey (CHIS) to form a sampling frame. Additional sampling frames and other probability and non-probability methods were also used to various degrees, including Respondent-Driven Sampling (RDS), convenience sampling, and conducting follow-up interviews with respondents from multiple other population surveys. This paper will present the methods used, describe the effectiveness and challenges for each, and explore how the data were processed to release data for the NHPI and LGBTQ populations.

Address-Based Sampling for Jewish Community Studies

Zachary Seeskin, *NORC at the University Of Chicago*
David Dutwin, *NORC at the University of Chicago*
Leonard Saxe, *Brandeis University*

Jewish community studies, which provide estimates of the size and characteristics of Jewish populations in different metropolitan areas, provide critical information for Jewish organizations to inform planning and programming. Sampling procedures for studies of this rare population require thoughtful planning. Until recently, these studies have most often been conducted using random digit dialing, and address-based sampling for Jewish community studies is a new development having only been conducted for a few studies to date. We describe the efficiencies enabled by the address-based sampling approaches for the 2020 Metropolitan Chicago Jewish Population Study and the 2021 Study of Jewish Los Angeles. Both designs employed dual-frame sampling from combined, deduplicated Jewish organization membership lists and the U.S. Postal Service Computerized Delivery Sequence File. We discuss stratifying geographic areas for sampling based on available information about the density of the Jewish population and using household-level information related to the likelihood of being a Jewish household. In particular, we review the method for and results of employing a predictive model for Study of Jewish Los Angeles sampling using vendor data to score households' likelihood of being Jewish households, as well as implications for the study of rare populations.

Going Beyond Demographics

Brianne Gilbert, *Loyola Marymount University - Center for the Study of Los Angeles*
Lindsey Hendren, *Loyola Marymount University - StudyLA*

Los Angeles County is one of the largest and most diverse regions in the country, spanning more than 4,000 miles and encompassing more than 10 million residents in 88 cities and in unincorporated LA. However, its people are often lumped together into reductive groupings based on demographic categories such as race/ethnicity or household income. Since 2014, researchers at the Center for the Study of Los Angeles (StudyLA) at Loyola Marymount University have queried more than 18,000 Los Angeles County residents as to their attitudes and opinions on a number of topics, including homelessness, finances, policing, COVID regulations, and trust levels in our annual LA Public Opinion Survey. This survey queries at least 2,000 adult residents during a mixed mode (phone, online, and face-to-face), multiple languages endeavor. Using eight years of survey data (2014-2021), this research delves into which geographic areas (e.g., cities and neighborhoods) are the most similar or different. In addition to looking at statistical analysis of these variables, researchers also geocode respondents and map their attitudes and opinions using Geographic Information Systems (GIS) for a more drilled down analysis. This research offers a different take on public opinion of Angelenos, going beyond standard demographic variables. This county just may have more in common than we think.

Big Population, Low Incidence: Estimating COVID and Long-COVID Symptoms from a National Study of US Veterans
Robert Petrin, *Ipsos*
Mirian Adly, *US Veterans Administration*
Robert A. Petrin, *Ipsos Public Affairs*
Alan Roshwalb, *Ipsos Public Affairs*
Megan Shaheen, *Ipsos Public Affairs*

A common challenge in survey research is estimating statistically small but practically important incidence rates for essential health or behavioral indicators.   For example, as the COVID-19 crisis evolves and enters new phases concerned with long-term impacts for US health systems, researchers are seeking information on long-COVID symptoms incidence rates and severity.   Yet many symptoms which have sizeable clinical and policy implications may be hard to estimate even with traditionally high-quality survey samples due to sampling zeros in two- and three-way contingency tables.   We propose that quality estimates for scenarios such as these can be obtained via two methods: (1) adapting traditional ratio estimators for applications to contingency tables; and (2) making Bayesian adaptations to traditional loglinear models for cell counts.   The purpose of this paper is to not only provide incidence rates of long-COVID symptoms for an important demographic for which limited long-COVID information is available (US military Veterans who can represent 20% or more of high-COVID risk groups), but illustrate the two methodological approaches indicated above.   Data for this paper will be taken from a nationally representative panel of 3,340 US Veterans fielded on KnowledgePanel.   Focal outcomes will include selected long-COVID symptoms included on the survey; adjusting for the demographic factors age, race, and principle locus of care (i.e., VA and non-VA care).   We conclude by providing essentially information on the US Veteran population, as well as practical advice for extending our proposed methods to other survey, policy, and health applications.

Methodological Research Using the General Social Survey
Survey Timings Across Measures and Modes: Examples of Survey Length Estimation from The General Social Survey (GSS)
Benjamin Schapiro, *NORC at the University Of Chicago*
Rene Bautista, *NORC at the University of Chicago*
Jaesok Son, *NORC at the University of Chicago*

The General Social Survey (GSS) is a nationally representative survey, typically conducted face-to-face every two years, which collects data on the attitudes and opinions of the American public. Over the course of its history, the GSS has included a paper questionnaire with and without a leave-behind component, a computer-assisted personal interview, a computer-assisted telephone interview, and beginning in 2021, a web-based self-administered questionnaire. Since the shift to a CAPI instrument in the early 2000s, the GSS has been tracking how long it takes respondents to complete their interviews. These measures of timing are an essential component for designing subsequent rounds of the study,

helping the GSS increase likelihood of completion and decrease chance of breakoff over the course of the interview. Over the history of the GSS, different measures of timing have been employed, from noting start and end times, to measuring length of specific sections, to item-level timing. Here, we will examine all these methods of survey timing side-by-side, demonstrating how each of them reflects an overall picture of the length of a GSS interview. Comparing across modes of administration, we will demonstrate how each method of timing produces different estimates for the length of the interview, and how those estimates correlated with partial interviews in the 2021 GSS and 2016-2020 GSS Panel data. Finally, we will discuss how the GSS manages timing estimates across multiple-session interviews, and how the GSS corrects for errors introduced in these interviews.

Methodological Experiments in the Web-based 2021 General Social Survey

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*

Conducted since 1972, the General Social Survey (GSS) is celebrating its 50th anniversary in 2022. It is the only nationally representative survey of non-institutionalized adults in the United States that has systematically studied major social topics. In the year of 2020 and 2021, the GSS was redesigned as a web survey to respond to the unprecedented challenges caused by the COVID19 pandemic. A series of methodological experiments were included in the redesigned web survey to understand the impact of survey design features such as grids, inclusion of middle categories (typically coded as "volunteered" responses in the face-to-face mode; for instance, "It depends"), and more inclusive wording for nouns —namely, gender-neutral wording. The sample for the 2021 GSS survey was released in three batches. Batch 1 was released to the field on December 1, 2020. Batches 2 and 3 were released on January 21, and February 24, 2021. This paper extends on preliminary analysis of the experiments conducted on batch 1 only. The preliminary analyses were conducted based on univariate unweighted statistics, as complex survey design variables were not available at the time. The preliminary analyses (batch 1) suggested that" volunteered" categories are more likely to change distribution responses compared to grid items or gender-neutral wording (Bautista et al, 2021). This paper aims to conduct analysis on the three batches accounting for survey design variables (strata, clusters, and weights) and be based on multivariate regression analysis. The analysis aims to identify socio demographic patterns associated with measures of quality. Also, the paper investigates sociodemographic patterns of the selection of "volunteered" responses. Finally, the paper also investigates the subgroups that may be more sensitive to gender-neutral wording in survey questions. This research helps to understand respondent dynamics as large studies such as the GSS pivot from face-to-face modes to web-based instruments

Neighborhood and Respondent Determinants of Panel Attrition

Tiffany Neman, *UW MADISON*

Attrition in longitudinal or panel surveys has long served as a prime area for investigators to explore the potential determinants of unit nonresponse more closely. Existing research largely focuses on the role of respondents' demographic and political characteristics, survey experiences, and dispositions in predicting panel attrition, suggesting that the overarching causes of attrition occur strictly at the level of the respondent. Relatively less attention has been paid to the structural determinants of attrition—specifically, the neighborhoods in which respondents reside, and the interplay of neighborhood and respondent characteristics. This is especially important given growing evidence of significant spatial patterning in unit nonresponse and in respondent attitudes and judgments. Using a three-wave rotating panel of the General Social Survey (GSS), the current study examines the extent to which neighborhood characteristics, net of respondent characteristics, as well as the interplay of neighborhood and respondent characteristics, predict attrition in the second and third wave of panel datasets. To do this, I link restricted, geocoded datafiles from the 2006, 2008, and 2010 three-wave GSS panels with five-year estimates from the American Community Survey (ACS) using respondent Census tracts, which approximate neighborhoods in this analysis. Considering the extent to which residential segregation occurs along racial and socioeconomic lines in the U.S., I specifically focus on racial and socioeconomic neighborhood characteristics (racial composition, median household income, and other features of the income distribution). To account for the extralocal context (i.e., larger geographies in which Census

tracts are nested), I carry out a multilevel analysis of the structural predictors of attrition, as well as the interactional effect of structural predictors and respondent characteristics

May 11, 2022 | 3:45 pm - 4:30 pm | Poster Session #1

Interest in Volunteering at Mega Events: Attitudes and Motivations for the Qatar 2022 FIFA World Cup
Abdoulaye Diop, *Qatar University*
Semsia Al-Ali Mustafa, *Qatar University*
John Lee Holmes, *Qatar University*
Engi ElMaghrabi, *Qatar University*
Maitha Mohammed O J Al Naimi, *Qatar University*

People engage in voluntary work for a wide variety of reasons and motivations. Understanding the determinants of volunteering decisions requires an examination of people's demographic characteristics, potential altruistic and egoistic motivations, previous volunteering experience and satisfaction. There has been a breadth of literature examining various factors leading to volunteering, whether for humanitarian purposes or volunteering for specific occasions such as mega-sporting events. However, much of the plethora of research about volunteering has been conducted in the context of developed, Western societies and only few studies have examined attitudes toward volunteering in other regions of the world. This is particularly the case for Gulf countries with their peculiar socio-demographic setting, as they have received limited academic attention in this regard despite successfully having hosted several global-scale events in the recent years. We address this research gap by exploring the determinants of Qatari citizens and expatriates' interest in volunteering for the upcoming FIFA 2022 World Cup in Qatar. Drawing on two waves of nationally representative surveys from Qatar, we explore various factors that could predict Qataris and expatriates' interest in volunteering. The results indicate that, even after controlling for a wide range of demographic variables, previous volunteering experience, interest in soccer, and positive attitudes about Qatar hosting the event have a strong effect on the respondents' interest in volunteering for the upcoming FIFA World Cup Qatar 2022. Practical implications and recommendations for event organizers are also discussed.

The Changing Landscape of Survey Design. A Review on the Use of Survey Modes in the Academic Field
Adam Stefkovics, *Harvard University Faculty of Arts & Sciences*
Chase Harrison, *Harvard University*
Angela Eichhorst, *Harvard University*
Dominic Joseph Skinnion, *Harvard University*

Technological and societal changes in recent decades have largely changed how survey data is collected. In the face of a variety of challenges, previously predominant modes, such as face-to-face or telephone data collections, are transitioning to self-administered or mixed-mode designs. Widespread use of online and smartphone data collection has transformed survey research to a strongly technology-mediated instrument. More recently, the COVID 19 pandemic further pushed many survey takers, including some large-scale surveys to leave their traditional modes behind. Adapting to the rapid changes of the research environment is inevitable, although, best practices and systematic empirical evidence is scarce regarding how to optimally design a survey. The recently published AAPOR's task force report provided some guidelines and insights into these changes (Olson et al. 2021). This paper aims to complement this report. We first review the literature to outline how surveying practices, specifically the use of different survey modes, have changed in published academic research in political science and sociology in the last decade. We systematically review articles that involved an analysis of survey data and were published in top journals in these fields between 2011 and 2021, and quantify and document changes in survey modes used in these published studies. Additionally, we provide an overview on the methodological changes of the most significant large-scale surveys. Second, interviews are conducted with survey experts, methodologists, and pollsters. The goal of the former approach is to provide a big picture in the academic field about the past and the present of survey modes, whereas our qualitative approach aims to provide more nuanced insights into how surveyors can design better surveys in the present challenging times.

Public Opinion Pedagogy: A Theory-Based Approach to Curriculum Design and Development
Ann Williams, *Georgia State University*

The purpose of this presentation is two-fold: the first aim is theoretical and the second is applied. With respect to theory, the paper highlights the value of bringing public opinion research to bear in college and university classrooms – a focus that is particularly relevant in our current socio-political landscape. To meet this goal, the paper introduces a theoretical approach to pedagogy that centers on understanding the expression of public opinion as an essential discursive form of communication, citizenship, and social advocacy. With respect to application and practice, the presentation includes exercises through which public opinion research can be employed in the classroom. The activities are designed to be utilized in classes that center on public opinion as well as courses in related disciplines, like journalism and media studies, where learning about the creation, interpretation, and dissemination of public opinion research is a desired outcome.

"More Essential Than Ever" or "Just Babysitting": Public Perceptions of the Childcare Workforce
Claire Kelley, *American Institutes for Research*
Sarah Kelley, *Child Trends*

The pandemic has created urgent problems with the early childhood workforce as center closures, the need to care for their own children, complaints about pay and conditions, and pandemic risk-mitigations drove many childcare workers out of the industry. The childcare workforce is down almost 130,000 people since the start of the pandemic (Washington Post, 2021). This issue - along with the shock to many families of being unable to rely on planned childcare options during the height of the pandemic - has brought the issue further into the public eye. At the core, many of these debates center around perceptions of the childcare workers themselves. On the one hand, childcare workers have been historically paid poorly (with a median wage of just over 12 dollars per hour (BLS, 2021)), as many predominantly female caring professions are. On the other hand, there are growing calls for childcare workers to be paid a 'living wage' and valued for the important role they play in children's development. These debates are further animated by the initial inclusion of 400 billion dollars in funding for increased childcare subsidies and pre-K in the Build Back Better bill. In this context, we use Big data -- news articles from local news media and posts on Twitter and Reddit -- to understand the public conversation around this particular issue. Analysing data from these sources using Natural Language Processing techniques including short text topic modelling and sentiment analysis, we identify key themes; which themes are associated with calls for higher wages vs calls for retaining the status quo; and the language and topics associated with each political camp on this issue. This exploratory study lays the groundwork for developing formal questionnaires, identifying issues for further study, and beginning to develop our understanding of this complex issue.

Extending Jungian Archetypes to Quantitative Survey Techniques
Edward Johnson, *Harris Poll*
Laurie, *McGrath*

Prominent psychologist Carl Jung created archetypes or common themes that exist in our collective unconscious that are still used today (Stevens, 2012). Many brands will use these designations to define themselves and establish a brand identity in its marketing and product development strategy (Dominici et al, 2016). Most of this is done in qualitative assessments as it takes quite a bit of effort to educate respondents on what the archetypes are and be able to accurately identify them without a moderator helping. We explore expanding this method to online quantitative surveys with implicit association testing (Priluck et al, 2010). We start by using a brand implicit association test to measure how closely each brand is associated with different emotional words. We then have an internal team of experts do the same to the Jungian archetypes. Lastly, we look at the difference in the scaled responses to the association testing between the brands and the Jungian archetypes. We scale that distance to get a measurement of association between the brands and the archetypes. Those who attend this presentation should be able to: 1. Understand the basic Jungian archetypes 2. Understand how implicit association tests are implemented to measure brands 3. Be able to implement these archetypes in an implicit association test   Stevens, A. (2012). The archetypes. In The handbook of Jungian psychology (pp. 88-107). Routledge. Dominici, G., Tullio, V., Siino, G., & Tani, M. (2016). Marketing archetypes: Applying Jungian psychology to marketing research. Journal of Organisational Transformation & Social Change, 13(2), 109-122. Priluck, R., & Till, B. D. (2010). Comparing a customer-based brand equity scale

with the Implicit Association Test in examining consumer responses to brands. Journal of Brand Management, 17(6), 413-4285

Too Protective To Grant Permission? Understanding Factors Impacting Parental Permission in Surveying Adolescents

Jiangzhou Fu, *UCLA Center for Health Policy Research*
Todd Hughes, *UCLA Center for Health Policy Research*
Royce Park, *UCLA Center for Health Policy Research*
Ninez Ponce, *UCLA Center for Health Policy Research*

The California Health Interview Survey (CHIS) has been consistently collecting health data regarding adolescents aged 12-17 for over a decade. However, due to decreases in the parental permission rate, which is the most substantial contributor to adolescent nonresponse, CHIS redesigned the adolescent recruitment procedure to improve parental permission rate and adolescent cooperation in 2019. This redesign has been proven efficient to increase overall parental permission rate and adolescent response rate. However, the factors impacting parental permission in CHIS adolescent recruitment merit further exploration. This paper analyzes data from CHIS 2017-2018 and CHIS 2019-2020 focusing on surveyed households with an eligible teen, and explores parent and adolescent demographic and socioeconomic characteristics impacting permission, through use of logistic regression. The interaction effect between parent and adolescent characteristics is also examined. Statistical results show that in the 2017-2018 teen recruitment design, older parents are more willing to grant permission than younger parents, while the situation reverses in the 2019-2020 redesign. Male and Asian parents are less willing to grant permission regardless of teen recruitment design. Under the 2019-2020 redesign, older parents with younger teens are trying to be more "protective", with lowest permission rates. Finally, this paper also concludes with a discussion of the recommendations on future CHIS adolescent data collection approaches from these findings.

Social, Psychological and Medical Consequences of Using Facebook: Much Ado About Nothing?

Jonathan Kelley, *International Survey Center*
Sarah Kelley, *International Survey Center*
MDR Evans, *University of Nevada*

Many Americans read news on Facebook (a quarter almost every day and 16% many times a day), news that is often politically slanted and sometimes scientifically misleading. There are widespread fears about the platform's partisan politics (favoring the left); about its role in spreading COVID-related misinformation (echoing fringe 'science'); and about it harming health (widely publicized but statistically undocumented). This has spurred calls for Facebook to censor ('moderate') content, in a (politically correct) revival of the House Un-American Activities Committee's infamous 1950s attacks on free speech. If (as often assumed) reading news on Facebook has as powerful effects, the consequences for partisan politics and for medical control of the coronavirus might well be important. But does news-viewing on Facebook actually shape people's views? We address this question with extensive national data (N>2000) from the International Social Science Survey (Round 21, Cohorts 7, 8 and 9), with reliable, well-documented multiple-item measurement, using regression methods to asses Facebook's effects on eight diverse outcomes, net of a dozen preexisting demographic, SES, and political 'usual suspects'. We find that reading news on Facebook has no practically important effect on feelings about Biden and Trump; nor on suspicion that the government faked the Apollo moon landing; nor on support for same sex marriage; nor on psychological wellbeing. Nor does it have any substantial effect on medical issues: No practically significant effect on willingness to wear masks; nor on intention to get vaccinated against COVID or the flu; nor on social distancing. In short people now, as did their forebears in the 1950s, hold views largely unaffected by media exposure. This is likely because trust in Facebook and other media is low and because people choose to read what accords with their own views, rather than what they read shaping what they think.

Is State-Level Variance in Modern Alcohol Policy Explained by Path Dependency? How Historic Support for Prohibition Explains Modern Policy.

Kathryn Setty, *Colorado State University*

Across the United States there is a great deal of variance regarding alcohol regulations from state to state. More restrictive states allow for dry counties, Sunday sale bans, and tightly regulate the types of alcohol that can be sold

outside of liquor stores. Conversely, less restrictive states allow for 24/7 purchase, drive through liquor stores, open container zones, and decreased regulation on type of alcohol sold at convenience stores. This variance is not explained by modern or historic partisanship, nor general religiosity of a given area. Religiosity classified by high amounts of Protestantism along with historic rates of Protestantism have the highest amount of explanatory value for states with more restrictive alcohol laws. Utilizing the theoretical foundations of path dependency and increasing returns this paper posits that current variance in alcohol policies across states can be accounted for by the historic levels of support for the 18 th Amendment and the temperance movement as represented by historic Protestantism in a given state. That is, after reaching a critical juncture in the passage of the 18 th Amendment, states which had been more inclined to prohibit alcohol sales reached policy equilibrium that enabled regulation to persist. This paper will use an original data set that combines historical data from the U.S Census and archival data with modern measures of religiosity, along with constructed composite variables that rank each state's alcohol policy over time as most restrictive to least restrictive. This data, presented in a time series cross section analysis, will illustrate the historic relationship between Prohibition support and modern alcohol policy. History has a prevailing, lasting impact on the modern era which can be illustrated through policy and the power paradigms that persist within our society.

The World Mental Health Survey During the Coronavirus Pandemic: Can a Long Phone Survey Be an Alternative to Face-to-Face?

John Lee Holmes, *The Social and Economic Survey Research Institute (SESRI)*
Salma Khaled, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Abdoulaye Diop, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Lina Bader, *Abdoulaye Diop, Qatar University Social and Economic Survey Research Institute (SESRI)*

The world mental health (WMH) surveys have been carried out in 30 countries. Because of its lengthy and complicated questionnaire related to mental health issues, these surveys have been implemented with face-to-face interviewing. However, the Coronavirus pandemic made this survey mode infeasible in 2020 for many countries around the world. It therefore became necessary to explore an alternative mode. In this study, we present the survey results from two WMH pilots conducted in Qatar: one just before the virus spread using the traditional face to face mode (315 respondents) and another pilot during the pandemic using a phone survey (345 respondents). Our main findings are that the response rate for the phone survey was about half that of face-to-face, and this lower response rate for the phone mode can be attributed to its higher break-off rate during the interview. This is as expected due to a lengthy and complicated questionnaire. However, even with the lower response rate, the field cost per completion for the phone survey was less than half that of the face-to-face, showing a potential advantage for the telephone. Another concern, however, is that the lower response rate with the phone survey could impact data quality. Nevertheless, our comparative analysis did not find significant differences between the two pilots with respect to the demographic variables (such as gender, age, marital status, and education). In addition, the results for most mental health indicators were similar between the two pilots. Our comparison was limited by the small sample sizes in both pilots. But it supported the feasibility of moving forward with the telephone as an alternative for the WMH survey, especially given circumstances where there is a constraint to using face-to-face. As a result, SESRI moved forward with the WMH study using the phone to reach 5000 respondents.

Development of a Public Health Surveillance System for New Fathers: Building on Over 30 Years of Experience From the Pregnancy Risk Assessment Monitoring System (PRAMS)

Katie Kortsmit, *DRH, NCCDPHP, CDC*
Clarissa Simon, *Northwestern University*
Fay Stephens, *Epidemiology Section/Maternal and Child Health Epidemiology Unit, Georgia Department of Health*
Patricia Castro, *Epidemiology Section/Maternal and Child Health Epidemiology Unit, Georgia Department of Health*
Michael Bryan, *Epidemiology Section/Maternal and Child Health Epidemiology Unit, Georgia Department of Health*
Ruben Smith, *DRH, NCCDPHP, CDC*
Katie Kortsmit, Bea Salvesen von Essen, Letitia Williams, Ada Dieke, Martha Kapaya, *DRH, NCCDPHP, CDC*
Wanda Barfield, *DRH, NCCDPHP, CDC*
Craig Garfield*, Department of Pediatrics, Northwestern University, Lurie Children's Hospital*

Fathers play key roles in the health and development of their families; however, there is no ongoing state-based, surveillance system examining men's experiences during the perinatal period. PRAMS, a site-specific, population-based surveillance system, has collected data from mothers on their experiences and behaviors before, during, and shortly after delivery for over three decades. Building on the PRAMS infrastructure, a novel public health surveillance system for new fathers was developed and piloted in the field in Georgia. This presentation reports on the design and methodology of this surveillance system and results of implementation efforts. Two methodological approaches were compared to identify the most effective approach for reaching fathers: "Indirect-to-Fathers" (surveys sent to fathers to complete via mothers' survey packets) and "Direct-to-Fathers" (surveys sent directly to fathers). A pilot study was conducted to assess differences in the proportion of fathers who completed online versus traditional mailed paper or telephone surveys. The study followed the general PRAMS protocol for mothers, and eligible fathers were sampled during October 15, 2018–July 3, 2019. During the study period, 857 participants were randomized to one of the two approaches (429 Indirect-to-Fathers; 428 Direct-to-Fathers) and invited to participate. Overall, 266 surveys (31.7% weighted response rate) were completed (134 Indirect-to-Fathers; 132 Direct-to-Fathers), with no statistical difference between arms. Most surveys were returned via mail (58%), followed by online (28%) and phone (14%) surveys. PRAMS for Dads provides a framework for the development and implementation of a public health surveillance system for new fathers. Implementation of such a system provides state- and population-based estimates on the health and behaviors of fathers before, during and shortly after pregnancy, which may better inform efforts to improve the health and well-being of fathers, mothers, and their families.

Self-Identification and Experiences of Discrimination Among Asian-Germans

Mariel Leonard, *DeZIM*
Jonas Koehler, *DeZIIM*
Kimiko Suda, *DeZIM*

In 2020, the German government authorized the creation of the National Discrimination and Monitoring (NaDiRa) project as part of a broader process of understanding the enduring legacy of colonialism and racism in Germany. Quantitative research which targets ethnic minorities and communities of color in Germany is still fairly new, particularly in the context of experiences of racism and discrimination. Using data from a convenience sample of Asian Germans conducted in 2020 by researchers at the German Center for Integration and Migration Research, this paper considers both (1) types of self-identification by Asian Germans and Germans of Asian-descent, and (2) what factors affect this self-identification. In particular, I consider factors such as age, national/ethnic origin, generation, legal status, and experiences of anti-Asian discrimination as contributing factors to the making of Asian German identities. Findings from this study will be used as part of a larger ongoing project to understand the terminology used by and within communities of color in Germany and to improve sample description and recruitment of these communities into studies of racism and discrimination.

Effectiveness of Social Media Recruitment Before and During COVID-19

Rachel Kinder, *ICF*

Recruiting a diverse group of participants for quantitative and qualitative research is a challenging but crucial task for ensuring the inclusivity in the study. However, reaching target audiences can present numerous challenges, including the cost of recruitment and demographic coverage of the method, especially during a pandemic when snapshots of health, social, and financial impact are essential for evaluating impact. One way to help guarantee planned sample sizes is to recruit from multiple sources, including social media. This study summarizes alternative methods of recruitment for cognitive interviews, surveys, and focus groups, including social media (Facebook and Instagram), and online marketplaces (Craigslist). We present the advantages and disadvantages of each recruitment method, comparing demographic coverage, data quality, and cost effectiveness. We will also present our experiences with the impact of COVID-19 on social media and Craigslist recruitment. Initial results show that click rates for ads posted during COVID-19 increased compared to before COVID-19 and continue to increase. An increase in older, low-income participants was also observed in Craigslist postings during COVID-19. Moreover, the initial results show that while Craigslist is by far the most cost-efficient approach, it lags behind Facebook and Instagram in terms of obtaining completed questionnaires or interviews with usable data. The findings of this analysis will be useful to survey researchers looking for feasible, inexpensive, and rapid means of accessing target audiences.

Form Over Function: Does Acquiescence Bias in Agreement Scales Really Exist?
Randall Thomas, *GfK*
Megan A. Hendrich, *Ipsos Public Affairs*

Some researchers have declared that agreement scales are a response format to be avoided—they believe that asking for extent of agreement or disagreement activates acquiescence bias, meaning that people's tendency to get along with others and be agreeable will cause them to select 'Agree' more often than 'Disagree.' We believe that the research to date has consistently confused scale polarity with acquiescence bias (cf. Dykema et al., 2021). Research over the past 30 years, we have shown that unipolar scales consistently lead to lower means, lower top-two-box scores, and more normal distributions. Bipolar scales, by contrast, often suffer from higher means, higher top-two-box scores, and skewed distributions. Most examinations of agreement scales employed bipolar agree-disagree scales and compared them with item-specific scales that are most often unipolar. In a series of two studies, each with over 2,000 respondents, we randomly assigned respondents to evaluate several different topics using either unipolar or bipolar agreement scales or unipolar or bipolar item-specific scales. We also developed criterion measures for each topic (e.g., if we asked how much people liked doing a particular behavior, we also asked how often they did the behavior in the past 30 days). In addition, since the midpoint is often of interest for bipolar scales, we randomly assigned respondents to either a 4-category scale or a 5-category scale. When controlling for scale polarity, we found that agreement formats led to equivalent results as item-specific formats. Looking at range-adjusted values, 4- and 5-category scales showed no mean differences as well. Further, unipolar formats led to higher levels of criterion-related validity than bipolar formats (with no substantial difference due to the number of scale points). Based on these results, we believe that agreement scales can be recommended across a wide range of applications.

The Best Laid Incentive Plans are Sometimes Visible Through the Envelope (Even When They Shouldn't Be)
Robyn Rapoport, *SSRS*
Sarah Glancey, *SSRS*
Elizabeth Sciupac, *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 recruitment materials sent via mail – specifically, the invitation letter and the incentive included with the initial outreach. Specifically, researchers want to ensure mailings are being opened, and also that any enclosed incentive monies are not being discarded. We will report on two experiments conducted in 2021 as part of a single survey designed to better understand health care experiences of adults in the US. The two experiments explored the effects of incentive amounts and sponsor branding. The first experiment tested two different incentive amounts and presentation: · $1.25 incentive in a standard envelope, where the quarter could be felt through the envelope $1.00 incentive in an envelope with a customized window, where the $1 bill could be seen prior to opening the envelope The second experiment tested whether identifying the sponsor organization had an impact on responses. The sample was assigned one of the following conditions: Sponsor organization identified in the invitation letter, Sponsor organization NOT identified in the invitation letter While our research found that the $1.25 incentive performed better, the first experiment did not perform as expected since the quarter in the $1.25 treatment was visible through the address window on the front. By comparison, the $1 bill was visible, but through a window on the back of the envelope. For the second experiment, identification of the organization did not seem to have a meaningful impact on whether recipients completed the survey. These results from the incentive experiment highlight the fact that experiments do not have to perform as designed to yield insights that can better inform future work.

Trade-offs in Mode Migration: Shifting from Face-to-face to Phone Surveys in Mexico and Russia
Adam Wolsky, *Pew Research Center*

In low- and middle-income countries, face-to-face surveys have been the go-to mode for decades. However, the COVID-19 pandemic accelerated the transition away from face-to-face surveys to alternative approaches to data collection, as social research organizations tabled in-person interviewing in response to safety concerns and travel lockdowns while leveraging high phone penetration rates and the broad adoption of mobile devices to reach publics.

This paper assesses the implications of transitioning from face-to-face to phone surveys as part of Pew Research Center's cross-national Global Attitudes Survey in two middle-income countries, Russia and Mexico. Our paper analyzes different facets of the performance of recent phone surveys in both countries. We evaluate sample representativeness by comparing demographic performance against known population parameters and sample estimates against available (non-demographic) benchmark questions to assess survey accuracy. We also contrast our findings to other, contemporaneous attitudinal surveys of similar design. By conducting these phone surveys via both landline and mobile phones, we evaluate the relative value of the two frames in terms of representativeness and accuracy. Moreover, our approach allows for a review of the dialing efficiency of each frame in terms of dispositions and associated metrics (e.g., numbers dialed per complete), as well as the value of an extensive contact protocol on each frame. We aim to provide practitioners with an overall assessment of the viability of phone surveys in both countries, including whether outcomes suggest an advantage to a dual-frame or mobile-only design. To evaluate the impact of a mode migration, our presentation will compare our phone results against outcomes from prior face-to-face surveys in both countries in terms of sample outcomes and long-standing attitudinal measures. The paper concludes with remarks on the potential benefits and challenges of a mode migration, including the premium researchers should place on in-country vendor capacity and telephonic infrastructure.

Are Generational Gaps or Life Stage Segments Better at Explaining Societal Differences

Aimee Vella Ripley, *Harris Insights & Analytics (The Harris Poll)*
Edward Paul Johnson, *The Harris Poll*
Jamie Lehr, *The Harris Poll*
Amy Mulvey, *The Harris Poll*
Alyssa Haskins, *The Harris Poll*

For many years, society has been defined by generational value conflicts (Twenge, 2010). As people are living longer and having children later in life (Smith, 1999) within a quickly changing technological world, we hypothesize that life stages rather than generational differences might be better predictors of attitudes and behaviors across a wide variety of social topics. We used a large modular survey of 23,595 respondents fielded in June of 2020 to look at which variable, life stage or generation, explains the most variance across a wide variety of social topics. The topics explored directly through behavior and attitudinal statements in the survey included Automotive, Food & Beverage, Finance, Fashion, Health and Beauty, Health, Self-care, Media, Advertising, Sports, Technology and Travel. We ran a one-way ANOVA across all 800+ variables and compared how much variation can be explained by either life stages or generation. As an extension, we ran the same tests on a look-a-like modeled dataset (Ma, 2016) of 131,000 respondents to see if the results still hold in these modeled databases commonly used as seeds for activation purposes in direct marketing. The primary research questions are: 1. Is life stage or generation a better predictor of attitudes and behaviors across a wide variety of social topics? 2. How consistent are these results, not only on original data, but on the look-a-like modeled dataset? Ma, Q. (2016). Modeling users for online advertising (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick). Twenge, J. M. (2010). A review of the empirical evidence on generational differences in work attitudes. Journal of Business and Psychology, 25(2), 201-210. Smith, T. W. (1999). The emerging 21st century American family. Chicago, IL: National Opinion Research Center.

Impact of Sending Text Messages on Survey Completion Rates Among a Hard-to-Reach Population of Youth with Disabilities and Their Families

Alex Saunders, *Mathematica*
Holly Matulewicz, *Mathematica*
Forest Crigler, *Mathematica*
Karen Katz, *Mathematica*
Anna Hill, *Mathematica*
Karen CyBulski, *Mathematica*

A wealth of research surrounds the effects of traditional outreach methods, including phone calls and mailed survey reminders, on survey completion rates. But there are far less data available on the efficacy of text message reminders on survey completion rates. This study investigated the impact of sending text message reminders to a population of youth receiving Supplemental Security Income and their parent or guardian. We sought to answer the following

questions: (1) Are there times during the field period when text reminders are more effective? (2) If a respondent completes the survey, are they most likely to call on the day that they receive the text reminder or in the days after receiving it? Mathematica sent respondents four texts over a 25-week field period reminding them to call in and complete a survey. The study team sent texts only to parents and youth who gave permission to text, and the texts acted as a supplement to other outreach methods, including emails, letters, postcards, and phone calls. We found that completion rates in the days following each text message were highest for the first text, sent early in the field period, and decreased steadily in response to the successive text messages, with very low completion rates following the fourth text reminder, particularly among parents. This aligns with other nonresponse strategies used in survey research, such as mailings or telephone reminders. We also found that respondents were most likely to complete the survey on the day they received the text reminder and that completion rates tapered in the days following the text. These findings could help researchers understand strategies surrounding text reminders, including populations that might be receptive to text reminders and best timing for sending these reminders.

The Bots are Taking Over – Lessons Learned from Battling Bots on a Social Media Survey

Allison Tippery, *Market Decisions Research*
John Charles, *Market Decisions Research*
Patrick Madden, *Market Decisions Research*

A "Bot" is a type of software application that performs automated tasks on command. They can impersonate people by interacting with computer systems and can independently complete a wide range of online operations such as clicking through an online survey for compensation. The impact of bots on surveys and methods to prevent them has been well researched. However, a recent study by the Maryland Department of Health and Market Decisions Research (MDR) provides a case study on how bots are adapting to traditional methods of detection and prevention and are becoming more difficult to stop.   The goal of the Maryland Youth Pandemic Behavior Study was to measure the impact of the COVID-19 pandemic on health behaviors and risk factors among high school students. The survey was conducted in Qualtrics and included a $10 post survey incentive via email. Respondents were recruited through targeted Facebook/Instagram ads. The survey included standard data quality checks such as speed and attention checks and use of a ReCaptcha score. Soon after the ads launched, an unusual amount of surveys were completed and the preliminary data showed rates of drug use as much as five times higher than anticipated.   Over the course of data collection, MDR implemented a series of more comprehensive security measures to catch fraudulent surveys. This included multiple consistency checks, checks for duplicate and suspicious emails, multiple logic checks questions, and open-ended question checks. A total 5,978 completed surveys resulted in only 608 validated surveys. This case study demonstrates the challenge preventing bots from fraudulently attacking surveys for monetary gain, the issues of using a social media recruitment strategy for surveys, and how current bots can easily bypass most traditional prevention methods. It also provides lessons learned and recommendations for better and more consistent data security measures in online surveys*.*

Controlling for Response Order Effects

Alyssa Haskins, *Harris Insights & Analytics (The Harris Poll)*
Paul Johnson, *The Harris Poll*
Lena Schafheimer, *The Harris Poll*
Jamie Lehr, *The Harris Poll*

Although all online surveys are susceptible to primacy effects, we seek to reduce these as it often isn't standard for many surveys to rotate response ordering. While previous research (https://ojs.ub.uni-konstanz.de/srm/article/view/7263) hasn't shown a strong correlation between satisficing and primacy effects, we extend that research to other design elements, quality checks, and balancing. We hypothesize that the following four steps reduce primacy effects, mitigating the need to rotate the display order: 1. Selecting certain sample sources 2. Adapting the scale for device type (PC versus smartphone or tablet) 3. Removing fraudulent and low-quality responses 4. Post-stratification weighting on demographics   We interviewed approximately 7,000 respondents in June 2021 from nine different online opt-in samples from seven different top online sample providers and measured primacy effects through random scale order presentation. We compare the primacy effects across sample sources, device type,

quality check removals, and with demographic balancing, to test our hypothesis. Attendees will learn key tips for reducing primacy effects when using a fixed scale order.

When Near or Far, Should You Logo Wherever You Are? Branding Envelopes To Increase Response Rates

Alyssa McCoy, *University of Northern Iowa*
Kyle Endres, *University of Northern Iowa*
Mary Losch, *University of Northern Iowa*
Kelli Larson, *Arizona State University*
Jeffrey Brown, *Arizona State University*

This study evaluates how the use of logos on envelopes affected the response rate for a self-administered mail-back survey with a Push-to-Web option. Across three waves of survey packet distribution, we randomly assigned addresses to one of three envelope conditions. The first condition included logos for both the in-state sponsor and the out-of-state collaborator along with the return address. The second condition included only the return address (no logo), while the third condition included the out-of-state collaborator's logo with the return address. Participants who had previously taken part in the study as well as a newly drawn address-based sample received invitations to take part in the study. Across both samples and all mailings, the return address only envelope (condition 2) had a significantly higher response rate compared to condition 3 (out-of-state logo and return address). However, additional exploratory analyses revealed that participants who previously participated in the study were more likely to respond when randomly assigned to either the condition with logos of an in-state sponsor or the condition without logos compared to envelopes with only the out-of-state collaborator's logo. Among those who previously participated in the study, there was no difference in response rate between envelopes with the in-state sponsor's logo and the condition without logos. This research adds to the understanding of when it may be beneficial to include logos on mailed envelopes and when to avoid logos.

"No, No, No. Not Today": Dynamics of Respondent Demographics and Non-Response in Open-ended Survey Questions

Amina Dunn, *Pew Research Center*
Vianney Gomez, *Pew Research Center*

Open-ended questions provide powerful and useful qualitative data. Yet existing research has shown how some types of open-ended survey questions can be prone to high rates of non-response and wide variation in the quality of responses given. Through this analysis, we seek to better understand whether there are other factors beyond cognitive and time burden that affect non-response on open-ended questions, as well as identify how differential non-response to open ends may have implications for analysis of open ended responses. Using the Pew Research Center's American Trends Panel, we examine the extent to which certain demographics (such as, gender and age), device use, and other factors are associated with a respondent's willingness to engage with open-ended questions or the quality of their response. By focusing our attention to response behaviors based on these different characteristics, we may better understand how these factors affect the quality of the data gathered by these questions

Increasing Equity in Data Reuse through Better Planning and Data Sharing

Amy Pienta, *University of Michigan*
Tamara Qawasmeh, *University of Michigan*
Sara Britt, *University of Michigan*
David Thomas, *ICPSR, University of Michigan*

The Health and Medical Care Archive at ICPSR contains a range of surveys and related data useful for researchers wishing to conduct scientific endeavors to better understand the underpinnings of how promoting a culture of health for all can help reduce health disparities and improve health equity in the US. However, if access to a broad range of data is itself unequal, archives themselves can without intending it delay or inhibit progress made on these important societal questions. This poster will outline common barriers introduced unintentionally by data archives and some of the steps taken by HMCA to further broaden access to its data to support great equity in how data is shared -- an important sometimes overlooked part of the data life cycle. The project team and the sponsor worked on this by redefining the

collection development policy of HMCA to give greater priority to archiving and disseminating a broader range of data and adding the virtual data enclave access option for users wishing to analyze restricted-use data at HMCA. Information about additional steps that can be taken to increase equity in how data are shared more generally will be provided. We give an overview of equity issues that arise through data sharing and reuse in order to increase awareness among researchers collecting data so that they can develop and implement data management plans that better attend to data equity throughout all stages of the data life cycle to broaden and improve access. Finally, those wishing to archive data will learn about aspects of repositories that increase or reduce inequities in how data are used.

Recruiting Cancer Survivors from SEER Registries – Findings from the 2021 HINTS-SEER Study

Andrew Caporaso, *Westat*
Terisa Davis, *Westat*

This analysis will present data collection outcomes and preliminary findings from the 2021 HINTS-SEER study, sponsored by the National Cancer Institute (NCI). HINTS, the Health Information National Trends Survey, is an ongoing cross sectional postal survey designed to monitor health information seeking behavior, attitudes about health topics, and correlates of health outcomes in the United States. HINTS-SEER was a special HINTS cycle which focused exclusively on cancer survivors who were recruited from three SEER registries. SEER (Surveillance, Epidemiology and End Results) registries are maintained by NCI and have collected information about cancer incidence and cancer patients in the United States since the 1970s. This report will present the HINTS-SEER methodology, data collection results, non-response bias analysis, and compare key estimates from the HINTS-SEER sample to those from other more representative samples of cancer survivors, including those sampled in the 2020 HINTS cycle and other national US health studies. The findings will be informative for health researchers interested in evaluating methods for recruiting cancer survivors for research. SEER registries are organized by region and each have different approaches for releasing registrant data for research. Differences across regions led to data collection, integration and harmonization challenges for the study. We found that passive consenting methods led to higher consent rates but lower response rates than active consenting methods. Among the 1,234 HINTS-SEER respondents, just over one percent reported never having cancer. Similarly, agreement in demographics between administrative registry data and survey reports was high but not perfect. We discuss these similarities and differences when considering if it is possible to combine the registry sample with the HINTS national sample into a single analysis*.*

A comparison of US adult profiles based on attitudes about abortion being moral vs legal

Ronna Turner, *University of Arkansas*
Ronna C. Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*
Kristen N. Jozkowski, *Indiana University*
Brandon Crawford, *Indiana University*
Beyza Buyuker, *Indiana University*

Prior research indicates that US adults' attitudes about whether they believe abortion is moral does not always align with whether they believe abortion should be legal. Differences in attitudes about the morality and legality of abortion have been demonstrated for participants of different backgrounds such as racial/cultural groups (Dugger, 1998), political parties (Turner et al., 2021), and religious identity (Montenegro et al., 2020). Using a decision tree analytical approach, participant profiles are first developed for identifying demographic, sociological, personal experiences, and attitudinal variables related to beliefs about abortion morality and legality. Analyses are then conducted to identify which background variables differentiate between participant subgroups whose attitudes align on the morality and legality variables (i.e., it is moral and should be legal, it is immoral and should be illegal) versus participant subgroups who differ on these two outcomes (i.e., it is immoral but should be legal). Background variables such as prior reproductive health experiences, attitude about government involvement, and belief in when life begins are also included. The results are intended to provide a more nuanced understanding of how background factors may be related to complexity in adults' attitudes about abortion. National data collections obtained from a GfK panel collected in 2019 (n = 1583) and a nationally representative Amerispeak panel (n = 1025) collected in 2021 are used for cross-comparisons

Unreal Respondents? Using Machine Learning to Identify Role Players in Online Surveys

Sarah Kelley, *American Institutes for Research*

As public opinion researchers increasingly turn to online surveys - for cost savings and because of the pandemic - concerns have arisen about data quality. Online samples are known to be younger, more educated and more liberal than the populations as a whole, but these kinds of biases are relatively easily identified (though comparison with population statistics) and corrected via weighting. A subtler - and thus potentially more dangerous - problem is people taking online survey disingenuously: not representing their real attitudes or playing the role of someone else. This is surprisingly - and increasingly - common. One large commercial survey panel from 2019 reported 20% of respondents admitting to role playing during the survey and more than 30% admitted to doing so in recent Mechanical Turk panels (Kelley, Evans and Kelley, under review). Given this new threat to survey research, it is imperative to begin developing solutions to help identify and eliminate such respondents. This paper tests a machine learning approach to identifying role-playing respondents based on differences in response patterns including meta features (like time taken to respond) and responses to survey questions themselves. This model exploits the fact that role-players are systematically different than 'real respondents' with regard to the multivariate structure of their respondents.

Should I Ask You to Provide This Information in One Question or Two Questions?

Jonathan Katz, *U.S. Census Bureau*
Elizabeth Nichols, *U.S. Census Bureau*
Erica Olmsted Hawala, *U.S. Census Bureau*

Some survey questions ask respondents to report additional information in a follow-up question. For example, a question about where the respondent was born might ask if they were born in or outside of the U.S. and then be followed by a question about what state or country they were born in, depending on their previous answer. In paper surveys, the follow-up detail (state or country of birth in this example) is asked not as an explicit question but simply with instructions to enter the detail in the write-in field embedded within the response choice. This type of design saves space on the paper questionnaire, which can reduce costs in printing and give a visual clue of the skip sequence to respondents. However, research shows that sometimes respondents do not provide the follow-up detail, most likely because they do not notice the instructions and/or write-in fields for their answers. In web surveys, concerns about saving space are no longer relevant. One possible alternative design for the web is to ask an explicit follow-up question. Respondents would answer the initial question and the appropriate follow-up would display as its own question on the same page. Our hypothesis is that implementing a two-question design, rather than a one-question design with an embedded instruction and write-in field for the follow-up detail, will increase the likelihood of respondents answering the additional item, as the design follows a more explicit reading and answering question process. To test this hypothesis, we conducted two split-ballot web survey experiments testing a one-question design against a two-question design. In this presentation we will share our findings on item nonresponse, user burden as measured by time-on-task, and user preference between the two designs. Findings from this research will help inform design of similar questions that ask for additional follow-up information.

A Census Survey of IRS Appeals Customer Satisfaction

Juan Bermudez, *ICF*
Matthew D. McDonough, *ICF*
Sherri Mamon, *ICF*

The IRS Independent Office of Appeals helps resolve tax disputes fairly and impartially without litigation. Appeals conducts an annual customer satisfaction survey as one measure to assess the service it is providing to the more than 70,000 taxpayers who enter the appeals process. The results of this survey allow Appeals to assess customer satisfaction with their experience and make informed service improvements. ICF has administered this survey for more than 15 years and has further assisted in making improvements to the sampling, questionnaire, and mode of administration. The current Appeals survey is administered quarterly using mail push-to-web protocol with telephone follow-up. However, the survey is evolving in the Fall of 2021 by integrating the survey invitation with the Appeals case close-out. Instead of being mailed separately, the web survey invitation and URL will be included as part of the closing document provided to each Appeals customer. This transition further means the survey will be exclusively web based

with no telephone follow-ups. The Appeals customer satisfaction survey will be the first IRS survey to use this methodology. This presentation will describe the transition of the Appeals customer satisfaction survey from quarterly survey of a sample of Appeals customers to a census point-of-closure survey. Initial data from FY 2022 survey administration will be compared to data from previous survey administration, and implications for future administration will be discussed.

Using the NoStat File To Supplement Address-Based Sampling Frames: How To Make Informed Decisions Regarding Inclusion and Exclusion of NoStat Records

Katie Archambeau, *NORC at the University Of Chicago*
Colm O'Muircheartaigh, *University of Chicago/NORC at the University of Chicago*
Ned English, *NORC at the University of Chicago*
Lee Fiorio, *NORC at the University of Chicago*

Address-based sampling (ABS) has evolved over the years; early ABS frames were created using in-person listing efforts, whereas current ABS frames are almost entirely extracted from the USPS Computerized Delivery Sequence (CDS) file. However, there remain pockets of undercoverage scattered throughout the United States. This paper attempts to further existing research on the utility of the NoStat file in supplementing areas of poor coverage. Past research has been done on the NoStat file, but much of this work has been inconclusive or has not been conducted in recent years. This research takes a fresh look at variables on the NoStat file and provides preliminary recommendations on determining whether or not a record from the NoStat file can be reliably used as a supplementary line on the CDS file. We began by reviewing a random sample of NoStat lines using publicly-available internet resources. Selected addresses were reviewed to determine whether they appeared to be a valid residential address that should be included under standard ABS coverage or if there were indications that the address referenced a business, vacant lot, or other type of address that should be excluded from a residential ABS frame. We used generalized internet searches to categorize the lines as likely for inclusion or exclusion, then created basic models using components of the NoStat file to categorize all lines on the NoStat file as likely inclusions or exclusions. We followed up the in-office research with in-field checks of addresses on the NoStat file to determine the accuracy of our prediction model. We use the outcomes of our field checks to provide estimates on the value-add of NoStat lines to CDS file coverage.

Hop In! Evaluating Our Efforts to Drive to a More Engaged, Informed, and Satisfied Public

Kayla Dean, *Minnesota Department of Transportation*

Transportation planning is a lengthy and tedious process that often doesn't allow public involvement until later in a project's planning efforts. In 2020, the Minnesota Department of Transportation launched a statewide survey surveying a variety of different construction projects after their completion. The post-construction evaluation survey measures the publics' response on satisfaction with the project, communications, and their opportunities for engagement after they've had a chance to use the roadway post-construction. The data will help us understand how effective our communications and engagement efforts are (at the project-level and macro state level) so improvements can be made where necessary, and build upon successes. In addition, the survey is being used to collect Title VI demographic data and FHWA reportable data. Goals of the survey are to eliminate participation barriers by engaging Environmental Justice (EJ) communities and low-income populations and creating a baseline understanding of public engagement and communications-related measures across the agency.

Benefits of Migrating the Quality Control Program to Centralized Reinterview

Kenisha Steward, *U.S. Census Bureau*
Candice Barnes, *U.S. Census Bureau*

In 2021, the New York City Housing and Vacancy Survey (NYCHVS) changed its method of data collection from a Paper-Assisted Personal Interview (PAPI) to Computer-Assisted Personal Interview (CAPI). This change allowed the survey to streamline the data collection efforts by automating more processes and enabling the survey to migrate its Quality Control (QC) reinterview program to the Centralized Reinterview System that supports other demographic surveys. The Census Bureau established Centralized Reinterview in 2015 as an independent operation to administer QC reinterviews through Computer-Assisted Telephone Interviewing (CATI) web-based system at centralized call centers, or Computer-

Assisted Personal Interviewing (CAPI) in the Regional Offices (ROs), as needed. This minimized possible sources of non-sampling error identified when implementing QC reinterviews solely in local ROs. In addition, this operational change allowed the NYCHVS Quality Assurance program to expand the focus and objective beyond coverage of vacant units to monitor additional non-sampling errors that could impact the quality of data collection including, within household coverage (roster accuracy), classification of household status, and suspected data falsification. This information normally gauges how well interviewers followed established procedures and non-adherence could influence the data collected. Analysis of the quality indicators was limited to the 2021 NYCHVS reinterview data and compared to current surveys, from data sources containing results from states within the New York RO boundary. The data was restricted to results from the first interviewing cycle the current survey switched to a CATI-CAPI reinterview mode. Keywords : NYCHVS, Centralized Reinterview, CATI-CAPI.

Keeping Up With Emerging Agricultural Data Needs: Development of the Hemp Acreage and Production Survey

Kenneth Herrell, *U.S. Census Bureau*
Larry Whaley, *National Agricultural Statistics Service, USDA*
Joshua Bates, *National Agricultural Statistics Service, USDA*

The USDA's National Agricultural Statistics Service (NASS) produces statistics from a suite of surveys and censuses that cover nearly all of the nation's agricultural commodities. Depending on the commodity, these questionnaires can ask varying levels of detail on a wide variety of topics, ranging from basic information such as income, prices, acreage, and harvests, to more specific items such as the potential for disease spread from livestock. One notable exception to the commodities collected has been hemp: until now. Following the federal legalization of hemp through the 2014 and 2018 Farm Bills, NASS launched its first-ever Hemp Acreage and Production Survey. The core data needs identified for this survey were the acreage, amount harvested, and price received. The questionnaire was developed using cognitive interviews that sought to bridge gaps between respondents and researchers in terms of concepts, terminology, and practices. Through these interviews, we learned many aspects of hemp as both a product and an industry that made us reevaluate what data are needed for this dynamic and emerging commodity. Some topics that shaped the questionnaire include the diverse uses of hemp and the harvesting needs for each of these different uses, the fluctuating market for a product that was legalized nationally only three years prior, the various ways hemp processing can impact the prices received, and the relationship between growers and their buyers. We also encountered the challenge of collecting data from a large number of respondents who were previously unknown to NASS. This case study paper seeks to capture a unique event at NASS–the development of a questionnaire for a new commodity–from conception to launch.

Predicting App Adoption Rate Curves Using Time Series Modelling and Nearest Neighbors

Tomer Zur, *Harris Insights & Analytics (The Harris Poll)*
Paul Johnson, *The Harris Poll*

Obtaining a wide user base with frequent usage is one of the most important aspects of monetizing a mobile or desktop application. We take the app usage history of a panel of 10,000 permissioned respondents and predict future breadth and depth of usage.   We start by fitting multiple types of time series models to each app using demographic covariates such as % of users under 30.   For each app that gets released on the app store today, there are other similar apps that have been released in the past. We use cluster analysis to identify nearest neighbor apps to predict how many users an app will have. Ultimately, the models were successful in doing this, and can be used in the future to help app developers predict the success of their apps. The models also generated new insights into which metrics are the most important predictors of an app's future success and what apps are most similar to each other.

The Effect of Corona Study Participation on Its Parent Study - Results From a Controlled Experiment

Ulrich Krieger, *University of Mannheim*
Carina Cornesse, *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 responses over the 16 week field period. We examine who took part in the additional survey and nonresponders characteristics in regards to their survey answers in the GIP before and after the additional MCS survey waves. Using members of a control group that was not invited to the MCS we can also look at the long term impact of the additional survey requests on the participation in the GIP. 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.

LOOK, This mail is important! Can Priority Mail or FedEx be an Effective Adaptive Design Intervention in ABS Surveys?

Vanessa Meldener-Harrell, *SSRS*
Cameron McPhee, *SSRS*
Arina Goyle, *SSRS*

In the era of diminishing response rates and rising data collection costs, survey researchers are turning to study designs that allocate resources in tailored and targeted ways, with the goal of increasing response and improving sample representativeness. Adaptive and responsive designs are not new, but the growing availability of auxiliary data appendable to sampling frames has enhanced survey researchers' ability to tailor specific methodologies to sampled individuals in new ways. One method that has proven successful at increasing response to ABS surveys (with mailed recruitment invitations) is the use of FedEx or Priority Mail to gain the attention of potential respondents and communicate the importance of their participation in the study. While substantial research has demonstrated the overall benefit of this approach, less has been documented regarding the differential use of this more expensive contact method with targeted, harder-to-reach, subgroups. The success of this tailored design relies upon both the ability of the survey researcher to identify in advance which sampled units likely belong to the targeted subgroups, and the ability of the intervention to increase the likelihood of response among those individuals. This poster presents the results of two experiments designed to assess whether the use of this type of mailing as a nonresponse follow-up strategy, specifically targeted to "harder-to-reach" groups, is effective at improving response rates and, most importantly, the incidence rates of these specific groups of respondents. Both experiments sent a random portion of targeted sample cases a nonresponse follow-up mailing via FedEx or Priority Mail, with the remainder receiving the same mailing via USPS First Class delivery. Preliminary results from one study suggest that while the use of Priority Mail can improve response, it may vary in its ability to increase the incidence of hard-to-reach subgroups.

Client-Side Researchers Versus Outside Contractors: How In-House Market Researchers Collaborate with External Survey Agencies

Wojciech Jablonski, *Netherlands Interdisciplinary Demographic Institute*

Survey market research agencies are important stakeholders for in-company research professionals. Marketing research departments tend to follow a "do-it-together" or a "partial DIY" approach, rather than relying solely on outside market research providers or conducting the research fully in-house. Some of the quantitative research tasks take place in-house while others are outsourced to vendors. For example, fielding the survey and tabulating survey responses can be outsourced, and the remaining tasks can be performed by the in-house staff. Another strategy is using in-company research staff for all conceptual tasks and outsourcing to panel and survey software providers. We focus on potential improvements in client-agency collaboration, as seen by research executives who outsource market research services for their companies. We conducted qualitative study (in-depth interviews) in 2018, contacting 39 professionals employed by various companies in the European Union (EU). Participants were recruited among senior professionals and heads of customer insight or market research departments. Respondents found the major problem in client-agency collaboration to be the inability of agency researchers to envision the research process as a whole. This involves two issues. One is that researchers lack the business-specific knowledge at the level of a particular client or even the entire field of business. As a result, they might not be able to smoothly transform research output into actionable insights. The other is that research agency staff is not involved enough in socializing research results within business organizations. In-house marketing research managers are most often believed by the agency to be their end

clients. However, in a vast majority of situations, it is the marketers and C-level executives who ultimately receive the results.

Followers or Learners? Investigating the Role of Source on Partisan Sensitivity to Elite Cues
Zachary Hertz, *University of Chicago*

To what extent is vote choice driven by underlying issue preferences, psychological partisan attachments, or cues from political elites? Studies present conflicting views of the American electorate: some cast voters as steadfast partisans, whose vote choice is driven by their partisan identities, while others find evidence that voters update policy positions in response to new information, even from opposing party leaders. Issue importance, I suggest, may resolve the tension between these seemingly contradictory findings. I conduct four survey experiments and solicit respondents' preferred policy position over the minimum wage, national security budget, health care policy, and foreign tariffs.   I show that while partisanship is a driving force behind position-taking among salient issues, on less salient issues voters update their.

May 11, 2022 | 4:30 pm – 6:00 pm | Concurrent Sessions B

Measuring Attitudes & Opinions Across the Decades with the General Social Survey
The Perception of White Precarity and Longing for the Wages of Whiteness
Madeleine Straubel, *University of North Carolina - Chapel Hill*
Mosi Ifatunji, *University of Wisconsin-Madison*

Classical theories on the 'wages of Whiteness' stipulate that working class Whites are often compensated with their privileged racial status in the United States. That is, while they do not have access to certain economic assets and opportunities, working class Whites take solace in the idea that their social and political status remains greater than non-Whites of similar economic means vis-a-vis the privileges allotted to Whiteness. Therefore, in the wake of the Great Recession and the election of the first Black President, many have speculated that working class Whites felt historically high levels of social, political, and economic precarity. Arguably, many Whites responded to this context with elevated levels of racially antagonistic attitudes intended to re-secure the wages of Whiteness, including more explicit claims regarding Whiteness as the social and political norm and elevated prejudice toward non-Whites. Drawing on the 2006 to 2014 waves of the General Social Survey Panels, we examine the relationship between White precarity and the racial attitudes of working class Whites. Our longitudinal mixed effects models show that Whites in the lower rungs of the American economy were most likely to place a greater value on being White and to express more prejudicial attitudes toward non-Whites. We take these findings to support classical theories concerning the wages of Whiteness and speculate on the ways in which these trends contributed to the election of Donald Trump.

The Politicization of Disagreement: Assessing the Changing Structure of American Public Opinion, 1972–2018
Stuart Perrett, *New York University*

In recent decades, responses to survey items have increasingly sorted along party lines. This alignment represents one important feature of the structure of American public opinion. Another is the level of disagreement that characterizes response distributions. Across a broad spectrum of opinions, beliefs, and identities, what is the relationship between political sorting and disagreement - that is, to what extent is disagreement political? And how has this changed over time? To answer these questions, I use 350 subjective, ordinal response items fielded across three or more decades in the General Social Survey. Taking political sorting as the correlation between issue positions and either party or ideological identification and using a measure of ordinal dispersion to estimate disagreement, I find an increasing positive association between the two phenomena. In the 1970's, disagreement was almost unrelated to how neatly opinion was sorted by party identification. Since then, items with higher levels of disagreement are more likely to be sorted by party. Disagreement has been more consistently associated with higher levels of ideological sorting, though this relationship has also grown since the 1980's. I then ask which issues or domains have propelled the politicization of disagreement in public opinion. I decompose the estimated coefficients between sorting and disagreement by item to

quantify their contribution to this relationship for each decade. I find that items from two domains play a large role throughout the period: public spending, and sexuality and abortion. Nevertheless, there is a great deal of heterogeneity within domains and across time. Though disagreement in American public opinion has become more political, this relationship is driven by a relatively small number of issues at any one point in time. A substantial proportion of disagreement remains unrelated to ideological or partisan divisions.

Party Differences and Thermostatic Responsiveness on the Environment 1973-2018

Lindsey Hendren, *Center for the Study of Los Angeles*

Political scientists have long examined the effects of partisanship in various arenas of political culture in the US and abroad. On the issue of climate change, a scientific consensus exists concerning the presence of anthropogenic climate change and a sense of urgency to combat some of its effects is beginning to emerge. However, the struggle climate scientists and those who are concerned about this issue face include individual's reliance on party cues in forming opinions of climate change. Adding to their struggles, most proposed policies are partisan due to the messaging of party elites. This research uses a two-pronged approach to provide a comprehensive study of public opinion on the environment evaluating both aggregate and individual level data using the General Social Survey from 1973-2018. Moreover, this research is among the first to examine whether the public is thermostatically responsive on issues of the environment at the individual level. The results provide support of a thermostatically responsive public on the issue of the environment.

Conceptualizing & Measuring the Complexity of Abortion Attitudes in the United States

Conceptualizing and Measuring the Complexity of Abortion Attitudes in the United States

Kristen Jozkowski, *Indiana University*
Kristen N Jozkowski, *Indiana University*
Brandon L. Crawford, *Indiana University*
Ronna C. Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*
Sara McClelland, *University of Michigan*

According to findings from several national polls and surveys, attitudes toward abortion in the United States have remained relatively stable since the 1970's. Despite these trends, abortion legislation has increased dramatically. Given such shifts in the abortion landscape, it is important to consider how abortion attitudes are measured. Abortion attitudes have primarily been assessed via two broad frameworks--the extent that abortion should be legal and is moral. Within these frameworks, people's attitudes are often contextual. Factors such as gestational age and circumstances regarding the pregnancy and the people involved in the pregnancy influence opinions. There may also be variability in strength of endorsement or support for abortion across these frameworks and contexts. For example, some people hold strong beliefs about abortion under some conditions but may hold less strong convictions under other circumstances. Additionally, some people may feel ambivalent about abortion in general or ambivalent across frameworks and circumstances associated with abortion. Given these findings, we argue that people's attitudes toward abortion are complex and multi-dimensional; current measures are not designed to capture such nuance. In this panel, we provide evidence for this complexity across five studies. First, we present findings from in-depth interviews, highlighting salient frameworks and contexts relevant to abortion. Next, we compare people's attitudes across two dominate frameworks—morality and legality—demonstrating how people may hold completing perspectives. Third, within the legality framework, we demonstrate how people's attitudes may vary based on whether they are asked about abortion as being legal versus illegal. Fourth, we provide evidence of complexity in terms of how people identify with regard to pro-life and pro-choice labels. And finally, we demonstrate how complexity in attitudes may vary depending on item wording and response format. We will conclude our panel by proposing ideas for capturing such complexity in abortion attitude measures going forward.

Complexity in Abortion Attitudes: Findings From in-depth interviews With US Adults

Kristen Jozkowski, *Indiana University*
Kristen N Jozkowski, *Indiana University*

Barbara Dennis, *Indiana University*
Xiana Bueno-Garcia, *Indiana University*
Brandon Crawford, *Indiana University*
Ronna Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*

We theorize people's attitudes toward abortion are complex and multi-dimensional. By complex, we mean people may feel conflicted in their stance toward abortion such that their opinions may be contextual (e.g., abortion should be legal in instances of rape, but subsequent abortions should be illegal) or their strength of endorsement may vary based on circumstances (e.g., people feel strongly in some circumstances, but ambivalent or indifferent in other circumstances). We also theorize that an aspect of complexity may manifest in people potentially holding competing views of abortion due to the multi-dimensional nature of abortion attitudes. That is, people may feel conflicted in their stance based on the framework through which they are being asked to consider abortion. Representing competing frameworks—some people may hold certain beliefs about the morality of abortion as a particular framework and those beliefs may conflict with people's perspective on whether abortion should be legal or the agency of abortion seekers. Current mechanisms to assess abortion attitudes tend to primarily focus on legality as a framework and often lack assessments of contextual factors or ability to assess strength of endorsement. Such limitations may yield conceptualizations of abortion that lack nuance. To address these limitations, we conducted in-depth qualitative interviews with English (N = 110) and Spanish-speaking (N = 60) adults from NORC's AmeriSpeak panel to generate constructs that underlie people's attitudes toward abortion; we are currently analyzing data. In this presentation, we will discuss relevant (1) frameworks for conceptualizing abortion beyond legality, (2) contextual factors people perceive as relevant in terms of shaping their attitudes, and (3) other important factors relevant to attitude formation. Data from these interviews will be integrated with findings from other formative studies to identify common themes and constructs that will drive item development for new abortion attitude scales and polling items.

Assessing Abortion Complexity Across the Dimensions of Legality and Morality

Brandon Crawford, *Indiana University*
Kristen Jozkowski, *Indiana University*
Ronna Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*

Abortion is a complex multi-dimensional social issue. Findings from public opinion polls and survey research often frame abortion attitudes in terms of support or opposition to abortion across various dimensions (e.g., legality, morality). Another way to evaluate abortion attitudes is by examining abortion complexity, instead of support or opposition. Abortion complexity refers to a value free assessment of abortion attitudes that is not a measure of support or opposition but rather a measure of the extent that someone may be conflicted or ambivalent about abortion. Stated differently, abortion complexity can be thought of as the opposite of polarization. Abortion complexity has primarily been assessed within one dimension (e.g., complexity of abortion legality across different scenarios). In the current presentation, we examine abortion complexity across dimensions by identifying the number of scenarios where a participant has disagreement between similar items dealing with the morality and legality of abortion. Specifically, participants were provided with two sets of 10 scenario specific abortion questions. One set of questions asked about the legality of abortion and the second pertained to the morality of abortion. For our complexity measure, we summed the number of instances when a participant's answers for legality and morality did not match. Data were obtained using a probability-based panel designed to be nationally representative (n=1,025). Approximately 25% of the sample showed no complexity across the 10 items, with approximately 20% being complex on at least four of the items. In our presentation, we will present overall findings regarding legal – moral complexity within the sample and discuss which socio-demographic characteristics are the best predictors of legal – moral complexity. We will conclude with a discussion of the relevance of abortion complexity when assessing abortion attitudes.

When Reverse-Oriented Item Formats are Not Polar Opposites: An Experiment Comparing Attitudinal Responses when Asking if Abortion Should be Legal vs Illegal

Ronna Turner, *University of Arkansas*
Brandon Crawford, *Indiana University*

Kristen Jozkowski, *Indiana University*
Wen-juo Lo, *University of Arkansas*

Research has indicated that reverse-oriented questions are not necessarily contradictory (Spector et al, 1997; Tourangeau & Rasinski, 1988). In this study, we conducted an experiment of wording effects to investigate participants' levels of abortion support or opposition based on question orientation: should abortion be legal versus should abortion be illegal?   We randomly assigned two sets of abortion circumstance questions to a volunteer sample (Qualtrics© n=2,042) with one group receiving the illegal format followed by the legal format, and the second group receiving the reverse order. Administering both sets to all participants allowed for an experimental comparison between the initial item wording formats, with comparisons to the second format to investigate order effects. Results indicate no significant order effect, however there was a significant wording effect. The impact of the wording effect was different for the types of abortion circumstances. Two types of circumstances were used to create composite scores (health-related and socioeconomic-related) as they have been demonstrated to function as two correlated, but multidimensional constructs (Muthen, 1981). For the health-related composite score, participants' responses to the legal format indicated significantly higher support for abortion than when they responded to the illegal format. In other words, participants were more likely to say yes when asked if abortion should be legal than to say no when asked if abortion should be illegal in circumstances such as health endangerment, rape, and genetic diseases. Conversely, participants indicated lower support for abortion in socioeconomic circumstances (e.g., unmarried, cannot afford, does not want more children) when asked if abortion should be legal (compared to illegal). Participants were less likely to say yes that abortion should be legal in socioeconomic circumstances than to say no when asked if abortion should be illegal. Results indicate item wording impacts may differ based on construct location.

Abortion Identities: Bringing Qualitative Insights to Pro-Choice and Pro-Life Categories

Sara McClelland, *University of Michigan*
Kristen Jozkowski, *Indiana University*

In two survey studies, we asked participants to report their abortion identity using a novel procedure which combined a quantitative measure with a qualitative follow up question. We asked participants to evaluate two questions: "How 'pro-life are you?" and "How 'pro-choice' are you?" on a self-anchored scale (McClelland, 2017), where they first indicated their level of agreement with each item and then provided qualitative data on how they defined each abortion idenity. Study 1 (N =72) was conducted in-person using paper and pencil; Study 2 (N =1,060) was conducted online using a sample from Prolific. Both studies were conducted with samples purposively constructed to represent diverse abortion and political identities, as well as gender and racial diversity. Across both studies, we found two patterns: (1) participants relied on extremely different definitions of pro-life and pro-choice when endorsing their level of agreement with each identity, indicating that these identities are more complex and also are less stable than most measures can capture and (2) the majority of participants endorsed at least some level of agreement of each identity, indicating more overlap between the two than most survey items allow for. These findings draw attention to the complexity of abortion identities which are often considered only in binary forms, and to the potential for mis-measurement of abortion identity in national polls where binary options are most commonly used.

Comparing Response Labelling Between Agreement and Judgment Anchors: Using a Multi-Group Structural Equation Model Approach

Wen-Juo Lo, *University of Arkansas*
Ronna Turner, *University of Arkansas*
Brandon Crawford, *Indiana University*
Kristen Jozkowski, *Indiana University*

The summative rating scale has been widely adapted in social science research after Likert designed his 5-point scale to quantify and measure underlying attitudes. The well-delineated guidelines for scale development have also been established and recommended by scholars (e.g., Krosnick & Presser, 2010). However, there is limited research investigating whether the response anchor labeling produces different response patterns, which may lead to measurement bias. Therefore, the purpose of this study is to compare two of the most frequently used anchors (Casper et al. 2019): the context labels of agreement (e.g., strongly agree – strongly disagree) and judgment (e.g., definitely legal

to definitely illegal), and evaluate if both anchors could be used interchangeably while measuring people's attitudes. A total of 1,345 participants were recruited through the Prolific panel and randomly assigned to one of four sets of online survey s . The survey adopted 6 GSS items related to abortion circumstances with 2 underlying factors. These sets differed in the number of response categories (i.e., 4 and 5) and the context labeling (i.e., agreement and judgment). To evaluate psychometric properties between group differences in anchor labelling under each response category, the multi-group structural equation model (Thompson & Green, 2013) was used for comparing latent variable means and reaching coherent conclusions at the construct level. The procedure sequentially assesses configural, metric, and scalar invariance in factor means between groups. The results indicated there were no significant differences between agreement and judgment anchors in either 4 or 5 response categories. The full paper presents a discussion of the findings of these studies within the context of existing research and common practice.

Survey Questions & Beyond: Collecting Biomarkers from Survey Respondents

You Want WHAT?! Mimicking Engagement, Creating a Sense of Trust and Community-Building Via a Web Survey Video to Boost Consent to Biomarker Collection
Shana Brown, *NORC at the University Of Chicago*
Erika Greenia, *NORC at the University of Chicago*

Convincing High School and Beyond (HS&B)'s longitudinal midlife participants to consent to an in-home health visit for the first time in the study's 40-year history would be challenge at any time, let alone in the middle of a global pandemic. After witnessing a lower consent rate in the self-administered web survey mode compared to the telephone interview mode during a 2020 pretest, the HS&B 2021 Follow-up Study team pivoted in hopes of obtaining a more comparable consent rate across web and phone modes in the main study launching in March 2021. Given the lack of published research on administering informed consent for biomarker collection in web surveys, the team embarked on an in-depth investigation of audio recordings from the pretest telephone interviews to inform improvements and identified a video as a potential improvement for motivating respondents. The goal was to mimic the engagement, sense of community, and trust-building that a conversation with a live telephone interviewer provides via this video embedded within the health visit consent screens of the web survey.   This presentation will describe the process of creating the content for the video, including the investigation of audio recordings, and bringing the consent to life visually, including providing safety reassurances in a pandemic. It will also compare consent rates between modes - and from the pretest to the main study - and examine view rates and breakoffs around the video, and explore demographic and other patterns to those consent, view, and breakoff rates.

Gaining Cooperation for In-person Biomarker Collection After a Web/Phone Survey
Christopher Wong, *NORC at the University Of Chicago*
Rachel Canas, *NORC at the University of Chicago*
Katie O'Doherty, *NORC at the University of Chicago*
Eric Grodsky, *University of Wisconsin-Madison*

Introducing a new data collection component into the administration of a longitudinal study may create challenges for gaining cooperation, particularly if the new component includes a pivot in research focus that may be surprising to longtime study participants. The High School and Beyond (HS&B) 2021 Follow-up Study has historically focused on collecting education and employment outcomes via web or phone surveys and is now asking respondents to participate in an in-person biomarker collection component to identify markers of mild cognitive impairment (MCI.) In this presentation, we describe the gaining cooperation strategies that were implemented in the late-2020 HS&B pretest for recruiting study participants for a follow-up health visit where blood, saliva, and other physical health measures would be collected. We discuss how and why these strategies were updated for the main data collection that began in spring 2021. We give specific attention to three changes in the data collection protocol: (1) adding a brief video to the Web survey that explained the procedures and purpose of the health visit; (2) instituting a new refusal conversion protocol, with an increased monetary incentive offer, for re-contacting respondents who initially declined to participate in a health visit; and (3) introducing a new incentive for those providing a blood sample to receive assay results that included COVID-19 antibody tests. Consent rates for the in-person visit increased by 16% during the first six months of main data collection when compared to the pretest. We consider the extent to which the increase in consent rate may

be attributed to the updated interventions, including by looking at different analytic subgroups. We discuss the generalizability of the lessons learned for other in-person biomarker collection studies, as well as how the COVID-19 pandemic that coincided with the pretest and the main data collection influenced strategies for gaining cooperation.

Evaluating Consent and Return Rates for Biomeasure Self-Collection
Katie O'Doherty, *NORC at the University Of Chicago*
Sara Walsh, *NORC at the University of Chicago*
Hanvit Park, *NORC at the University of Chicago*

Population-based health surveys have increasingly incorporated biomeasure collection into surveys to complement self-reported estimates of health. When an array of biomeasures are collected for a study, the measures are typically collected by a field interviewer during a face-to-face interview or medical examiner during a face-to-face health visit. Recently the National Social Life, Health and Aging Project (NSHAP) asked respondents to self-collect a variety of biomeasures ranging from sensory measures, anthropometrics, mobility measures, and biological specimens. After completing a web, phone or paper survey, respondents were invited to receive a "BioBox" containing all the supplies and instructions needed to self-collect these measures. The BioBox collection process took approximately one hour for respondents to complete. Once the respondents self-collected the measures, they were instructed to mail back the BioBox with their specimens and booklet where they recorded the outcomes. Respondents that did not initially agree to receive the BioBox during the survey received a follow-up call from an interviewer to address their questions and attempt to gain their cooperation. This presentation will discuss how the BioBox consent rate and return rate compares to in-person biomeasure consent and completion rates. We will explore whether respondents that consent and return the BioBox differ from those that do not consent or return the BioBox. Finally, we will discuss the outcomes of the post-survey BioBox refusal conversion effort.

Designing the BioBox: Innovations in Remote Biomeasure Collection
Kaitlyn O'Keefe, *NORC at the University Of Chicago*
Sara Walsh, *NORC at the University of Chicago*
Katie O'Doherty, *NORC at the University of Chicago*

The National Social Life Health and Aging Project (NSHAP) is a longitudinal population-based study focusing on the interaction between social relationships and health outcomes in older adults. Biomeasures ranging from sensory measures, anthropometrics, frailty measures, and biological specimens have been collected from respondents every 5 years since 2005. During the first 3 rounds of data collection, trained interviewers collected biomeasures as a component of an in-home interview. During NSHAP Round 4, a selection of biomeasures were adapted for remote self-collection and respondents were sent a "BioBox" containing all the supplies and instructions needed to self-collect these measures. The BioBox has been tested and tweaked over the course of a pilot study in 2018, a pretest in 2019, and a second pretest in 2020. Each test round of data collection provided valuable insight that ultimately shaped the BioBox version utilized in the NSHAP 2021 Remote Main study. This presentation will follow the evolution of the remote biomeasure collection protocols and procedures throughout each round of data collection.

Multimode Questionnaire and Biomeasure Data Collection with Older Adults: Round Four of the National Social Life, Health, and Aging Project
Anna Wiencrot, *NORC at the University Of Chicago*
Katie O'Doherty, *NORC at the University of Chicago*
Ned English, *NORC at the University of Chicago*
Rebecca Curtis, *NORC at the University of Chicago*
Sara Walsh, *NORC at the University of Chicago*
Stephen Smith, *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 study of adults born between 1920 and 1965. It has been conducted face-to-face since its first round in 2005-2006. For the current round, NSHAP was already developing ways to collect complex questionnaire and biomeasure data remotely to reduce costs,

allow for participation from respondents who prefer remote modes, and to provide the opportunity to consider a hybrid remote/face-to-face approach for future work. Due to the pandemic, the need to adapt the study for remote data collection from face-to-face was accelerated. In the first three rounds of NSHAP, an interviewer visited the respondent's home to administer a questionnaire collecting data on their social network, cognition, medication use, physical health, and other aspects of health and wellbeing. Interviewers also collected a number of biomeasures during the visit, including biological specimens, anthropometrics, and physical and sensory measures. To collect data remotely, NSHAP has adapted the questionnaire to be administered by web, phone, and self-administered paper questionnaire. To collect the biomeasures, NSHAP developed the "BioBox," providing the supplies and instructions for respondents to self-collect their health measures at home and return for laboratory analysis. This paper describes the methodological and operational challenges in adapting a lengthy, complicated longitudinal survey for remote administration in a manner that balances methodological rigor while minimizing respondent burden. We compare data from the remote data collection 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.

Are We Asking Too Much? Measuring & Reducing Respondent Burden

Harder to Bear: Understanding Respondent Burden on Perceptions of Difficulty

Ki Park, *University of Northern Iowa, Center for Social & Behavioral Research*
Ki H. Park, *CSBR-UNI*
Eva Aizpurua, *Trinity College*
Mary E. Losch, *CSBR-UNI*

Designing questionnaires which are accessible to their respective survey populations is essential to mitigate measurement and nonresponse error, and obtain quality data. In this study, we explore two distinct, but related threats: respondents´ perceived difficulty answering survey questions and perceived burden. For this, we use data from a mixed-mode (web and phone) data collection from a dual-frame RDD sample. The survey of adults in a Midwestern state (N =1,003) asked about their awareness of and attitudes toward STEM education. We start by exploring the magnitude of these two issues, and the relationship between them. Based on prior research, we anticipate that perceived difficulty would be associated with greater burden. However, we anticipate this relationship to be weak to moderate (r < .50), providing support to the hypothesis that perceived burden encompasses more than difficulty answering survey questions (Yan, Fricker, & Tsai, 2020). To further assess the overlap between those two concepts, we compare correlates of difficulty and burden. We expect to find both shared correlates, and others that are unique to each metric. The results of this study will shed light on the relationship between task difficulty and survey burden in the context of interviewer-administered telephone surveys.

Commitment is Hard Work: The Association Between Respondent Commitment, Perceptions of Effort, and Survey Engagement

Ki Park, *University of Northern Iowa, Center for Social & Behavioral Research*
Erin Heiden, *CSBR-UNI*
Mary E. Losch, *CSBR-UNI*

Many strategies are routinely implemented to minimize measurement error in survey research. These involve choices around data collection mode(s) and questionnaire design. In this study, we test the impact of a motivational statement and a respondent commitment request on multiple indicators of self-reported engagement (i.e., effort providing accurate responses, multitasking) and experiences answering the survey (i.e., perceived difficulty, enjoyment, and comfort). For this, we use data from CRONOS, a probability-based online panel implemented in Estonia, Slovenia, and the UK during Round 8 of the European Social Survey (2016/2017). Using a between-subjects design, respondents were randomly assigned to one of three experimental groups: (1) standard introduction (control group), (2) motivational text (treatment 1), and (3) motivational text with explicit commitment request (treatment 2). Based on previous research (Cibelli, 2017), we hypothesize that motivational statements and respondent commitment will be associated with greater effort and reduced multitasking behaviors. The results from this presentation will provide helpful recommendations for practitioners considering the use of such strategies in their surveys.

Comparison of Page Time and Overall Survey Time as Measures of Data Quality in Online Panel Surveys

Emily Geisen, *Qualtrics*

The overall time to complete an online survey is often used as a quality measure by survey researchers. A common rubric is to consider respondents that complete the survey beneath a certain benchmark (e.g., one third the median time) as fraudulent. However, this approach can be problematic because researchers use display, skip, and branching logic to show subsets of questions to respondents which leads to some respondents receiving fewer questions and consequently finishing the survey in a shorter time. To address this limitation, we explore whether individual page timing can be used as a more accurate predictor of quality than overall timing.   In this study, we surveyed 4,000 respondents from five different online nonprobability panel providers. For each survey page that a respondent saw, respondents were considered speeders if they submitted the page in less than one-third of the median time for that page. We then created a composite timing metric by calculating the percentage of pages the respondent saw that they sped on. We limited our analysis to pages that presented the same number of questions to all respondents. We then created an overall quality score comprising more than a dozen other quality indicators including both digital fingerprinting technology (which uses IP address and online behavior to identify bad actors) as well as straightlining, inconsistent responses, missing responses, and a review of open-ended responses.   In this presentation, we will compare whether individual page timing or overall timing more accurately predicts "poor" responses as defined by the overall quality score. The results of this study will help researchers better identify and prevent fraudulent respondents in online surveys.

Engaging Reluctant Respondents With a Shorter "Critical Items Only" Survey Instrument for Enhancing Response Rates

Brenda Rodriguez, *Abt Associates Inc*
Debi McInnis, *Abt Associates Inc*
Stan Hsieh, *Abt Associates Inc*

Follow-up studies for evaluations of social interventions pose three key challenges at each wave of data collection: maximizing response rates; minimizing item non-response; and minimizing the response rate differential between treatment and control group members. As longitudinal studies update histories in the wake of lengthy prior interviews, it becomes increasingly difficult to obtain the cooperation of study members for another long interview. The National Evaluation of the 2nd Generation of Health Profession Opportunity Grants Program (HPOG 2.0) survey was first administered 15-months after random assignment. The HPOG 2.0 study participants include low-income individuals, a population that is often difficult to engage for data collection. As the team prepared for the second wave of survey data collection, 36-months after random assignment, the team discussed alternative ways to address these challenges. Led by the Director of Analysis, the team added a shorter "critical items" version of the instrument (shorter version averaged 16 minutes; full length version averaged 57 minutes). The team offered the "critical items" version to reluctant respondents, focused on key outcomes of interest, as a final attempt to complete the survey. The evaluation team used this shorter instrument to 1) maximize overall response rates, 2) balance the response rate differential between treatment and control groups, and 3) minimize item nonresponse for the key outcomes of interest. Data collection took place after the onset of the COVID-19 pandemic, which removed the options for in-person interviewing. This shorter version of the instrument meaningfully increased the overall response rates and reduced the response rate differential between treatment and control group members. This presentation will describe the methods used to prepare and implement the "critical items" instrument and assess its impact on the response rates overall. We will also discuss challenges encountered and important lessons learned.

Are Mediocre Respondents Entitled to a Little Representation and a Little Chance?

Jason Sclar, *HarrisX*
Erik Green, *HarrisX*
Woojae Chung, *HarrisX*
Katherine Bradshaw, *HarrisX*
Dritan Nesho, *HarrisX*

The goal of survey research is to capture the most accurate view of the population. Many researchers recognize that one limitation of online panels is that there are poor quality respondents who do not pay attention and may provide different answers than their counterparts who pay closer attention. However, this discounts the possibility that the type of respondent who speeds through surveys may also hold different beliefs than the type of respondent who does not. To disentangle these possibilities, we use a collection of thousands of survey respondents from the 2020 Presidential election who were disqualified from surveys for speeding. Using multilevel regression and poststratification (MRP), we compare district level estimates that both include and exclude these lower quality respondents to election returns to determine the importance of including less attentive respondents in finding unbiased population estimates.

Privacy Attitudes Toward Mouse-Tracking Paradata Collection

Felix Henninger, *Ludwig-Maximilians-University of Munich*
Pascal J. Kieslich, *University of Mannheim*
Amanda Fernández-Fontelo, *Humboldt-Universität zu Berlin*
Sonja Greven, *Humboldt-Universität zu Berlin*
Frauke Kreuter, *Ludwig Maximilian University of Munich*

Survey participants' mouse movements provide a rich, unobtrusive source of paradata, and offer insight into the response process beyond the observed answers alone. However, the use of mouse-tracking may require participants' explicit consent that their movements are recorded and analyzed. Thus, the fundamental question arises how this affects the willingness of participants to take part in a survey at all – if prospective respondents are reluctant to complete the survey, paradata collection may do more harm than good. Previous research found that other paradata collection modes reduce the willingness to participate, and that this decrease may be influenced by the specific motivation provided for collecting the data. However, consent towards the collection of mouse movements has not been addressed so far. We show that willingness to participate in a survey decreased when mouse-tracking was part of the general consent. However, a larger proportion of the sample was willing to both take part and provide mouse-tracking data when these decisions were joined, compared to an independent opt-in separated from the decision to complete the study. This indicates that survey practitioners may face a trade-off between maximizing their overall participation rate and maximizing the number of participants that also provide mouse-tracking data. Providing motivations did not have a positive effect and in some cases even reduced participants' willingness to participate.

"Time of Day Matters": Examination of Quality Differences in Survey Participants Based on Time of Participation

Erik Green, *HarrisX*
Jason Sclar, *HarrisX*
Katherine Bradshaw, *HarrisX*
Woojae Chung, *HarrisX*
Dritan Nesho, *HarrisX*

Online surveys allow for 24-hour data collection. This improves representation by permitting access to respondents who may not be available for face-to-face interviews or telephone interviews during more conventional hours. However, these off hours also present opportunities for respondents who are more tired or distracted. When engaging in round-the-clock data collection, it is imperative that researchers understand what types of responses they are collecting at all hours of the day. Building on previous work about respondent quality (Tworek, Bisaha, Green, Nesho 2019; Green, Tworek, Ji, Nesho 2020), we look at the interaction of time of survey and quality of respondent, defining low quality respondents as those who engage in lazy respondent behaviors such as straight lining questions or rushing through the survey (Malhotra 2008). A novel data set of tens of thousands of survey responses per month collected during all times of day provides a robust foundation for interrogating the link between time and quality and helps to illuminate the tradeoff between opening a survey to a wider audience against allowing respondents to take the survey under suboptimal conditions.

Evaluating Travel Burden Items in the National Household Travel Survey Through the Lens of Item Response Theory

Alec Biehl, *Oak Ridge National Laboratory*
Ho-Ling Hwang, *Oak Ridge National Laboratory*

The general purpose of the U.S. National Household Travel Survey (NHTS) is to assist transportation planners, policymakers, and researchers in understanding travel behaviors patterns regarding mode choices and trip purposes across different population segments. Originally conceived in 1969, the NHTS has evolved over the years to reflect new knowledge and technologies in both the transportation sector and survey science. The 2017 NHTS is the latest iteration of this data collection effort and, notably, it features several new questions to capture travelers' experiences and opinions. Among these questions are five 5-point agreement items that target the notion of travel as a financial burden. One of these items assesses the overall burden of travel on responding households, another item inquires about the impact of gasoline prices on travel behavior, and the remaining three items focus on if walking, biking, and public transit are utilized to save money. While the Federal Highway Administration report titled Changing Attitudes and Transportation Choices contains charts illustrating the percentage of respondents falling into each agreement category, there has yet to be a comprehensive analysis on these items, which could potentially shed light on the increasingly important issue of transportation equity. The purpose of this research is to utilize techniques from item response theory—which has minimal application in transportation literature thus far—to uncover key properties of the five agreement items, including their capacity to reflect a single latent "travel burden" trait. Hence, graded and nominal response models will be constructed to compare results based on the inclusion of unordered response categories (e.g., "I don't know") in the analysis. Furthermore, differential item functioning will be employed to determine any significant differences in response patterns across various socio-demographic and geographic groups. Implications for the design of psychometric scales in future NHTS instruments will also be discussed.

Diary Design: Mode, Device, & Diary Alternatives

The Effects of Interviewer-administered Recruitment and Personalized Feedback on Participation and Data Quality in an App-based Household Budget Survey Across Three Countries

Bella Struminskaya, *Utrecht University*
Evelien Rodenburg, *Utrecht University*
Barry Schouten, *Statistics Netherlands and Utrecht University*
Tom Oerlemans, *Statistics Netherlands*

The wide-spread and increased use of smartphones for daily activities enables innovations in data collection in the social sciences and official statistics. For diary studies, smartphone-based data collection can potentially reduce recall errors and response burden. Using smartphone's cameras, participants can take pictures of receipts for budget surveys and food intake questionnaires; passive geolocation measurement can replace questions in travel surveys. However, to ensure representation, participants have to be willing and able to use their smartphones to perform such tasks. Moreover, participants' motivation to provide information is key to ensuring high measurement quality. A randomized experiment was implemented in the general population app-based Household Budget Survey (part of the Eurostat's ESSnet Smart Surveys) in three countries (Netherlands, Luxembourg, Spain; N=4,000). Three factors were manipulated: (1) whether interviewers recruited participants vs. mail-based recruitment, (2) whether participants were promised personalized feedback on their spending vs. no feedback mentioned, and (3) whether participants were shown insights on the automated text extraction from the uploaded photos of receipts vs. no insights. Interviewer assistance can lower the initial hurdle of app installation and increase participants' motivation to provide data. Personalized feedback can also increase motivation: people are used to receiving insights about their behavior from commercial apps (e.g., on physical activity, screen time). We will report on the recruitment success and the data quality by these conditions across countries. We use app navigation paradata to infer respondents' ability and motivation. For the Netherlands, will assess nonparticipation bias using the administrative data linked to the app data. As providing insights can improve data quality (e.g., participants correct information about purchases) or have negative effects (e.g., participants change spending behavior), we compare the reported spending behavior among the two insights groups.

Web Diary Data Collection to Capture Out-of-Home Media Measurement

Brian Wells, *Nielsen*
Hani Zainulbhai, *Nielsen*
Seth Brohinsky, *SSRS*

Passive data collection from personal and in-home electronic devices is growing and becoming more powerful. However, there are certain activities and behaviors outside of a person's home that cannot be fully described through passive data collection alone. While smartphones can passively determine location and wearable meters can detect encoded sound from a radio or TV at that location, these only capture fragments of people's exposure (or potential exposure) to programming out-of-home. Say we know you are at your local bar, but is the "big game" on? Perhaps the TV is audible, but did you actively watch the screen? Many dimensions of out-of-home media measurement currently rely on human input. To better understand out-of-home media consumption, we designed a survey to capture many of these components while trying to minimize recall bias, especially for brief, unanticipated viewings. This paper details a brand new 7-day web diary built to capture out-of-home television viewing from a nationally representative sample. Data was collected over four weeks from mid-October to mid-November 2021, with each week having an independent and representative sample covering a different 7-day period. Each day respondents were prompted to record all of their out-of-home activity and associated media consumption including what programs they viewed, whether sound was available, and how attentive they were to the program and advertisements. We intend to report on the current state of out-of-home activity and media consumption given the continued disruption of the COVID-19 pandemic, discuss which aspects of the survey diary collection worked well and which needed improvement, and provide guidance for other researchers exploring the use of a web diary for detailed, daily measures.

Innovative Use of a Probability Panel to Conduct a Web Diary Survey
Hani Zainulbhai, *Nielsen - Chicago*
Seth Brohinsky, *SSRS*
Brian M. Wells, *Nielsen*

Over the past decade, public opinion researchers have decreased their reliance on non-web based modes of data collection and are increasingly relying on probability-based panels for generalizable survey results. As the use of panels has grown, new and innovative applications for panel data have become increasingly common. Accordingly, traditional, non-web based diary studies have often proven difficult to recruit, burdensome to administer, and costly to conduct. Recently, Nielsen partnered with SSRS on an innovative study utilizing the probability-based SSRS Opinion Panel to recruit and administer a 7-day diary survey to measure out-of-home television viewership. This study consists of two parts – a Recruitment phase and a Diary phase. During the Recruitment phase, selected panelists completed a short survey and were asked to participate in the 7-day diary. Those who agreed were enrolled in the Diary and were asked to complete a once-daily diary, logging all their out-of-home activity and television consumption. The Recruitment and Diary surveys were conducted in four waves from mid-October through mid-November 2021. This presentation will provide an overview of the methodology and process used to administer the Recruitment and Diary surveys, including sample design, recruitment procedures, panelist communication, web-based diary survey design, and participation and response rates for each survey. We will discuss how effective a probability-based panel is in the recruitment of eligible and willing respondents to participate in a 7-day diary survey. Moreover, we will analyze response rates at each stage and measure the impact attrition had over the course of the study among various demographic sub-groups. This information can be useful when considering the sample design for future panel-based diary studies. Finally, we will provide overall recommendations and process improvements for future panel-based diary studies.

Surveying Hard-to-Reach & Rare Populations
NSHAP Remote Data Collection: Web Survey Experiences among Older Adults
Lauren Sedlak, *NORC at the University Of Chicago*
Jennifer Satorius, *NORC at the University of Chicago*
Anna Wiencrot, *NORC at the University of Chicago*
Kelly Pudelek, *NORC at the University of Chicago*
Hanvit Park, *NORC at the University of Chicago*

The National Social Life, Health, and Aging Project (NSHAP) is a longitudinal, population-based study of adults born between 1920 and 1965. Data are traditionally collected by interviewers every five years in the respondent's home. The fourth round of in-person data collection was scheduled for 2020 but delayed due to the pandemic. Meanwhile, NSHAP conducted an interim sub-study on COVID-19 by web, phone, and paper. In 2021, the fourth round of NSHAP resumed. A

combination of web, phone, and paper was used to administer the traditionally in-person interview to half the existing sample, holding the other half for in-person interviewing in 2022. This presentation focuses on conducting a web survey among NSHAP respondents. Web surveys are generally more cost-effective than in-person interviewing and may reach respondents who would hesitate to allow an interviewer into their home, especially in unprecedented times like a pandemic. However, little research has been done into how older adults experience and respond to web surveys. A common assumption is that older adults are unlikely to participate via web, but this has not been fully examined. We will help fill this gap with regards to best practices in design, outreach, and administration of web surveys targeted to older adults. Exploring how NSHAP respondents participated in both the fourth round and COVID-19 web surveys may inform NSHAP data collection by web in future rounds and other panel studies of older adults considering adopting remote data collection. We will describe the design of this round of NSHAP, examining response to the survey across modes, including a breakdown of respondent characteristics by mode. We will also look at metadata to compare aspects of respondents' web participation to other modes. Finally, we will review respondent comments collected during the field period to try and further contextualize the respondent web experience.

How Mobile-based Research Design and Advanced Scripting Tools Enable the Inclusion of Persons with Disabilities in West Africa
Fran Bodine, *GeoPoll*

In the midst of the coronavirus pandemic in 2020, Sightsavers, an international NGO that works with partners in developing countries to treat and prevent avoidable blindness and promote equality for people with visual impairments and other disabilities, launched a large behavior change campaign. The campaign aimed at educating communities on various preventative measures for COVID-19 and demystifying false sentiments about the virus. Staying true to their mission, the goal of the campaign was simple – "leave no one behind." The plan was to survey respondents classified as vulnerable groups and persons with disabilities in the most remote and challenging parts of Africa. Due to the pandemic, traditional face-to-face research methodologies were not an option, presenting challenges around both acquiring respondents and ensuring survey comprehension for respondents with a wide range of visual and other disability profiles. To address these challenges, Sightsavers partnered with GeoPoll, a leading provider of remote data collection in historically underserved markets and populations. In this presentation, representatives from GeoPoll will discuss how the campaign leveraged GeoPoll's remote research methodologies and advanced scripting tools, in collaboration with local partners and Sightsaver outreach facilities, to reach and accommodate respondents with varying degrees of cognitive and physical disabilities from across West Africa. We will review how the survey was rendered across multiple modes and employed with numerous formatting adjustments to ensure all respondents were able to participate and provide feedback.

Examining Mode Effects in a Sample of Jewish Households: Differences Beyond Standard Demographics
Jazmyne Sutton, *SSRS*
Eran Ben-Porath, *SSRS*
Cameron McPhee, *SSRS*

Studies targeting low incidence populations often require multiple modes of administration in order to effectively reach respondents. In this study, adults in Jewish households in the greater New York area were reached through Address Based Sampling (ABS), stratified by likely Jewish incidence. In order to increase cooperation, a multi-mode multi-contact strategy was employed so that respondents could choose to take the survey online, by phone, or on paper. Because respondents could choose the mode of completion, systematic differences in selection into mode were expected. For example, more than half of phone respondents were age 65 or older compared to less than a third of web respondents. However, even when statistically controlling for demographic differences such as age, education, gender, and parental status, significant differences by mode were observed for key outcomes such as feelings of connectedness to the Jewish community, religious participation and membership, experiences with substance abuse, and mental health outcomes. In this paper, we explore specific factors that may help explain these modal differences. Notably, Jewish denomination significantly contributed to the observed effects by mode. Additionally, issues of social desirability may also be at play when considering differences in response pattern by mode. We consider the implications of these findings for future usage of multi-mode designs in unique low-incidence populations.

Understanding Latinx Consumer Behaviors Through an Equity-Driven Approach

Alicia Garza, *Fors Marsh Group*
Natalie Morrissey, *Fors Marsh Group*

Diary studies are a powerful and accessible tool to conduct equitable ethnographic studies, especially with Spanish dominant populations who have been traditionally underrepresented in research and may be hard to reach. In this paper, we present a case study of how we combined equity-focused research approaches with a web-based diary to learn about media consumption behaviors among the Latinx population in the United States, conducted completely in Spanish. We discuss how equity-focused strategies and translation solutions can be used to promote meaningful engagement and participation in qualitative research to amplify the voices of communities who have been typically underrepresented in research. We will share flexible approaches, technical assistance practices, and translation methods used by the bilingual study team to promote participant agency and engagement throughout the data collection period. The case study illustrates the value of employing responsive and tailored research approaches to collect in-depth and participant-driven data and improve equity and representation in research.

Using Qualitative Data to Develop Interviewer Training Materials for Surveying Limited-English-Speaking Households

Mikelyn Meyers, *U.S. Census Bureau*
Patricia Goerman, *U.S. Census Bureau*
Kathleen Kephart, *U.S. Census Bureau*
Patricia LeBaron, *RTI International*
Marcus Berger, *U.S. Census Bureau*

Monolingual English-speaking interviewers in the U.S. face challenges when they encounter a Limited English Proficient (LEP) household, such as correctly identifying which language is spoken so cases can be reassigned, and finding and working with appropriate interpreters to help them complete an interview.   Current training on overcoming language barriers is often inadequate. Similarly, Census Bureau bilingual interviewers receive extensive training to administer surveys in English, but in most survey operations, they receive minimal to no training administering the survey in their non-English ("target") language.   Some surveys lack the resources to design training materials in multiple languages, and coordinating bilingual training can be difficult. The Census Bureau's Center for Behavioral Science Methods (CBSM) has conducted 42 focus groups and 38 interviews with respondents who speak a variety of languages to research topics like doorstep interactions between interviewers and LEP respondents, and LEP respondents' reactions to messages designed to overcome reluctance.   In addition, we have shadowed interviewers in the field to observe their surveying of LEP households, and conducted 41 focus groups with field staff about topics like the strategies they use for completing surveys with LEP respondents.   CBSM has recently begun using the data collected from interviewers and respondents to create, test, and deploy training materials for different survey operations to help interviewers more accurately and efficiently survey LEP households.   These efforts are targeted to needs that interviewers expressed during debriefings, and are informed by research findings from pretesting and field observations. This paper includes insights from qualitative data collection with interviewers and respondents and observation of survey operations.   We present exemplars from our training materials and cover recommendations for training monolingual and bilingual interviewers on working with LEP households, with the goal of developing training materials that are responsive to interviewer needs, evidence-based, and effective.

Advances in Automated Coding of Text & Voice Data

Automated Prediction of Emotional States During Survey Completion with Voice Data for Assessing Respondents' Engagement in Smartphone Surveys

Christoph Kern, *University of Mannheim*
Jan Karem Höhne, *University of Duisburg-Essen*
Konstantin Gavras, *University of Mannheim*
Stephan Schlosser, *University of Göttingen*

Collecting voice answers to open questions in smartphone surveys represents a promising new avenue for collecting rich and in-depth information about respondents' attitudes, behaviors, and beliefs. In addition to the spoken content, voice answers allow survey researchers to extract a variety of non-verbal cues, such as amplitudes and pitches, from the collected voice data. These non-verbal cues can be used to predict respondents' emotional states (e.g., boredom), and their level of interest, which in turn opens new opportunities for researching respondents' engagement during web survey completion. In this study, we test whether and to what extent this prediction approach results in valid and informative measurements of respondents' emotional states that are reflective of their survey participation experience. For this purpose, we conducted an experiment in a smartphone survey in Germany in July and August 2021 and asked respondents to provide voice answers before and after invoking positive or negative emotions by using two picture conditions (i.e., showing a healthy or unhealthy environment). We also collected voice answers shortly before and after respondents were asked to evaluate the survey in terms of interest, difficulty, and perceived length. We used the open-source "SurveyVoice (SVoice)" tool for collecting voice answers and the open-source openEAR toolkit for predicting respondents' emotional states. On this basis, we make use of the experimental conditions to validate the predicted emotional states of respondents and analyze their associations with the survey evaluations (in terms of interest, difficulty, and perceived length). Our study provides new insights on whether respondents' predicted emotional states are indicative of their survey participation experience. This allows us to assess the potential of utilizing voice answers for unobtrusively understanding respondents' engagement in smartphone surveys.

Toward a Semi-Automated Item Nonresponse Detector Model for Open-Response Data

Kristen Cibelli Hibben, *US National Center for Health Statistics*
Zachary Smith, *US National Center for Health Statistics*
Travis Hoppe, *US National Center for Health Statistics*
Valerie Ryan, *US National Center for Health Statistics*
Paul Scanlon, *US National Center for Health Statistics*
Kristen Miller, *US National Center for Health Statistics*

Open-ended survey questions or web probes can be valuable because they allow respondents to provide additional information without the constraints of closed-ended options. They also offer advantages when little is known about a topic, such as using results to inform the design of closed-ended response options or exploring the full breadth of interpretations to better understand a topic. However, collecting open-text responses is not without challenges. Of particular concern is data quality; open-ended questions are more prone to item nonresponse, as well as inadequate and irrelevant responses. Further, coding and analyzing large sets of qualitative data can be time-consuming and labor-intensive.   To address these challenges, we developed an item nonresponse detection model that draws on recent technological advancements in combination with targeted human-coding. The model is based on a Bidirectional Transformer for Language Understanding model, which we have fine-tuned using Simple Contrastive Sentence Embedding. This approach is powerful because it uses natural language processing, as opposed to previous nonresponse detection approaches that have relied exclusively on rules or regular expressions.   Data are from the Research and Development Survey During COVID-19 created at the National Center for Health Statistics. NORC collected data in summer 2020 using a probability-based panel representative of the US adult English-speaking non-institutionalized population. The sample was supplemented with data from a non-probability online-only opt-in sample. Using open-text responses from two web probes as case studies, we present our process of evaluating and refining the model and prospects for a generalized version we hope to make available to others. We also present results looking at the amount of item nonresponse detected for the two probes overall, by type of nonresponse based on a typology we developed, and by key socio-demographics. Initial results indicate that patterns of item nonresponse are not randomly distributed across key socio-demographic or methodological subgroups.

Qualitative Insights from Social Media Data: Seeing the Forest and the Trees

Johann Gagnon-Bartsch, *University of Michigan*
Frederick Conrad, *University of Michigan*
Michael Schober, *The New School*
Robyn Ferg, *Westat*
Juejue Wang, *University of Michigan*
Rebecca Dolgin, *The New School*

Mao Li, *University of Michigan*
Paul Beatty, *US Census Bureau*

Social media has been used, with varying degrees of success, to quantify public opinion on various topics. An implicit analogy underlying this work is that social media data may be regarded, at least in some ways, as similar to data gathered from traditional designed public opinion surveys. At the same time there has been enthusiasm about using the content of social media posts as a source of qualitative data, to complement the information produced in traditional sample surveys. This work may be motivated by an alternative analogy: that social media data may be regarded as similar to data gathered from a very large focus group. In this framework, the primary challenge is the sheer volume of data; it is simply not possible for a human to read and digest all of the relevant material posted to social media. A spectrum of approaches to this problem are possible, with varying levels of automation: bottom-up topic modeling algorithms, semi-supervised topic modeling, and human coding of different sub-samples of social media content. We seek to address three questions: (1) what is the optimal level and type of automation to allow a human analyst to obtain novel, timely, and credible insights; (2) to what extent do the insights obtained from social media posts reproduce those obtained from traditional focus groups; and (3) to what extent do social media posts provide insights that complement or are entirely distinct from those obtained from traditional focus groups? We explore these issues as part of a larger effort to leverage social media content to better understand barriers to participation in the US Census. We present an interactive software tool that enables users to explore a large corpus of social media posts, with an emphasis on allowing the users to see both "the forest and the trees."

Population Estimates Based on Social Media Scraping

Cong Ye, *American Institutes for Research*

Social media has become a major communication channel for school districts to reach their audience. Important district announcements are usually broadcast on social media. This creates an opportunity for studying district policy at the population level with no burden on districts and much less effort from researchers as compared to surveys. In addition, much of the process can be automated so that statistics can be generated quickly for timely reports on important topics, such as school closures during the COVID-19 pandemic. One challenge for this approach is to develop an automatic tool to code the announcements. Human coding would be relatively expensive and time-consuming. This challenge is addressed by topic modeling techniques which categorize documents (i.e., announcements in this case) into different topics where the probabilities of fitting each document to different topics are estimated through an iterative process. Another major challenge is that school districts use social media differently. Some have dedicated managing staff, and some may not use it at all. Therefore, simple mean estimates of the school district population are likely to be biased. This challenge is addressed by weighting adjustments as population data (e.g., the Common Core of Data) on this population are rich and high quality. The results are validated by comparison with aggregated administrative data.

Data Science Informed by Survey Science: Collecting More Accurate Labels

Stephanie Eckman, *RTI International*
Frauke Kreuter, *LMU*
Jacob Beck, *LMU*
Rob Chew, *RTI International*

Machine learning models rely on high-quality input data, for example, images labelled as dogs vs cats or text labelled as positive or negative sentiment. The instruments used to collect these labels are similar to web surveys, except that the questions are about images or text rather than about the labelers themselves. Our study tests whether the principles of data quality in web surveys also apply to the collection of labels for machine learning models. We fielded six versions of an instrument to code tweets as containing hate speech or offensive language. All tweets have been previously coded by 4 or more coders. By comparing the labels collected in the different versions, we provide the first evidence that instrument design matters in the collection of labels for data science. Our results will interest data scientists who want to save time and money by collecting high quality labels.

How Did the Survey Do? Recent Assessment of Total Survey Error

Are We There Yet? Unrepresentative Big Surveys Significantly Overestimate US Vaccine Uptake

Valerie Bradley, *University of Oxford*
Shiro Kuriwaki, *Department of Political Science, Stanford University*
Michael Isakov, *Harvard College*
Dino Sejdinovic, *University of Oxford, Department of Statistics*
Xiao-Li Meng, *Harvard University, Department of Statistics*
Seth Flaxman, *University of Oxford, Department of Political Science*

Surveys are a crucial tool for understanding public opinion and behavior, and their accuracy depends on maintaining statistical representativeness of their target populations by minimizing biases from all sources. Increasing data size shrinks confidence intervals but magnifies the impact of survey bias – an instance of the Big Data Paradox (Meng 2018). Here we demonstrate this paradox in estimates of first-dose COVID-19 vaccine uptake in US adults: Delphi-Facebook (Barkay et al 2020, Kreuter et al 2020) (about 250,000 responses per week) and Census Household Pulse (Fields et al 2020) (about 75,000 per week). By May 2021, Delphi-Facebook overestimated uptake by 17 percentage points and Census Household Pulse by 14, compared to a benchmark from the Centers for Disease Control and Prevention (CDC). Moreover, their large data sizesled to minuscule margins of error on the incorrect estimates. In contrast, an Axios-Ipsos online panel (Jacksn et al 2020) with about 1,000 responses following survey research best practices provided reliable estimates and uncertainty. We decompose observed error using a recent analyticframework (Meng 2018) to explain the inaccuracy in the three surveys. We then analyze the implications for vaccine hesitancy and willingness. We show how a survey of 250,000 respondents can produce an estimate of the population mean that is no more accurate than an estimate from a simple random sample of size 10. Our central message is that data quality matters farmore than data quantity, and compensating the former with the latter is a mathematically provable losing proposition. This paper has recently been accepted to Nature, and will appear Nov/Dec 2021.

COVID-19 Vaccination Rates in the U.S.: Comparison among National Probability-Based Surveys – National Immunization Survey Adult COVID Module, Household Pulse Survey, and the AmeriSpeak® Panel

Vicki Pineau, *NORC at the University Of Chicago*
Jason Fields, *U.S. Bureau of the Census*
David Dutwin, *NORC at the University of Chicago*
Dani Heide, *NORC at the University of Chicago*
Elizabeth Flanagan, *NORC at the University of Chicago*
Kim Nguyen, *Centers for Disease Control and Prevention*
James A. Singleton, *Centers for Disease Control and Prevention*
Michael Chen, *Centers for Disease Control and Prevention*
Benjamin Fredua, *Centers for Disease Control and Prevention*

This presentation will cover results from an investigation into differences in estimates of the proportion of adults reporting receipt of ≥ 1 doses of a COVID-19 vaccine in the U.S. reported weekly or biweekly by three national probability-based surveys during 2021: National Immunization Survey Adult COVID Module since April 22, 2021, the Household Pulse Survey since January 2021, and questions included in the AmeriSpeak Panel biweekly Omnibus sample since February 2021. Though each survey is probability-based in design, each has unique features that are in some respects expected to produce somewhat different estimates of COVID-19 vaccination in the U.S. (e.g., sample frames, survey modes); while other features, such as question wording, would be expected to produce similar estimates to each other and to official vaccination data housed in the CDC COVID-19 Data Tracker. First, we will compare the sample coverage, sample design, survey mode, fielding procedures, question wording/instructions, and sample weighting across the three surveys. Next, we will compare estimated vaccination rates for subpopulations across the three surveys to CDC's reported estimates of receipt of ≥ 1 doses of a COVID-19 vaccine for subpopulations. Comparisons will be made for population subgroups (age, Hispanic ethnicity/race, sex) and lower geographic levels when possible.   We will then seek to quantify the potential impact of differing sample designs, nonresponse, survey mode, and estimation methods that may impact the differences observed between the reported estimates from the three surveys and CDC's COVID-19 Data Tracker

Total Survey Error Analysis: National Immunization Survey Adult COVID Module COVID-19 Vaccination Rates

Elizabeth Ormson, *NORC at the University Of Chicago*
Vicki Pineau, *NORC at the University of Chicago*
Kirk Wolter, *NORC at the University of Chicago*
Jason Boim, *NORC at the University of Chicago*
James Singleton, *Centers for Disease Control and Prevention*
Michael Chen, *Centers for Disease Control and Prevention*
David Yankey, *Centers for Disease Control and Prevention*
Jennifer Kriss, *Centers for Disease Control and Prevention*
Yi Mu, *Centers for Disease Control and Prevention*

The difference between an estimated survey outcome and the true population value can be impacted by a variety of sources. This overall difference is defined as the total survey error which arises from both sampling and nonsampling errors, such as nonresponse error, sample-frame coverage error, and measurement error. In this analysis, we will present results from a total survey error evaluation of national estimates of the proportion of adults reporting receipt of ≥1 dose of a COVID-19 vaccine in the U.S. as reported each week by the National Immunization Survey Adult COVID Module (NIS-ACM) since April 2021. We will examine the potential stages in which survey error can occur in the NIS-ACM process, and seek to measure the error associated with each survey stage. External data sources, such as the American Community Survey, the National Health Interview Survey, and COVID-19 vaccine administration data from CDC will be used to assess and quantify the potential sources of error.

Has the Sun Finally Set on Landline Interviewing? Data Quality, Representativeness, and Inclusivity as a Function of Dual-Frame RDD Sample Allocation and Calibration
Luis Tipan, *SSRS*
Arifah Hasanbasri, *SSRS*
Jonathan Best, *SSRS*

Approximately 97% of US adults have a cellphone. In 2016, when that figure was at 90%, Pew Research Center analysis found that results based only on cellphone interviews are similar to results based on a combination of landline and cellphone interviews. We revisit this question six years later as the proportion of landline interviews in dual frame random digit dialing surveys continues to decline and advise on single- and dual-frame telephone designs to ensure diverse and representative samples. Related to the optimal allocation of landline and cellphone sample, we explore whether calibrating to phone use is necessary to ensure sample representatives. Given that phone use is correlated with socio-demographic variables generally included in weighting (Dutwin and Buskirk 2019; Bloomberg, et al. 2007; Strauts 2010), we examine to what extent calibrating to the population distributions of cellphone only, dual frame, and landline only users impact data quality. We ran experiments where (a) different proportions of cellphone sample were drawn (including a cellphone only design), and (b) calibrations were done both including and excluding phone use. Bias and variance on dual- and single-frame RDD data were assessed relative to health, political, social, and demographic metrics from high quality benchmark surveys such as the NHIS, BRFSS, ACS, GSS, and CPS.   This paper expands on 2016 work about the potential "twilight" of landline samples and builds on research presented at the 2021 AAPOR Virtual Conference (Best & Tipan, 2021) on phone use weighting. This research contributes to scarce experimentation on differences in data quality and weighting between single frame and dual frame telephone surveys and provides guidelines on how to design and weight telephone samples in this new era.

Better Than a Coin Flip? Prediction Accuracy in Election Polls

Social Capital and Participation: Examining Hypotheses About Polling Problems in the 2020 Election Cycle
Robert Griffin, *Democracy Fund*
John Sides, *Vanderbilt University*

In the 2020 election cycle, polling continued to show signs of bias - specifically, overestimating the democratic vote share nationally and in many states. A number of hypotheses have been put forward to explain this phenomenon as well as offer potential fixes. Notably, one story has centered around non-response by those low in social capital and/or trust while another has noted the high rates of survey participation by highly active Democrats. Using unique data from

Democracy Fund VOTER survey and the Democracy Fund + UCLA Nationscape project, we show: There is little evidence that introducing social capital and/or trust into the weighting scheme increases representativeness. There is evidence that reweighting by in-cycle measures of participation and activity increases representativeness. While these findings cannot reasonably be portrayed as "fixes" to the problems faced by modern election polling given the ever-changing nature of those issues, they do shed light on the nature of polling problems in the 2020 election cycle.

Evaluating Multi-Mode Exit Polls
Patrick Tucker, *Washington University - St. Louis*

Exit polls provide researchers with relatively instant data to identify the motivations of voters and to make predictions regarding the outcome of elections. In recent years, however, the representativeness of exit polls has been questioned for at least two main reasons. First, like many political surveys, exit polls have suffered from partisan differential non-response that biases the sample. Second, voters have changed their method of vote in large numbers, making it difficult to identify who is and who is not a voter. In this paper we evaluate the effectiveness of employing different exit poll technology, such as texting and email, to identify early and absentee voters in two recent salient elections.

"A Hit and a Miss": An Investigation into the Accuracy of Polling the 2021 Gubernatorial Elections in VA and NJ
Justine D'Elia-Kueper, *Catalist*

Heading into Election Night 2021, the public polling available for the only two gubernatorial elections of the cycle predicted a close race for the open-seat contest in Virginia, whereas in New Jersey, incumbent Democratic governor Phil Murphy looked to be heading for comfortable reelection. In an Election Night surprise, however, the Virginia gubernatorial race was called for Republican Glenn Youngkin, while the gubernatorial race in New Jersey remained too close to call until the following evening. This leads to the question of why were the pre-election polls in Virginia fairly accurate, while the New Jersey polls completely missed the mark. So badly, in fact, that a leading New Jersey pollster penned an op-ed apologizing to both campaigns and questioned the wisdom of continuing to publish pre-election polls at all. Since in all likelihood public pre-election polls are not going away anytime soon, it is important to understand what factors contributed to the polling miss in New Jersey this year. Was it the unexpected surge of voters in Central Jersey who were motivated to vote by their opposition to Murphy's mask mandates and lockdowns, as some in the media have speculated? Or could it be explained by other factors, such as the differences in pollsters' methodologies or different levels of non-response bias between the two states? This presentation will leverage voterfile data along with public and private polling data to examine how these and other factors may have contributed to the polling miss in New Jersey, with comparisons to the more accurate polling in Virginia where appropriate. It is only by learning the lessons of this cycle, that we can hope to improve polling in the future and make the need for pollsters to apologize publicly a thing of the past.

The Geography of Racially Polarized Voting: Calibrating Surveys at the District Level
Shiro Kuriwaki, *Stanford University*
Stephen Ansolabehere, *Harvard University*
Angelo Dagonel, *Harvard University*
Soichiro Yamauchi, *Harvard University*

Laws designed to prevent minority vote dilution in the United States are motivated by the premise that voting divides sharply along race and geography. We present the first study of vote preferences of racial groups in every congressional district using surveys (or to our knowledge any method). We develop calibration techniques that improve the existing multilevel regression and poststratification method. This avoids the ecological regression assumption, standard in prior work, that preferences of racial groups are independent of geography. We make two technical innovations --- survey-assisted synthetic target estimation and two-way calibration --- that enable a richer and better calibrated MRP model. Patterns of racial polarization in the US reflect both groups and geography. Sixty percent of the variation in preferences is attributable to nationwide differences across the racial groups, while 30 percent is attributable to variation within groups across geography. The Deep South and some Midwestern districts have the highest racial polarization between White and Black voters. Differences between White and Hispanic voters' preferences vary considerably across and even within states.

In Memory of Eleanor Gerber: Her Contributions to Our Understanding of the Role of Culture in Survey Question Interpretation & Survey Participation

In Memory of Eleanor Gerber: Her Contributions to Our Understanding of the Role of Culture in Survey Question Interpretation and Survey Participation

Alisu Schoua-Glusberg, *Research Support Services Inc*
Patricia Goerman, *U.S. Census Bureau*

This session reviews and honors the contributions of Eleanor Gerber to survey research. As a research anthropologist at the U.S. Census Bureau, she advocated for increased use of qualitative and ethnographic research methods, and made important contributions to including hard-to-count populations in surveys, design of survey questions on race and ethnicity, census over and undercounts, and research on respondent privacy and confidentiality. Her cognitive interviewer training courses were legendary and the foundation she laid for this work has lasted long past her time at the agency.   The session will include five presentations. The first one (Miller) will discuss how Gerber was the first to bring an anthropological viewpoint into understanding the question response process, and how this has given future direction to the theory and practice of cognitive testing. A second presentation (Childs), will showcase Gerber's ethnographic research on privacy and confidentiality, in particular as these affect survey participation among the hard to count, andwill relate it to the work currently being done at Census on privacy and confidentiality and its implications for survey methodology. The third presentation (Goerman) will focus on Gerber's innovative training methods and approaches to training interviewers for cognitive testing and how these methods are currently used at the Census Bureau. The fourth presentation (Cidade) will examine Gerber's contribution to the conceptualization and operationalization of complex social phenomena, using race and ethnicity as an example. Finally, the last presentation (Schoua-Glusberg) will discuss Gerber's qualitative examination of why and how children are undercovered or listed more than once in the decennial census. Gerber's contribution in this regard is arguably the best in-depth look we have, ten years later, at the problem of duplication. The session will conclude with an invitation to audience members to share their recollections and/or the impact of Gerber's work on their own research.

Privacy and Confidentiality: Eleanor Gerber's Legacy Research

Jennifer Childs, *U.S. Census Bureau*
Aleia Clark Fobia, *US Census Bureau*

Eleanor Gerber led important and seminal ethnographic research on privacy and confidentiality during the 1990 and 2000 Censuses. This research has inspired a long line of study using mixed-methods both for more recent censuses and more broadly applied to federal statistical data collection. This presentation will showcase Eleanor's work and draw connections between work that she conducted in the 1990s and 2000s and work being done at the U.S. Census Bureau today. Eleanor's work made critical connections between members of hard-to-count groups (HTC) and their views towards privacy, confidentiality and data sharing. Using data collected up through 2020 using a variety of methods, we will show how these views have changed over the years since the time of Eleanor's research. Her research has provided a crucial baseline, as technological and social shifts have made the issues of privacy and confidentiality and their intersection with HTC groups even more important for the field of survey methodology. From an experimentally designed test of confidentiality language during the 2010 Census through a multi-mode 2020 Privacy and Confidentiality Study and much intercensal work, this presentation will walk us through Eleanor's legacy of privacy and confidentiality research at the Census Bureau. The presentation will also discuss the future of this line of research as the Bureau moves toward the 2030 Census.

Eleanor Gerber's Contributions to our Understanding of Children's Overcount and Undercoverage in the Decennial Census

Alisu Schoua-Glusberg, *Research Support Services Inc*

Eleanor Gerber brought to the Census Bureau her ethnographic training and experience, and used it to enrich the methodological work she carried out at the Bureau. Although she was, by no means, the first anthropologist working in survey methods at the agency, she may arguably be the one who left the strongest mark in the work of a generation of researchers who benefitted from working with and learning from her. Her legacy lies in the integration of a cultural focus to the study of any issue in survey and instrument design, from research on privacy, to rostering, to race and ethnicity, to the issue of census undercoverage and duplication. Even a decade later, on some of these topics her work gives us the most complete look to date. A case in point is that of understanding how and why some persons are counted more than once in the decennial census. This presentation will illustrate Dr. Gerber's work on a couple of qualitative studies on duplication and on undercoverage that followed the 2010 census, and give examples of the study methods and findings from these research efforts. The presentation will specifically focus on what we learned about the duplication and undercoverage of children, a topic that the Census Bureau continues to study and grapple with.

Anthropology in Action: Eleanor Gerber's Contribution to the Operationalization of Race and Ethnicity
Melissa Cidade, *US Census Bureau*

Cognitive testing is a fundamental part of survey design and a requirement under the Census Bureau's data quality standards. But, to Eleanor Gerber, cognitive testing was more than a survey pre-testing method: it is a window into conceptualization and operationalization of complex social phenomenon. Using race and ethnicity as examples, Gerber shed new light on the ways that respondents perform identity in the survey response process. This paper explores some of that work from the early 2000s, and traces the impact of a social science-oriented approach to survey testing of Gerber mentees at the Census Bureau and beyond through personal narrative and structured interviews. We end with an update on the current work on race and ethnicity identification on Census Bureau surveys, and a look toward future research projects on identity and survey response.

Training the Researchers: Eleanor Gerber's Contributions to Cognitive Interviewer Training at the U.S. Census Bureau
Patricia Goerman, *US Census Bureau*

Eleanor Gerber was one of the early developers and practitioners of the cognitive interview method at the U.S. Census Bureau for use in pretesting survey questionnaires. Drawing on her experience teaching anthropology at various Universities prior to working at the Census Bureau, she designed one of the first Census Bureau cognitive interviewer training courses. One of the highlights of the course was her description of particular respondent archetypes, such as "Stan the Accountant," "Agenda Annie" and "the Professor" and hands on practice with how to interact with people who have differing communication styles. Researchers in the Center for Behavioral Science Methods have continued to use and build upon the cognitive interview framework and approach that she set out. This talk will present key components of the training with a focus on the contributions that Dr. Gerber made to the survey pretesting and research field. It will conclude with a discussion of current methods in use at the Census Bureau.

Improving the Measurement & Recruitment of Sexual & Gender Minorities in Surveys: Recent Findings from Qualitative & Quantitative Research

Improving the Measurement and Recruitment of Sexual and Gender Minorities in Surveys: Recent Findings From Qualitative and Quantitative Research
Nancy Bates, *US Census Bureau, Retired*
Mark Noyes, *Market Decision Research*
Dan Cassino, *FDU*
Thomas File, *U.S. Census Bureau*
Jennifer Ortman, *Census*

To drive research agendas and monitor trends, both the Institute of Medicine and National Academy of Sciences (IOM, 2011; NAS, 2020) have recommended the routine collection of data on sexual orientation and gender identity (SOGI) in surveys. Despite this recommendation, routine collection and research on best practices is scarce. Presentations from this session help fill this void by showcasing recent SOGI-related measurement research across a broad spectrum of platforms. These include: the Census Bureau's Pulse Survey; the Understanding America Panel Survey, state-level surveys; cognitive interviews; and a probability panel of older adults. Among other things, research findings will report on an order experiment for the two-part gender identity collection method; use of screeners to delineate between sexual and gender minorities (SGM) and non-SGM people; collecting GI by proxy; and use of scales to measure GI. Presenters reflect a wide range of institutions including the federal government, academia, and a private sector market research firm.

How Do People Want to Be Asked About Their Sexual Orientation and Gender Identity on National Health Surveys?

Anthony Pho, *Stanford University School of Medicine*
Juno Obedin-Maliver, *Stanford University School of Medicine*

Sexual and/or gender minority (SGM) people including lesbian, gay, bisexual, and transgender and gender expansive people, have demonstrated health inequities and vulnerabilities to poor health, and yet sexual orientation and gender identity are not routinely collected in national health surveys. Moreover, there is little evidence to inform how SGM and non-SGM people want to be screened about their sexual orientation and gender identity to screen for SGM status on national health surveys. We received funding from the National Institutes of Health to explore how people want to be surveyed about their sexual orientation and gender identity. We recruited and purposively sampled U.S.-based SGM and non-SGM participants through newspaper ads and social media to participate in 60-90 minute in-depth online cognitive interviews. Video interviews were recorded using Zoom software and were transcribed. We coded the interviews inductively and deductively using Dedoose software. Thirty of the forty-four participants were SGM people, and the majority were racial/ethnic minorities. We tested four versions of a sexual orientation and gender identity screener. SGM and non-SGM people preferred, "Do you identity as LGBTQ?" to alternative versions that included terms like sexual and/or gender minority, and someone with a transgender experience/history. The preferred screener was less confusing and was successful in delineating between SGM and non-SGM respondents (based on follow-up questions). Use of such screeners in general population surveys can have dual benefits: (1) reduce respondent burden among non-SGM populations, and (2) produce more nuanced and robust measures of SGM people. Many participants discussed that context of information collection (e.g., national health survey, community organization survey, local clinic form) might influence answers. Additional themes included trust/distrust, regard for government, fear, and privacy. Future phases of the grant will test the preferred screener in online surveys.

Evaluating Measurements of Sexual Orientation and Gender Identity on the Experimental Household Pulse Survey

THOMAS FILE, *U.S. Census Bureau*
Jason Fields, *U.S. Census Bureau*
Eric Stone, *U.S. Census Bureau*

For more than a year and a half the Census Bureau's experimental Household Pulse Survey (HPS) has provided researchers and policymakers with near real time information on the COVID-19 pandemic's impact. Prior research shows that sexual and gender minority (SGM) populations have differential health outcomes and access to physical and mental health care, and that their economic and educational experiences differ from the non-SGM population. In July 2021, the HPS began asking a series of three sexual orientation and gender identity (SOGI) questions: assigned sex at birth, current gender identity, and sexual orientation. Accurate survey measurement requires understanding how these new survey items perform, and this paper proposes three separate areas of important evaluation. First, using a built-in experiment from six independent collection cycles of the HPS (July 21-October 11, 2021), we will assess whether question order has any impact on reported SOGI outcomes across various demographic groups, specifically whether asking about sexual orientation before gender identity has any measurable effect on reported SOGI outcomes or survey breakoffs. Next, to assess whether certain groups of respondents understand the questions better than others, we will investigate the characteristics of respondents who "correct" their sex at birth and/or current gender identity answers when prompted by an automated "check" question built into the HPS instrument when reported sex at birth and gender identity do not align. Finally, to inform future questionnaire design and SOGI data dissemination strategies, we will

evaluate item nonresponse rates for the three individual SOGI items and assess their performance in comparison to other core demographic variables on the HPS. Improving SOGI data collection will enable researchers and public health officials to identify emerging areas of need and reduce key disparities long-term, thereby improving population health and quality of life for SGM populations.

Changes in Population-Level Response Patterns to Gender Identity Survey Items Across Administration
Mark Noyes, *Market Decisions Research*
Br. Brian Robertson, *Market Decisions Search*

Survey items that explicitly divorce assigned, biological sex from the more fluid construct of gender identity are becoming more common. While these items are growing more popular, they are a recent innovation and their performance in surveys that rely on proxy responses and on questionnaires not necessarily related to gender and sexuality is less well understood than binary, single-item gender assessments. This research attempts to address some of the questions that remain around the use of explicit g ender identity survey items, and form a realistic, probability-based estimate for the parameters of the population who identify as gender minorities.   This research tracks the performance of such a question series across multiple administrations of telephone-based, state-level household surveys from 2018 to 2021, to track differences in response and identification of individuals who are members of gender minority groups. These surveys gather detailed information about all members of a household, making them much more likely t o identify members of extremely small groups as they collect data on more individuals. We will examine questions such as item-level non-response, break-off rates, potential indications of interviewer error, changes in population identification over time and potential patterns in response as compared to data available in large scale, federal data collections. Once widely implemented as a common practice, results will reflect, and improve patterns in response.

Masculinity and Femininity: What Can We Learn From Self-Report Data?
Dan Cassino, *Fairleigh Dickinson University*

The paper looks at two inter-related questions. One: what aspects of gender identity (masculinity/femininity) can we measure through self-report scales, rather than through larger constructed scales like the BSRI? Two: Are the portions of gender identity that we can measure through self-report scales useful in structuring socio-political attitudes, controlling for existing demographics like education and partisanship? To address these issues, the paper draws from two sources: the Understanding America Survey, a large-scale national panel survey carried out over the past 18 months at USC that included one of my self-report gender identity items, and a nationally representative online survey carried out that used a number of gender identity measures. The results show that self-report scales do an excellent job of capturing one of the dimensions measured by constructed scales such as the BSRI, adherence to the gender identity most commonly associated with the individual's sex. That is, men can accurately (relative to the BSRI) rate their own masculinity, and women their own femininity, but they do a poor job of reporting traits associated with a different gender identity. However, this dimension serves to account for much of the variance in socio-political views, controlling for factors such as partisanship, and is thus valuable despite its shortcomings.

Summing It Up: Improving the Measurement and Recruitment of Sexual and Gender Minorities in Surveys
Jennifer Ortman, *U.S. Census Bureau*

Surveys that collect data on the characteristics and well-being of sexual and gender minority (SGM) populations allow for better understanding of SGM populations relative to the general or other population groups. These data meet critical needs to understand trends within larger population groups and can lead to potential resources and interventions needed to better serve the community. This discussion summarizes the recent SOGI-related measurement research presented in this session, highlighting the achievements of these endeavors, posing questions to delve deeper, and reflecting on how to push forward to support quality measurement of sexual orientation and gender identity (SOGI) across a broad spectrum of platforms.

Interviews, Interviews, Everywhere! Proxies, Question-Wording & Data Quality

Crosslinguistic Analysis of "Language Spoken at Home" Questions

Marcus Berger, *U.S. Census Bureau*
Betsarí Otero Class, *U.S. Census Bureau*
Angie Obrien, *U.S. Census Bureau*

Questions about language spoken at home are common on surveys, and data from these questions are used for a variety of purposes. This project analyzes how questions about language use are interpreted by speakers/users of different languages. The study involves two rounds of cognitive testing, in which we interviewed monolingual speakers of English and Spanish, and users of American Sign Language (ASL). Some previous research has shown that, in some cases, Spanish speakers misreport themselves as English speakers due to question wording that may be interpreted as implying English as the respondent's native language. However, there is an absence of research on how ASL users respond to such questions. Participants were divided into two groups: one group saw language use questions from a current U.S. Census Bureau sponsored survey, and the other saw alternative question wording. Participants were later probed on their understanding of the questions, and preference between two designs. This research investigates whether users of different languages interpret questions about language use in the same way. Particularly, we investigate: (i) whether "speak a language" or "use a language" is understood better by users of ASL, and whether changing "speak" to "use" changes how English and Spanish speakers respond; (ii) whether providing examples of different languages as potential response options affects how users report the language(s) they use at home, and whether the languages provided have an effect; and (iii) how users of ASL self-report when asked how well they speak English. The findings from this research will provide valuable insight into how language use questions are interpreted by speakers of different languages and what effect that has on survey data collection.

How Do You Know They Did That? Exploring Proxy Response Strategies in a Survey of Participation in the Arts

Alfred Tuttle, *U.S. Census Bureau*
Angela O'Brien, *U.S. Census Bureau*
Jonathan Katz, *U.S. Census Bureau*

Proxy responses are survey responses collected from one household member about or on behalf of another. Although accepting proxy responses in household surveys can improve response rates and efficiency of data collection, data quality may suffer, such as when the proxy respondent has limited knowledge about others in the household. Research on the congruence of proxies' and target respondents' answers has found that agreement of responses can be affected by various factors, including proxy respondents' social distance from the target respondent, proxies' motivation and level of cognitive effort, and characteristics of the survey (topic, task complexity, etc.). To explore the issue of data quality resulting from proxy response, we incorporated an analysis of proxy response strategies into a small-scale pretesting evaluation of a national CATI/CAPI survey. We conducted cognitive interviews with 18 pairs of related and unrelated participants who share a household. In addition to assessing respondents' comprehension of the survey questions, we also probed proxies' strategies for formulating responses to questions about target respondents, i.e., whether they drew on direct or indirect knowledge, or developed responses based on assumptions, reasoned inferences, etc. We examined the strategies used by proxies in various relationships with target respondents (spouse/unmarried partner, parent-adult child, siblings, unrelated housemate). We also examined the relationship between question topic/format and proxies' response strategies. We used the level of agreement between proxies' and target respondents' answers to the survey questions as an indicator of the quality of proxies' responses.   In this presentation we will present the results of our qualitative analysis of the paired cognitive interviews. We will discuss the response strategies used by proxies in various types of relationships and for different kinds of survey questions, and recommendations for maximizing data quality when using proxy respondents.

Mix and Match: Question Structure and Self-Proxy Agreement Across Household Type

Angie O'Brien, *U.S. Census Bureau*
Dave Tuttle, *U.S. Census Bureau*
Jonathan Katz, *U.S. Census Bureau*

Within household proxy reporting, or relying on respondents to report on the attributes and behaviors of other household members, is often used by survey designers because it is cheaper and more efficient than collecting information from the entire household. However, proxy reports tend to be less reliable than self-responses because they rely on respondents to both know and accurately report information on other household members, often without being able to consult them. Thus, it is essential to understand the potential pitfalls with certain types of questions and household relationships when designing questionnaires. In this presentation, we describe an exploratory study of self-proxy agreement through data collected in a CATI/CAPI survey pretest during late 2020/early 2021. This pretest consisted of cognitive interviews with 18 pairs of household members – some related and some unrelated-- about their participation in the arts. Each interview employed various question structures (e.g., open-ended numeric, closed-ended frequency, closed-ended select-all) and reference periods (e.g., past 12 months, past 24 months, lifetime) to solicit reliable data on individuals' artistic activities and interests to varying degrees of success.   In this presentation, we explore self-proxy match rates by question structure, reference period, and relationship type using a variety of matching techniques (e.g., yes/no, quartile). Preliminary results indicate higher self-proxy match rates from close-ended questions with fewer response options than select-all or fill-in-the-blank questions. Additionally, results show lower self-proxy match rates in questions using with longer reference periods, regardless of the type of relationship between household members. Furthermore, the start of the 2020 pandemic proved to be an anchoring point in the cognition process of all respondents when answering questions. Finally, the data indicate higher self-proxy agreement when reporting a lack of an event or a behavior than the presence of one, irrespective of the question structure or the relationship between respondents.

Say What You Mean to Say, Talk About What You Want to Talk About
Steven Snell, *Goldman Sachs*
Carol Sue Haney, *Qualtrics*

Efforts to improve survey data usually focus on steps that the researcher must take, but how can we better involve the respondent in ensuring data quality? We build on the growing literature on adaptive survey design to better understand the implications of making a survey unfold to respondents, adapting questions based on previous answers in order to elicit more and better information. Specifically, we explore what happens when we give respondents an opportunity to amend, review, and augment their answers. This research is facilitated by new, but underutilized advances in survey platforms and artificial intelligence that make it easier than ever to give respondents access to review and amend their responses before submitting them. We leverage two original surveys fielded across distinct audiences to better understand which respondents modify or expand their responses and in what context. The findings of this research have implications for how to better engage survey respondents, thereby mitigating against satisficing and inattention and ultimately promoting data quality.

Conceptualizing Unit Culture: Interviewing 500 Soldiers to Understand Culture in the US Army
Matthew McDonough, *ICF*
Steven Aude, *ICF*
Chelsey Jackson, *ICF*
Jennifer McLeer, *University of Hartford*
Gregory A. Ruark, *U.S. Army Research Institute for the Behavioral and Social Sciences*

The Army has framed its understanding of culture comparatively, training Soldiers on the enduring values, beliefs, and norms of groups that they might engage with outside the United States during deployment or missions. What remains to be known is what the enduring values, beliefs, and norms of Army units themselves are and how consistent or different this culture is across Army units. To conceptualize Army unit culture, the Army Research Institute (ARI) established a multidisciplinary research team (including expertise in anthropology, organizational psychology, and sociology). This presentation will describe the methods used by this team. Guided by grounded theory and multidisciplinary culture literature, the team developed semi-structured interview protocols to investigate unit culture. Thirteen researchers assigned to one or more four-person teams interviewed Soldiers from nine brigades located throughout the continental United States (CONUS) during data collections in 2017 and 2018. Each team spent up to 1 week with each brigade, where they conducted approximately twelve to twenty 1-hour interviews per day (three to five interviews per interviewer). In total, the research team conducted 504 inductive, semi-structured interviews with a

sample of enlisted, non-commissioned officers (NCOs), and officers that was generally representative of each brigade. The research team used validity procedures that are appropriate for grounded theory research, including triangulation, researcher reflexivity, prolonged engagement, audit trails, and peer debriefing. Additionally, to enhance consistency of interviewer practices and facilitate analytical comparability, researchers generated team-wide interview practices about how to pose interview questions and how to use on-the-fly probes to invite additional detail in answers. Findings from this work can guide unit commanders, leaders, and unit members to understand and describe their own unit's culture together with identifying the components and processes that reinforce and sustain that culture.

Data Collection in Low- & Middle-Income Countries

Do Referrals Improve Representation of Women in RDD Surveys in Low-Income Countries?

Steven Glazerman, *Innovations*
Karen Grepin, *University of Hong Kong*
Valerie Mueller, *Arizona State University*
Michael Rosenbaum, *Innovations for Poverty Action*
Nicole Wu, *University of Hong Kong*

Mobile phone ownership has grown dramatically in low-income countries, making approaches like Random Digit Dial (RDD) surveys increasingly feasible as alternatives to in-person sample recruitment. However, coverage bias remains a major risk to the representativeness of survey research in such settings, particularly when the focus is on women. We explore the tradeoffs associated with asking men to refer eligible women from their household by passing the phone. Referrals create the possibility of reaching women who do not own or carry phones themselves, but excludes women whose husbands, brothers, fathers, or sons mistrust them or otherwise harbor reservations about letting them speak to an interviewer. This study builds on an RDD survey on the impacts of COVID-19 on women in Kenya. We compare the characteristics of women we reached directly via RDD with those of women interviewed through referral by a male household member who himself was reached via RDD. We use a nationally representative reference dataset to show that the referral protocol does improve outreach among vulnerable groups, such as young women, the asset poor, and those living in areas with low connectivity. Among mobile phone users, we show a referral (rather than a direct dial) protocol includes more nationally representative proportions of women with these attributes. While adding a referral protocol improves representativeness of women and other subpopulations, we also show how much more expensive it is per complete than simply dialing more numbers. We explore the implications for recommended survey protocol design and offer practical suggestions for survey researchers seeking to use RDD to construct representative samples of women and other subpopulations that are underrepresented among mobile phone users. [Authors are listed in alphabetical order]

Adopting Computer-Assisted Telephone Interviewing (CATI) for Health Surveys in Latin America and the Caribbean

Alexander Nisetich, *RTI International*
Luis Sevilla, *RTI International*
Roberta Caixeta, *Pan American Health Organization*
Dolores Ondarsuhu, *Pan American Health Organization*
Carolina Chavez Cortes, *Pan American Health Organization*
Charles Lau, *RTI International*
Victoria Dounoucos, *RTI International*
Raquel Ovalle Romero, RTI International

Face-to-face interviewing (FTF) has been the main mode of data collection for health surveys in Latin America and the Caribbean. The COVID-19 pandemic forced public health experts to adopt computer-assisted telephone interviewing (CATI) in place of FTF. In the past 18 months, CATI surveys have proliferated throughout the region. Looking to the future, researchers are exploring how CATI surveys complement FTF surveys for public health surveillance. To understand how to best deploy CATI surveys, it is useful to understand how CATI and FTF samples differ across countries, and how to improve the quality of CATI surveys. In this paper, we study where and how CATI can collect nationally representative samples for public health research. We examine the results of a 2021 survey conducted for the Pan-American Health Organization in five countries (Antigua and Barbuda, Belize, Bolivia, Honduras, and Paraguay) on COVID-19 and non-communicable diseases. Respondents answered a 10-minute questionnaire about how COVID-19 has affected their lifestyle, risk behaviors for COVID-19 and non-communicable diseases, and use/access to healthcare and other services. Using random digit dialing (RDD), the survey interviewed 2,000 adults per country, age 18 and over with access to a mobile phone. We pursue two objectives. The first objective is to understand the representativeness of CATI surveys, comparing the CATI sample composition in each country with benchmark demographic data separately by country.

This analysis sheds light on which countries are best suited for CATI surveys, given that the five countries differ with respect to mobile phone penetration, income, and educational levels. The second objective uses paradata from the five countries to understand how to improve demographic representation in CATI surveys. Specifically, we investigate how the number of call attempts affects sample representativeness, describing how the sample composition shifts across contact attempts. We conclude with recommendations and directions for future research.

Asking Ugandan Adolescents about Sex Trafficking: Challenges and Advantages of Audio Computer-Assisted Self-Interview (ACASI)

Holly Koogler, *ICF*
Matt Jans, *ICF*
Suteera Nagavajara, *ICF*
Kato Francis, *Makerere University*
Laban Musinguzi, *Makerere University*
Eddy Walakira, *Makerere University*
Bart Robertson, *Global Fund to End Modern Slavery*

Sexual topics are particularly sensitive for minors. Using audio computer-assisted self-interview (ACASI) helps reduce sensitivity by creating a relatively private response mode and may increase accuracy and foster inclusion in data collection. However, this data collection method is rarely used, especially in rural contexts or areas with low technology exposure, due to perceived complexity and effort. This presentation describes a simple and practical ACASI application within a population survey of commercial sexual exploitation prevalence among rural Ugandan adolescents in the Napak District of Uganda's Karamoja Province. Sampled households were screened for adult and adolescent respondents, and up to two adolescents aged 12 to 17 were randomly selected within each household. Following a face-to-face interview on less-sensitive topics, the ACASI instrument included approximately 35 questions on parent-child relationships and sexual exploitation (about 40% of the total questions asked of adolescents). The ACASI instrument was implemented on the same tablet that the interviewer used, and it was designed for adolescents with no or low literacy or technology exposure. Respondents heard audio of each question read by a native speaker and were trained to select their responses by associating their answers with an image, which they touched on the screen. The final sample included 830 adolescents within 986 households. This presentation discusses the preparation and implementation of the ACASI instrument including challenges tailoring the program to local culture, technical challenges designing the program, ethical considerations, and overall lessons learned. To our knowledge, this is the first ACASI application designed specifically for children with low literacy and low exposure to technology. Insights gained will guide future sensitive data collection efforts in challenging settings and help promote inclusion and equity in surveys.

Representativeness in RDD Surveys in 10 LMICs

Shana Warren, *Innovations for Poverty Action*
Elliott Collins, *Innovations for Poverty Action*
Savanna Henderson, *Innovations for Poverty Action*
Michael Rosenbaum, *Innovations for Poverty Action*

The COVID-19 pandemic forced an abrupt shift in the way social scientists gather survey data in low- and middle-income countries (LMICs), leading many researchers to implement remote survey methods to recruit study participants. Random Digit Dialing (RDD) presents an opportunity to quickly build survey samples that include a wide range of respondents, even in populations with limited internet connectivity. However, such samples are subject to selection bias due to both coverage limitations and response bias. In this paper we present estimates of representativeness in samples obtained via random-digit dialing during the first months of the pandemic based on surveys conducted in nine LMICs. We consider the extent of selection bias for both demographic and socio-economic variables, and then assess the performance of simple statistical adjustment to correct for biases in these samples. We find that RDD respondents are disproportionately young, urban, wealthy, and more educated than the typical individual, resulting in estimates of employment and durable asset ownership that are not likely to be nationally representative. Simple post-stratification adjustment did little to improve representativeness of those estimates. Further work will expand data sources, variables considered, and methods for statistical adjustment to determine the robustness of these findings.

Sampling Outside the Lines: Does Passing the Phone to a Non Phone-Owning Household Member Improve Health Estimates in Uganda?

Abigail Greenleaf, *ICAP at Columbia University*
Sam Biraro, *ICAP - Uganda*
Ashley Crocker-Benn, *Department of Population and Family Health, Columbia Mailman School of Public Health*
Veronicah Mugish, *ICAP - Rwanda*
Dorothy Aibo, *ICAP - Uganda*
Qixuan Chen, *Department of Biostatistics, Columbia Mailman School of Public Health*

The expansion of mobile phone ownership in low and middle-income countries (LMIC) provides a platform for low-cost, rapid data collection. Mobile phone surveys are increasingly used to collect data for myriad topics in LMIC. Yet, in many countries, coverage error impedes the creation of valid estimates. In February 2022, we will call a random sample of respondents who participated in the Uganda Population-based HIV Impact Assessment (UPHIA 2020) study, a nationally representative survey in Uganda conducted from February 2020 to March 2021, to evaluate the efficacy of a sampling approach to reducing coverage error, and to generate estimates of key health outcomes. Phone ownership in Uganda is approximately 70%, with 90% of people in urban areas reporting ownership compared to 63% in rural areas. Offered in six languages, the phone survey will include a sample of 800 primary respondents 18 years or older who participated in the UPHIA survey, consented to future research and provided a valid mobile phone number. All primary respondents will be asked either to pass the phone to a "secondary" respondent in their home who consented to follow-up during UPHIA but does not own a phone or to pass the phone to a household member who was not present during UPHIA and does not own a phone. Using both the UPHIA and phone survey data, we will compare key health behaviors between UPHIA and phone survey respondents to establish if the estimates created in the phone survey match the UPHIA reference standard. We will also assess whether secondary respondents make the phone survey sample more representative of the reference population than a sample of solely primary respondents. The results of this analysis will inform the sampling of future phone surveys in LMIC

Cracked Foundations: Societal Problems Laid Bare by the COVID Pandemic

COVID-19 and the Experiences of Populations at Greater Risk

Katherine Carman, *RAND Corporation*
Carolyn Miller, *Robert Wood Johnson Foundation*
Anita Chandra, *RAND*

Since 2013, the Robert Wood Johnson Foundation has led a pioneering effort to advance a Culture of Health to enable "all in our diverse society to lead healthier lives, now and for generations to come". Now, in the context of COVID-19, RAND and RWJF partnered to build from the NSHA to implement a longitudinal survey to understand how these health views and values have been affected by the experience of the pandemic, with particular focus on populations deemed vulnerable or underserved, including people of color and those from low- to moderate-income backgrounds. The questions in this COVID-19 survey focused specifically on experiences related to the pandemic (e.g., financial, physical, emotional), how respondents viewed the disproportionate impacts of the pandemic, whether and how respondents' views and priorities regarding health actions and investments are changing (including the roles of government and the private sector), and how general values about such issues as freedom and racism may be related to pandemic views and response expectations. We collected data from the same group of 3,500 respondents four times between June of 2020 and October of 2021, a period which captured the introduction and removal of lockdowns and mask mandates, high rates of spread and low rates of spread, and great attention to racism in our society. In this paper we present results about how perceptions of systemic racism and perceptions of the impact of COVID-19 on people of color changed over the course of the pandemic. In our survey of vulnerable and underserved groups, we found that over time there was a trend away from recognizing the role of systemic racism across a number of questions. The latest wave of data collection just ended (in October 2021) and in our paper we will explore the groups that had the biggest changes in views.

Measuring Risk Factors for Intimate Partner Violence and Impacts of the COVID Pandemic in the United States

Kelly Bell, *Ipsos*

Nataly Johanson, *Ipsos Public Affairs*
Kelly Bell, *Ipsos Public Affairs*
Zachary Lewis, *Ipsos Public Affairs*

Intimate partner violence (IPV) is an important issue in the United States. The US Department of Justice reports that one out of every four women will experience sexual violence, physical violence, and/or stalking by an intimate partner during their lifetime [1] . Since the onset of the COVID-19 pandemic, researchers have found that domestic violence incidents increased 8.1% after jurisdictions-imposed pandemic-related lockdown orders [2] . This paper will augment existing research and federal statistics on IPV by measuring known risk factors for IPV (including mental health challenges, stress levels, lack of social support, substance abuse, low relationship satisfaction, and frequent relationship conflict [3]) and assess the impact of COVID-19 on these risk factors. Further, we will examine risk factors across demographic, behavioral, and socioeconomic characteristics. The data are derived from a custom survey fielded using an online probability-based panel (Ipsos KnowledgePanel) and are generalizable to the US adult population. This study is unique among research examining IPV in the context of COVID because it employs primary data collection via a survey of the general public. Much of the existing literature exploring IPV and COVID have relied on administrative data (e.g., police logs, hotline calls) and have focused on estimating changes in incidence rates for IPV. This paper offers a unique contribution by exploring Americans' attitudes, perceptions, and experiences with IPV risk factors and the COVID pandemic. This paper will serve to increase the knowledge and skills of scholars and industry professionals by providing context for interpreting changes in IPV incident rates in recent years and offer recommendations for further studies of IPV-related phenomena in the US.   [1] DOJ, 2014: https://bjs.ojp.gov/content/pub/pdf/ndv0312.pdf [2] CCJ, 2021: https://covid19.counciloncj.org/2021/02/23/impact-report-covid-19-and-domestic-violence-trends/ [3] Capaldi, Knoble, Shortt, and Kim, 2012: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3384540/

Commuter Behavior Change Factors and Processes Due to the Covid-19 Pandemic

Alec Biehl, *Oak Ridge National Laboratory*
Michael Hilliard, *Oak Ridge National Laboratory*

The Covid-19 pandemic spurred extensive survey research to identify the myriad of impacts from the disease itself and resultant government policies. One such survey is the COVID-19 and the Future Survey , a joint effort by researchers from Arizona State University and the University of Illinois Chicago. Wave 1 of the survey, released publicly in March 2021, contains 8,723 responses regarding actual and anticipated changes in travel behavior and activity participation, along with several psychological measurement scales and standard individual characteristics. A team of researchers from Oak Ridge National Laboratory utilized data from this survey, among other sources, as part of a rapid response project funded by the U.S. Department of Energy to capture the effects of the Covid-19 pandemic on transportation choices using the CommuterSim tool. More specifically, a 1,465-respondent subsample representing the metro areas with the five largest public transit systems—Boston, Chicago, New York City, San Francisco, and Washington, D.C.—was analyzed to understand (a) before, (b) during, and (c) after "mobility profiles." Thus, the proposed presentation will feature empirical findings addressing the following two questions. First, what information regarding attitudes, risk perceptions, and so forth, is captured using appropriate factor analytic techniques? Second, what role do the extracted latent variables play in a Markov model of changing commuting behavior, after controlling for important socio-demographic and geographic features? Moreover, the integration of findings into the broader simulation framework will be briefly addressed, particularly with respect to equity implications

Political Identity and Risk Politics in America: Evidence from a Pandemic

Eric Raile, *Montana State University*
Pavielle Haines, *Rollins College*
Amber N. W. Raile, *Montana State University*
Elizabeth A. Shanahan, *Montana State University*
David C. W. Parker, *Montana State University*

Risk perceptions have become politicized and polarized in the United States as an element of the "culture war." This political amplification of risk, which is similar to social amplification of risk, has serious consequences for public opinion, governance, and public policymaking. We propose an alignment, particularly over time, of political identities,

information sources, support for political elites and policies, and risk positions as a consequence of the political amplification and politicization of risk. Our more specific research questions and hypotheses are related to whether and how this alignment may have occurred. We conducted a survey in four Western U.S. states (Montana, Colorado, Utah, and North Dakota) in April 2020, early in the coronavirus pandemic. We conducted another survey in Montana in September 2020. The questionnaires asked about the topics necessary to sort out these relationships, including questions about views on stay-at-home orders, economic collapse, face coverings, and vaccines. We also asked a series of more complex questions about political identities. We supplement with data about the information environment during this period. We use structural equation modeling to assess the complex pattern of relationships among variables. While some of our analyses are confined to a single point in time, we also have data that allow for assessing changes over the April–September 2020 timeframe. Additionally, the April survey included a messaging experiment that allows for more careful evaluation of causality. The experiment randomly varied messages about wearing cloth face masks in public, with President Trump as an information source. Results show partisan-motivated risk perception that fits with the notion of public opinion leadership. Other initial results of the broader study support the idea of political factors and risk positions aligning at the individual level, as well as changing to align with information sources and political affinities.

Current & Historical Predictors of Abortion Attitudes
The Role of Racism and Sexism in Attitudes Towards Abortion in White, Latinx, and Black Individuals

Majel Baker, *University of Michigan*
Kristen N. Jozkowski, *Indiana University*
Sara I. McClelland, *University of Michigan*

Attitudes towards abortion play significant historical and contemporary roles in U.S. politics. While religious and political opinions are important predictors of abortion attitudes, we argue that other social attitudes—specifically "new and subtle" forms of anti-Black racism and sexism—could be predictive of abortion attitudes as well. In Study 1, we demonstrate the novel finding that individuals who endorsed more symbolic racism and modern sexism did, indeed, endorse more opposition to abortion in a sample of Black, Latinx, and White U.S. residing individuals (N=1,060). These relationships were true above and beyond religious factors, indicating the opposition to abortion is not simply a religious attitude as might be presumed based on prior literature and theory. In Study 2, we conducted a conceptual replication of these relationships using the 2012 ANES data (N=3,860) in order to: (a) test robustness of these novel findings using different measures of symbolic racism, modern sexism, and abortion; and (b) extend these findings by controlling for political ideology (liberal–conservative) in addition to religious factors. Preliminary results showed that the relationships between more subtle racism and sexism and opposition to abortion replicated in the ANES online sample of Black and White voters and partially replicated in the sample of Latinx voters. These results confirm our hypotheses that racist and sexist attitudes relate to individuals abortion attitudes for both the religious and the non-religious. Future analyses for the presentation will include examining these relationships alongside political ideology. These two studies demonstrate that abortion attitudes are related to more than just religious or political attitudes. Without this insight, researchers and policy makers risk interpreting data from opinion polls without a full picture of what individuals might be drawing on when voting, answering poll questions, or supporting political candidates.

Migrant generations and Abortion Circumstances: A Close Look at Latinx's Abortion Attitudes in the US

Xiana Bueno, *Indiana University*
Montenegro, María, *Indiana University*
Lo, Wen-Juo, *University of Arkansas*
Valdez, Danny, *Indiana University*
Crawford, Brandon, *Indiana University*
Turner, Ronna, *University of Arkansas*
Jozkowski, Kristen, *Indiana University*

Many Latin American countries are governed by abortion laws that are more restrictive than those in the United States. When moving to the U.S., integration processes into the new social context might progress at different paces among first-generation migrants while perhaps coexisting with beliefs and views conformed in their country of origin. Similarly,

second and subsequent generation migrants -born and socialized in a different context- might adopt views on social issues, such as abortion, of their parents' generation. Given the dual cultural context inherent to families of migrant origin, and under the general assumption that abortion attitudes are, themselves, complex and contextual, we explored and compared people's attitudes toward legal abortion across first, second, and third-plus generations of Latinxs population in the US. A total of 1,183 self-identified Latinxs was selected from a larger data pool collected through Qualtrics using quota-based sampling. We applied a Multiple Indicators Multiple Causes (MIMIC) approach to investigate the effects of Latinxs' generation, sex, education level, religious attendance, and political affiliation on a 7-item measure of abortion attitudes comprising a two-factor model similar to the abortion items from the General Social Survey (i.e., 'hard' and 'soft' reasons; Benin, 1985). Results indicated abortion attitudes across Latinx adults in the US from different migrant generations are not homogeneous, but sensitive to the reasons motivating abortion. There were no significant differences across generations for 'hard' circumstances (i.e., woman health risk, rape, fetal anomaly). However, there were significant differences in abortion support across Latinx migrant generations in relation to abortion motivated by social-related circumstances known as 'soft' reasons. Our findings lend support to previous research which suggests complexity in abortion attitudes by demonstrating differences in attitudes across generations and other demographics. We will discuss implications for research involving culturally diverse populations in light of their migrant origin or ancestry.

Are there Racial/Ethnic Disparities in Support for Abortion Access? Exploring the Role of Abortion Law Knowledge and Political Engagement

Xiana Bueno, *Indiana University*
Bueno, Xiana, *Indiana University*
Ezike, Nnamdi, *University of Arkansas*
Lo, Wen-Juo, *University of Arkansas*
Turner, Ronna, *University of Arkansas*
Crawford, Brandon, *Indiana University*
Jozkowski, Kristen, *Indiana University*

Given the social and health implications that unequal access to reproductive health has, it is important to explore whether a link exists between race-ethnicity and different factors involved in explaining beliefs towards abortion access. Little is known about whether racial-ethnic disparities within political engagement and abortion law knowledge might impact people's support for abortion access. Thus, this study aims to examine whether and how political engagement and abortion law knowledge, may inform racial-ethnic disparities regarding support for abortion access. We administered a web-based survey in English and Spanish to a national sample of n=2,383 participants that identified as Black/African American, Latinx, or White. The outcome variable asked participants about their position regarding how easy it should be to get an abortion and whether there should be fewer or more abortion restrictions in the US and in their state. We used political engagement, abortion law knowledge, and demographic (e.g. race-ethnicity) variables as predictors. Quantile regression was used to investigate the effect of political engagement and abortion law knowledge on support for abortion access across different racial-ethnic groups. Political engagement was significantly associated with support for abortion access at the 25th and 75th percentiles. Abortion knowledge was significantly associated with support for abortion access at all three percentiles. We found significant differences in support for abortion access between Black and White participants at the 25th percentile with Black respondents providing higher endorsements compared to White respondents. There was also a significant effect on support for abortion access among political affiliations and education levels.   We conclude that the association between racial-ethnic groups and support for abortion access entails certain complexity. There are not consistent significant differences among groups, and, as shown, observed differences were reduced only to Black/African Americans (compared to Whites) and in the most polarized segments of the scale distribution.

Dimensions of Abortion Attitudes and Their Relationship to Abortion Focused Policies and Politicians

Brandon Crawford, *Indiana University*
Kristen N. Jozkowski, *Indiana University*
Beyza Buyuker, *Indiana University*
Ronna C. Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*

Abortion is a complex social issue that yields attitudes and opinions that span multiple dimensions. These dimensions may include support for abortion generally, policies and laws related to abortion, or identities related to abortion (e.g., pro-choice/pro-life). However, because different firms and surveys often use different measures, the extent that these dimensions are examined relative to each other is limited. For example, less research focuses on what types of abortion policies someone who identifies as pro-life might support, or the extent that individuals who support the legality of abortion in various circumstances also think it is immoral. Understanding the relationship between these different dimensions helps to contextualize the complexity and nuance of abortion attitudes. Furthermore, given that in 2022 the Supreme Court will rule on abortion laws in Mississippi and Texas that may drastically alter the precedence of Roe v. Wade and the abortion landscape in the United States, it is crucial that we continue to expand our understanding of the complexity of abortion attitudes and how these attitudes may translate to support or opposition for abortion-focused policies and politicians. This presentation uses survey data obtained from both quota and probability-based samples between 2017-2021 ($N_1 - 3,000$; $N_2 = 919$; $N_3 = 1,583$). The surveys included various measures of abortion identity, scenario specific attitudes about the morality and legality of abortion, and support for different abortion-focused policies and politicians. First, we will use frequencies and crosstabs to highlight how the different questions and dimensions relate to one another. Next, we will explore which questions may be better predictors of support or opposition to abortion-focused policies and politicians. We will conclude by discussing why some measures are better predictors than others of abortion focused policies and politicians and make recommendations for key considerations in assessing abortion attitudes moving forward.

A Comparison of National Historical and Personal Retrospective Trends of Perceptions About Abortion

Nana Amma Asamoah, *University of Arkansas*
Ronna Turner, *University of Arkansas*
Wen-Juo Lo, *University of Arkansas*
Kristen Jozkowski, *Indiana University*
Brandon Crawford, *Indiana University*

When using General Social Survey data (GSS, Smith et al., 2018) for attitudes about abortion legality from 1972 to 2018, trends indicate that US adults are generally supportive of abortion being legal in medical-related circumstances (e.g., health endangerment, birth defects) and split on support of abortion being legal for socioeconomic circumstances (e.g., not married, cannot afford more children). From 1972 to 1996, there are also relatively consistent age-related trends where younger age groups (18-34 years and 35-49 years) are the most supportive of abortion legality and the oldest age group (65 years and older) is least supportive. However, these trends appear disrupted from 1996 to 2018, with the exception of the oldest age group typically less supportive of socioeconomic circumstances. In this study, participants marked responses on a 5-point Likert-type scale to indicate whether they are supportive of or opposed to abortion over four timeframes in their life: now, one year ago, in early adulthood (18-25 years of age), and in adolescence (13-17 years of age). Data using a GfK panel in 2019 (n=1583) and an AmeriSpeak panel in 2021 (n=968) are presented. Relevant age groups indicate significantly larger percentages of participants that are more supportive of abortion now than in their adolescence. In regard to changes in attitude since young adulthood, approximately one-third of the two older age groups (≥ 45 years) indicate they are either more or less supportive with approximately equal proportions of change in each direction. For the two younger age groups (< 45 years) a significantly larger percent indicate they are more supportive of abortion now than in their young adulthood, and even one year prior. Further comparisons will be presented to investigate how the retrospective results that indicate more US adults believe they are becoming more supportive of abortion, aligns with historical data trends.

Devices & Technology to Enhance Data Collection

Are Respondents on the Move when Completing a Mobile Web Survey? Evaluating Response Context through Combined Data Sources

Caroline Roberts, *University of Lausanne*
Jessica M. E. Herzing, *Interfaculty Centre for Educational Research, University of Bern*

Mobile devices are designed to be used while people are on the move, meaning they offer unique opportunities for in-the-moment data capture with huge potential to augment traditional surveys. By combining a mix of passively collected sensor data with active multimodal tasks, the hope is to better engage respondents, minimise recall biases associated with answering questions about events after the fact, and shed light on innovative research problems. However, it is not clear how willing respondents are to complete tasks while out and about and doing so may expose them to additional sources of distraction affecting whether and how they continue responding. Both these challenges represent significant potential sources of selection and measurement error in estimates derived from mobile data. This study investigates to what extent mobile respondents are more likely than PC respondents to complete questionnaires while on the move, and whether the possibility to respond via a smartphone app is associated with greater mobility than responding via a browser. Using a combination of self-report data, paradata and respondent photographs gathered in a methodological experiment testing an app-based survey alongside a browser-based design in a probability-based sample of the general population, we assess: 1) the time and location of survey participation by device and software; 2) whether location changes between starting and terminating questionnaire completion, and 3) whether differences in response context are related to completion interruptions. We find only a minority of respondents was on the move, but differences by device in the time-of-day questionnaires were completed, mobility and completion interruptions. App users tended to complete the survey at home, even for tasks they were asked to complete while mobile, highlighting the need to optimise onboarding procedures and task requests to facilitate mobile completion if this is a core study design feature.

Measuring Physical Activity in the General Population with Hybrid Data: Data Donation from Activity Trackers, Research-Grade Accelerometers and Survey Data

vera toepoel, *Utrecht University*
Annemieke Luiten, *Statistics Netherlands*
Anne Elevelt, *Statistics Netherlands*
Inge de Wolf, *Amsterdam University Medical Center*

National organizations typically measure PA in the population through surveys, however, surveys might not be the best method. Measuring PA is difficult because it requires remembering activities, estimation about intensity and duration of specific types of activities, complex calculations and intensity of PA is a relative issue. Furthermore, questions on PA are susceptible to social desirability bias. Accelerometers – research grade and consumer owned like smartwatches and Fitbits- can be used to augment or replace surveys. This presentation investigates how the use of accelerometers relates to survey measures when it comes to Total Data Error. Data come from the Dutch LISS Panel, a probability-based online panel consisting of about 8000 individuals. The panel was asked to complete the Dutch National Survey on Physical Activity, as well as questions on ownership on activity trackers, donation of activity tracker data, and willingness to wear a research-grade device in a follow-up study. Survey data were collected in October 2021. The follow-up study, where participants wore the research-grade accelerometer ActivPal for a week, will be conducted between November 2021-Februari 2022.Results show that about half of the participants abide to the norms of Physical Activity as determined by the World Health Organization. About half of the panel owned an activity tracker such as smartwatch of Fitbit. About half of that group was willing to donate their data. In our presentation we will focus on predictors of device use and physical activity based on covariates from the panel (analyses still to be done). In addition, we will report on the results of the accelerometer, and look at validity between survey data, research-grade accelerometer and personally owned activity trackers. This presentation shows how to use different types of sensor data to augment or replace survey questions. It informs on hybrid data opportunities in social research.

What's in a Name? Further Experiments in Increasing CATI Contact Rates

Missy Mosher, *Dynata LLC*
Jesse Reinhold, *Dynata LLC*

It is a well-known fact that contact and response rates for telephone surveys have decreased severely in the past number of years. Not only is it difficult to get a call connected in the first place, there is the added difficulty of getting the call answered (let alone getting the called party to agree to take a survey). For those in the research industry who still stand by the tried and true standard of telephone probability sampling, it is imperative that ways are found to increase contact and response rates in order to minimize bias and keep costs under control. In 2019, Dynata presented some

results of a test where outgoing phone numbers were 'whitelisted' in order to help them connect to potential respondents which met with mixed results. In 2021, we now focus our research on caller ID displays. We ran a test of two services, Hiya and First Orion (which are associated with two of the major phone carriers), to see if changing what is displayed as a caller ID will have any effect on respondent behavior. Even though these experiments are not novel and many have tested them before, what is constantly shifting is the legal landscape and what telephone carriers and the public itself are doing in response to those changes. For example, the Stir/Shaken legislation meant to verify valid calls just went into effect on June 30, 2021. Dynata has been monitoring the consistency of the verified check mark required by Stir/Shaken and can share some anecdotal results on what we are seeing as an addendum to our presentation. Attempts by the carriers to become compliant with new laws and/or the public's understanding of such laws will likely have a huge and ongoing effect on respondent behavior that begs for continued exploration.

Not Home? No Problem: Capturing Electronic Authorization for Multiple Household Members on MEPS
Ryan Hubbard, *Westat*

In addition to a CAPI interview centered on health care utilization and cost, data collection for the household component of the Medical Expenditure Panel Survey (MEPS) includes the collection of hard-copy authorization forms that allow for critical follow-up with pharmacies and providers. Traditionally, at the time of the household interview, respondents and other present household members are asked to sign paper forms. Forms for other household members are left along with instructions. Interviewers then engage in a time-consuming process of reminder calls, and if the forms are complete, they engage in costly travel to retrieve the forms, increasing burden for the household as well. The completed forms are then processed and scanned for use in the MEPS medical provider component to obtain detailed and accurate health care cost information. In an effort to modernize collection, reduce interviewer and respondent burden, and reduce processing time, MEPS is introducing two forms of electronic signature collection. First, respondents and household members present at the time of the interview will be asked to sign electronically on the laptop using a stylus. For household members not present, they will be contacted by email or text (if permitted) via DocuSign to complete the forms via electronic signature on a computer, phone, or tablet. While interviewers will continue to follow up by phone, the DocuSign system will also send reminders. Paper forms are then reserved for household members not present for the interview for whom there is no contact information. These methods allow for direct communication with household members other than the household reporter, helping to promote cooperation. The paper details the outcomes associated with the first four months of authorization form collection with regard to indicators of perceived interviewer and respondent burden*.*

Device Usage on the Internet Instrument of the American Community Survey
Gregory Mills, *U.S. Census Bureau*
Lindsay Longsine, *US Census Bureau*
R. Chase Sawyer, *US Census Bureau*

The internet instrument was first used in production for the American Community Survey in 2013. At that time, it was estimated that around 95 percent of logins were from personal computers (PCs), while 5 percent were from tablets and mobile phones. These figures were based on a 2011 ACS Internet Test. The purpose of this paper is provide an updated analysis of general device usage trends on the ACS, as much time has passed since such a systemic analysis was performed, and in that time, usage trends of PCs, tablets, and mobile phones in wider society have changed considerably. We use ACS internet instrument paradata, which gives some details about the device type and operating system used by a respondent, paired with case disposition and response data. We first explore the prevalence of various device types and operating systems using the ACS internet instrument over time. We then examine some behavioral patterns that are associated with device types, such as number of logins, device switching patterns, quickness of response after the survey invitation, and likelihood of completing the survey in a single login. Finally, we analyze some demographic characteristics associated with usage of each device type using a logistic regression model. Plans for additional research into device usage on the ACS will also be discussed.

Not Answering the Phone? The Door? Shifting from Interviewer-Administered to Self-Administered Surveys

Evaluating a Web/Mail Alternative to a National Face-to-Face Survey: Initial Results from the American Family Health Study

Brady West, *University of Michigan*
Mick P. Couper, *University of Michigan*
William G. Axinn, *University of Michigan*
James Wagner, *University of Michigan*
Rebecca Gatward, *University of Michigan*
Htay-Wah Saw, *University of Michigan*
Shiyu Zhang, *University of Michigan*

The rising costs of large-scale face-to-face data collections, when combined with rising survey nonresponse rates and general reluctance of the public to participate in time-consuming in-person data collections, are threatening the future of national face-to-face surveys. The COVID-19 pandemic has further eroded researchers' reliance on face-to-face surveys. These trends create an urgent need for the evaluation of innovative web-based data collection methods that are convenient for the general public and yield high-quality scientific information for population researchers. The development of an alternative data collection methodology is especially important for research teams without the resources of larger government agencies. The web mode is particularly appealing because it is relatively inexpensive and affords a high level of privacy and confidentiality when correctly implemented. We present initial results from a sequential mixed-mode web/mail data collection approach that was implemented on a national probability sample in 2020-2021 and compare them to results from a benchmark face-to-face national survey of population reproductive health: the 2017-2019 National Survey of Family Growth (NSFG). This allows us to test maximum design complexity, including a complex household screening operation (to identify households with persons aged 18-49). We demonstrate the ability of this national web-based survey that employs address-based sampling, sequential mixed-mode design, and modular survey design (i.e., completing a survey in several short sessions rather than a single long session) to 1) effectively replicate key 2017-2019 NSFG population estimates across a wide range of fertility and family formation topics (including contraceptive use) and 2) capture notable changes in estimates based on measures likely affected by the COVID-19 pandemic, at a significantly lower cost and within a shorter timeframe.

Comparing the Telephone to Web Data Collection Transition Between the 2021 NYC Community Health Survey of Adults and the 2021 NYC KIDS Survey

Martha McRoy, *Abt Associates*
Michael Witt, *Abt Associates*
Stephanie Zimmer, *Abt Associates*
Theresa Stroble, *Abt Associates*
Nicholas Ruther, *Abt Associates*
Amber Levanon Seligson, *New York City Department of Health and Mental Hygiene*
Steven Fernandez, *New York City Department of Health and Mental Hygiene*
Ariana Annibale, New York City Department of Health and Mental Hygiene
Michael L. Sanderson, Utah Department of Health

The NYC Community Health Survey (CHS) and KIDS Survey, key health surveillance surveys of NYC adults and children conducted by the NYC Department of Health and Mental Hygiene (DOHMH), transitioned from telephone-based sample designs to multimode list frame/ABS designs in 2021. The CHS is an annual survey that collects self-reported data from ~10,000 randomly selected adults that enables researchers to track the health of New Yorkers and measure the extent to which diseases, including COVID-19, and risk factors occur at the neighborhood, borough and city levels. The KIDS Survey collects data every other year from a parent/guardian on the health of 7,500 children randomly selected in NYC between the ages of 1 and 13 and provides data to understand and address the health care needs of children. The surveys were transitioned to optimize geographical targeting, provide additional stratification options, improve response rates, reduce data collection costs and to expand analytic options. The 2021 data collection protocol for both surveys involved mail solicitations asking respondents to complete the survey via web followed by a paper survey

mailing for CHS and a telephone follow-up effort for KIDS to further minimize nonresponse bias. Bridge studies for both surveys were conducted in 2021 using the prior design methodologies to measure the effect of design changes. This presentation will summarize the mode changes for both studies and their impact on the efficiency of collecting data and final survey estimates. The presentation will compare the effects of mode changes for similar questions between the CHS and KIDS surveys, for example general physical health status or receipt of mental health counseling, highlighting different mode effects seen across the studies. These differences may be attributed to different target populations (adult vs child), a different subject person (self-report vs proxy), or differences in the data collection protocols.

How the California Health Interview Survey Adapts Sampling and Data Collection To Better Include Underrepresented and Underserved Populations

Royce Park, *University of California - Los Angeles*
Todd Hughes, *UCLA Center for Health Policy Research*
Jiangzhou Fu, *UCLA Center for Health Policy Research*
Ninez Ponce, *UCLA Center for Health Policy Research*
Susan Sherr, *SSRS*

For a large and complex survey like the California Health Interview Survey (CHIS), many factors were carefully considered for the 2019 transition from the use of random-digit-dial (RDD) sampling with single-mode (CATI) interviewing to address-based sampling (ABS) and mixed-mode (web and CATI) interviewing. The redesigned sampling and methodological changes were beneficial overall and outweighed any observed limitations. After two years of data collection using the redesigned CHIS, CHIS 2021 explored other sampling, operational, and data collection strategies to complement the redesigned CHIS to improve data representativeness of the underrepresented and underserved. This paper presents the various methods CHIS 2021 explored to improve data representativeness of the diverse population of California. First, CHIS explored the use of a pre-paid cell sampling frame to supplement the ABS mail push-to-web design with a CATI non-response follow-up. We examined the complementary coverage provided by these two frames. Second, CHIS enhanced our modeled-prediction approach to target mailing and interview language support to better support the non-English monolingual language speaking population of California. Third, CHIS oversampled American-Indian and Alaska Native population in the rural areas of California. And, lastly, the CHIS conducted follow-on surveys to provide robust data on the underrepresented and underserved population. Utilizing CHIS as a screening mechanism, several follow-on surveys were conducted on Latino and Asian immigrants, Asian-American, Native Hawaiian, and Pacific Islanders, and persons in need of long-term services and support.

Moving Away From RDD: Shifting to ABS While in the Field

Marcus Berzofsky, *RTI International*
Caroline Scruggs, *RTI International*
Robert Ashmead, *OSU Government Resource Center*
Bo Lu, *Ohio State University*
Tim Sahr, *OSU Government Resource Center*
Tom Duff, *RTI International*

State-based surveys, which predominantly relied on a random digit dial (RDD) frame for the past 20 years, are now in need of an alternative frame due to ever decreasing response rates and increasing data collection costs. This shift began pre-pandemic, but the past 2 years have exacerbated problems experienced with RDD-based studies: people are not answering telephone numbers they do not recognize. Additionally, cellphone providers have introduced call blockers that identify some survey call centers as spam. These issues have led many state survey researchers to an address-based sampling (ABS) frame. Sample members selected from an ABS frame are recruited through mail to obtain survey responses via mail, web or inbound calling, bypassing the issues found in RDD. However, respondents obtained through an ABS sample are generally different from those obtained through an RDD sample. While survey weights can correct for some of these differences, they cannot correct for all. Therefore, if the transition from RDD to ABS happens too quickly, multi-decade time series can be broken. As such, bridge designs – designs that utilize both frames – are often used to adjust for the impact of the shift. As recently as 2019, the Ohio Medicaid Assessment Survey (OMAS) utilized only an RDD frame achieving a response rate of 20%. An RDD-ABS bridge design was planned for 2021 based on

results from a pilot ABS sample released in 2019. However, once in the field and without additional funds available, the RDD sample was projected to fall short of the targeted number of interviews. This presentation discusses how, mid-data collection, a shift in allocation was made to assign 80% of the sample to the ABS frame. We will discuss the implications to the available response options (e.g., adding paper responses), the impact to cost, and the impact to the population estimates.

Fake News? Nope, Fake Respondents & Answers

Building Better Mousetraps: Use of Trap Questions and Data Quality

Mina Muller, *Ipsos Public Affairs*
Randall K. Thomas, *Ipsos Public Affairs*
Frances M. Barlas, *Ipsos Public Affairs*
Megan Hendrich, *Ipsos Public Affairs*

Researchers often have concerns over data quality due to inattentive or unmotivated respondents. A number of measures have been developed to assess whether respondents are behaving in sub-optimal ways and jeopardizing data quality. These measures can range from speeding through the survey, response non-differentiation, incomplete responses, or greater use of non-substantive responses (e.g., 'Don't know' or 'Not applicable'). Additional measures are called respondent traps that can indicate when someone is not paying attention. One of the most common trap is a compliance trap, whereby the respondent is directed to select a particular response (e.g. "Select 'Somewhat agree' for this item") regardless of the question. The second most common trap is a consistency trap, which presents two very similar questions at different locations in the survey, expecting that people paying attention will respond in consistent ways. We were interested in whether cleaning out respondents who fail such traps could improve data quality. In a larger omnibus study, we designed an experiment with over 5,000 completes from the KnowledgePanel, an online probability-based panel. Respondents were randomly assigned to one of three groups: 1) no trap, 2) two compliance traps, or 3) two consistency traps. In addition, we had a set of 10 demographic items for which we had benchmark values which could be used to determine the extent of bias in the sample. We looked at the correspondence of traps with speed to complete (another indicator of sub-optimal behavior). For both trap conditions, we found that higher trap failure had a modest association with faster completion times. We then examined whether eliminating respondents due to trap failure improved data quality. We found that there was no difference in average bias between the full samples in each condition, and no reduced bias when eliminating participants who failed the traps.

Detecting and Deterring Information Search in Online Surveys

Matthew Graham, *The George Washington University*

This paper introduces a framework for measuring the prevalalence of information search (``cheating") in online surveys and evaluating methods of combatting it. Three studies (total $N>14,000$) yield several key findings with implications for practice. Cheating is common, affecting 8 to 18\% of responses on average and up to 30% on some questions. Deterrence works: a pledge not to cheat reduces cheating by half. Detection also works: a browser-based method identifies 70-85% of cheaters, while 60-85% of cheaters on knowledge questions also correctly answer ``catch" questions about obscure Supreme Court cases. Detection and deterrence are complements: in combination, the three methods (pledge, browser, catch) deter or detect more than 90% of cheating, leaving cheating to affect less than 1% of the remaining data. Together, detection and deterrence can almost completely overcome the cheating problem.

Bias Measurement Error due to Respondent Role-playing: A Serious New Challenge for Survey Research

Jonathan Kelley, *International Survey Center*
MDR Evans, *University of Nevada*
Sarah Kelley, *International Survey Center*

Role-playing has become a dangerous new source of measurement error in surveys. But luckily, at least in on-line surveys respondents are willing to confess. Asked at the end, most "Almost always gave my OWN TRUE VIEWS", but recently a frightening proportion role-played, "pretending to be someone else, playing a role instead of giving their own views". Role-playing is getting worse, climbing from <10% in mTurk samples in 2016 and 2017, to 20% in 2019 (in a highly

regarded commercial on-line panel), and ≈30% since 2020 (many mTurk samples). Role-playing is widely diffused through the population: Using reliable multiple-item scales with N>10,000, role-players are not distinct in demography, SES, or party but are distinctive otherwise: Correlations show them less scientifically knowledgeable (-.29), more risk tolerant (.32), more confident in government (.24), less trusting of medical science (-.25), more envious of the rich (.22), and more opposed to income inequality (-.41), social distancing (-.39), and mask wearing (-.28). Role-players are perhaps more ideological and less party political than conventional respondents but the biases are not systematically upward or downward. In multivariate analyses, standardized regression coefficients for role players vs "normals" are sometimes downward-biased (-.04 for role-players vs. expected -.30 for Democratic partisanship's effect on meritocratic attitudes; .07 vs for .35 for subjective health's effect on wellbeing; .11 vs. .39 for education's effect on occupation). But role-playing bias is sometimes upwards (e.g. a stereotypical correlation between father's and respondent's party of .67 vs .45). Overall in half a dozen multivariate analyses with different dependent variables, role-playing has statistically significant interactions with about 40% of two dozen "usual suspect" independent variables. To make matters worse, role-playing is not a fixed trait: Many respondents shift from one panel wave to the next. Dropping all role-players appears to be the only way of avoiding these errors.

Authentic Panels, Strong Projects: The Cleanliness Challenge of Public Opinion Research
Pam Cusick, *Rare Patient Voice*

Are You Willing to Share Your Life? Data Linkage & Consent
Data Analysis After Record Linkage: Sources of Error, Consequences, and Possible Solutions
Martin Slawski, *George Mason University*
Brady Thomas West, *University of Michigan*
Emanuel Ben-David, *United States Census Bureau*

Record linkage bears a lot of opportunities for creating richer data products, saving costs in data collection, reducing respondent burden, and avoiding response bias or measurement error. At the same time, the possibility of linkage error is often unaccounted for. Linkage error arises from uncertainty about which pairs of records residing in two separate files belong to the same statistical unit, and can result into mismatches (false positive matches) and missed matches (false negative matches). For example, in a recent project involving the Health and Retirement Study (HRS), survey respondents were linked to the business register (BR) of the U.S. Census Bureau, to identify the size of the establishment at which they are currently working. In the absence of the employer identification number (EIN), which was the case for about 40% of all records from the HRS, linkage to the BR was performed based on establishment addresses. This approach frequently yields thousands of possible matches in the BR per record in the HRS. In this study, we illustrate how linkage error can contaminate a variety of possible downstream data analysis tasks (e.g., regression or principal component analysis), which in turn leads to invalid inferences. We then provide an overview of possible statistical methods and software that can be applied to mitigate the impact of such errors. We show that there is no universal strategy for this task; instead, the mitigation method of choice depends on several factors such as the mismatch rate, goodness of fit of the model used by the data analyst, available knowledge about the linkage process, computational resources, and the analysis question (inferential goal, statistical model) of interest. The findings of this study can be used to guide practitioners in their choice of a suitable method in studies where errors in record linkage are suspected.

Survey Consent and Data Privacy Expectations in the Social Media Age
Jennifer Kelley, *Westat*
Brad Edwards, *Westat*

Survey research data collection and social media platforms share the same end goal: access to an individual's data. To gain access, one must obtain permission. However, the permission request process and the benefits to the individual and the requestor are quite different across these domains. Recent laws and policies have changed the process in the social media environment (e.g., EU GDPR; Apple's iOS Tracking). High-profile data breaches and egregious use of data have spurred them. Privacy advocates urge more transparency to give individuals agency over how their data are used. But the public may have come to expect data breaches and accept them as a fact of modern life, deciding that the value of a company's product or service is worth the risk. Conversely, the survey consent process has changed little over

the years, governed by strict federal laws and IRB requirements. Further, survey researchers who want to use social media data are often held to far more stringent standards than social media companies, even though survey research has not experienced catastrophic breaches like social media. Despite the differences in consent processes and how they are regulated, individuals may see survey research and social media requests for data through the same lens. This myopic view may be moving the needle of the individual's expectation of privacy in survey research closer to social media and may factor into survey response rate declines. Surveys cannot offer a direct ongoing benefit to the individual. The very concept of privacy and its value has changed greatly. What this means for the survey consent process, and more importantly, how this may impact respondent cooperation, is largely unknown. This paper discusses the shifting expectations of data privacy in the social media age and explores how these factors may affect future consent design and survey cooperation.

Asking Panelists for Their Twitter Handles (Again) and Comparing Consent Rates

Athena Chapekis, *Pew Research Center*
Emma Remy, *Pew Research Center*
Gonzalo Rivero, *Pew Research Center*
Aaron W. Smith, *Pew Research Center*

Linking survey data to digital trace data often requires the consent of respondents for the collection of their online activity. This additional step can introduce biases in our analyses if those who agree to share their data have distinctive attitudes or online behaviors. In this presentation, we analyze the decision of survey respondents to share Twitter data with researchers and the effect of two different consent languages fielded in 2019 and 2021 in Pew Research Center's American Trends Panel -- a nationally representative online panel of U.S. adults. We study differences across a wide set of demographic and attitudinal variables in terms of whether respondents reported to be Twitter users, whether they gave us a Twitter handle, and whether they provided a valid Twitter handle. We also characterize changes in behavior with regard to their willingness to share Twitter data for panelists who were present in the two waves. We show that the revised consent language from 2021 was associated with a lower consent rate overall but those who consented did not look significantly different from those who did not. These results suggest that studies linking attitudes to online trace data in social networking sites are representative of the broader population of interest but also offer a cautionary note about contextual effects on the decision to not consent to data linkage

Will Respondents Submit Their Sensitive Data and at What Price? Case Study on Student Loans Research

Ilan Levine, *The Pew Charitable Trusts*
Ama Takyi-Laryea, *The Pew Charitable Trusts*

The Pew Charitable Trust's Student Loan Research study focuses on the experiences of student loan borrowers who have ever had a federal undergraduate loan in default. In 2019, the study's qualitative research found that due to the complicated nature of the student loan system, some interview participants experienced confusion around topics central to the research, including default status, the names of servicer agencies that they have worked with or the types of repayment plans they have been on. In an effort to validate the surveys' self-reported data, respondents who completed a subsequent, nationally representative survey, were invited to submit their National Student Loans Data System (NSLDS) transcript from the Department of Education website for research purposes. Surveys are an excellent tool to defining and quantifying highly technical and confusing pain points in a given groups' experiences in a particular area. Yet gathering technical and sensitive information, such as dates and finances, through self-reported survey data can lead to inaccurate information due to social desirability bias, recall bias, response bias, and sampling bias. One solution to this challenge is through collecting administrative data in tandem with a survey. This proposal offers a case study on data collection when survey respondents themselves submitted highly sensitive administrative data. Key questions going into this data collection were: Will respondents share their sensitive data for research? And if so, at what incentive level? Nearly half (48%) of respondents in this case study consented to sharing such information and over a quarter (27%) ultimately followed through with their submission. Respondents were willing to submit administrative data at varying incentive levels, ranging from $175 to $0. This presentation will highlight the specific benefits, cost, and logistical challenges, as well as offer considerations for others interested in soliciting self-reported and sensitive administrative data simultaneously from survey respondents.

What's Your Age Again? Testing the Impact of Differing Introductions on Willingness to Provide Full Birthdate Information in the Consumer Expenditure Interview Survey

Graham Jones, *U.S. Bureau of Labor Statistics*

From April 2020 through March 2021, the Bureau of Labor Statistics' Consumer Expenditure Interview (CE) Survey fielded a fourth interview research question asking that respondents provide full date of birth information for their consumer unit (CU) members, aged 14 or older. The purpose of this research question was to test respondent willingness to provide birthdate information, with the ultimate goal of reintroducing the birthdate question to facilitate administrative data linkages, which have the potential to improve CE data quality, while reducing respondent burden and overall survey costs. Also embedded within the research question was an experiment to test the impact of question framing on compliance with the birthdate question. For the first two quarters of the fielding, respondents received an introduction explaining the purpose of the question (i.e., to improve linking to administrative data records) , while respondents in the last two quarters were simply told that CE would like to collect some more demographic information at the end of the interview. The inclusion of the birthdate research question also allowed CE to evaluate the impact of the question on respondent burden and explore measurement error by comparing collected ages asked in the first interview against ages calculated using member date of birth information. Preliminary results indicate that 80.5 percent of eligible CU members provided full birthdate information, and there was no difference between the rates at which CUs in the two introduction wording groups provided birthdate information for their members. This presentation will be of interest to household survey methods researchers, survey statisticians and federal survey data collectors

Monitoring & Evaluating the Quality of the 2020 Census

Monitoring and Evaluating the Quality of the 2020 Census

Michael Bentley, *US Census Bureau*

The 2020 Census, like each prior decennial census, was designed to count everyone in the United States once, only once, and in the right place. With over 330 million people to be counted, it was crucial for the U.S. Census Bureau to closely monitor all data collection progress to manage the program, track potential data quality issues, and identify possible areas of concern. This panel will present a collection of innovative research and analyses used during census operations to successfully monitor progress and ensure optimal data quality.

Real-Time Analysis of Data for the 2020 Census Self-Response Operation

Sarah Konya, *U.S. Census Bureau*

During the 2020 Census, the U.S. Census Bureau monitored data collection through a project called the Real-Time Analysis of Data (RTAD). RTAD allowed the bureau to manage the program, track potential data quality issues, and identify possible areas of concern. This presentation will focus on the metrics that were produced to monitor the Self-response operation. The metrics produced included response rates by mode, post-office delivery status, and demographic distributions. This monitoring project became even more valuable when COVID-19 disrupted census operations and activities and RTAD was able to quickly produce meaningful statistics to help leadership understand the pandemic impact.

Producing 2020 Census Operational Quality Metrics

Sarah Konya, *U.S. Census Bureau*

After data collection ended for the 2020 Census, the U.S. Census Bureau released three sets of operational quality metrics. These products were shared with the public much earlier in the decennial schedule than they have been in the past, demonstrating our commitment to be transparent and timely with the results of the 2020 Census. The releases summarized several aspects of the 2020 Census including final address resolutions, average household size by operation, sub-state summary statistics, and item nonresponse rates by operation. This presentation will highlight some of the results and talk about comparisons to the corresponding 2010 Census metrics.

Developing and Evaluating Self-Response Rate Projections for the 2020 Census

Jordan Misra, *U.S. Census Bureau*

Before the 2020 Census, the U.S. Census Bureau developed tract-level daily self-response rate projections to monitor the expected response during data collection. The self-response projections were developed using three components: 1) a set of linear regression models that predicted self-response rates for each tract based on demographic, socio-economic, and household-level variables from the American Community Survey (ACS) data; 2) expert opinion based on the 2010 Census, middecade tests, and general survey response trends; and 3) a daily response pattern that was modeled after the 2010 Census and middecade tests. This presentation will discuss the methodology used to develop the self-response rate projections, as well as the methodology used to evaluate it. The evaluation of the self-response rate projections focuses on comparing the predicted and observed rates in various ways, including: examining different response modes individually, as well as comparing predictions at different levels of geography and in the context of key operational and demographic variables.

Analysis of Internet Response Rates and Paradata During the 2020 Census
Jordan Misra, *U.S. Census Bureau*

The 2020 Census was the first to utilize online data collection on a large scale, making internet self-response the primary method of data collection. One advantage of an internet census questionnaire is the collection of online paradata, which allows managers to monitor user's interactions with the instrument in real time. This presentation describes how U.S. Census Bureau managers are using internet paradata to monitor the 2020 Census data collection. The talk will focus on the hurdles associated with monitoring a data collection as large as the decennial census. We will discuss reporting the publicly available 2020 Census Self-Response Rates and using paradata metrics, such as tracking device type, average time to complete, and respondents with multiple logins. We will also discuss how processing limitations affected which metrics we chose to monitor and how we programmed these metrics. We will describe how researchers identified real-time issues without having national or regional benchmarks with which to compare. We will end our presentation with a discussion of "lessons learned" that others should consider when using Internet paradata to monitor large surveys and censuses for the first time.

Methods to Track Data Quality in the 2020 Census
Michael Bentley, *US Census Bureau*

During the 2020 Census, the U.S. Census Bureau was committed to collecting and producing high-quality data products for the public. In addition to building steps into the data collection processes that ensured quality, the U.S. Census produced several metrics during and after data collection to quantify the quality of the results. This presentation will give an overview of the methods that the U.S. Census Bureau employed to monitor data collection and measure quality: projecting and reporting self-response rates, the Real-Time Analysis of Data project, and operational data quality metrics. Additionally, this presentation will introduce self-response concepts that will be discussed in subsequent presentations.

Innovations in Election Polling
Innovations in Election Polling
David Sterrett, *NORC at the University Of Chicago*

Election surveys have long helped researchers, policymakers, the media, and public better understand the attitudes of voters. These polls play a key role in American democracy as they help tell the story of who votes and why they vote for a certain candidate. However, election polls have faced challenges in providing reliable information during the last two Presidential elections. Surveys conducted by academics, media organizations, and partisan pollsters have all struggled to accurately represent the electorate in recent years. Given these issues, this panel will explore new approaches in election surveys designed to improve their accuracy and mitigate nonresponse bias, which occurs when those who do not respond to surveys have significantly different attitudes from survey respondents. The presentations will share the results of new methods for recruitment and weighting used in surveys conducted by both researchers and media organizations. Panelists will share insights from AP VoteCast, a modern approach to election polling developed by The

Associated Press and NORC at the University of Chicago, and two of the most widely used academic elections surveys, the American National Election Studies (ANES) and the Cooperative Election Study (CES). The panel aims to help contribute news ideas and approaches to the broader ongoing efforts of researchers and pollsters to improve the reliability of election surveys.

Exploring AP VoteCast's Experiments with Mail Recruitment Approaches for an Election Survey

Mariana Meza Hernandez, *NORC at the University Of Chicago*
Leah Christian, *NORC*
David Sterrett, *NORC*
Jennifer Benz, *NORC*
Benjamin Skalland, *NORC*

AP VoteCast is a survey of the American electorate conducted by NORC at the University of Chicago for The Associated Press and Fox News. Developed by the Associated Press and NORC, VoteCast features a large, nationally representative survey of voters and non-voters that allows for in-depth analysis of election attitudes and behaviors. The survey combines a probability-based sample of registered voters and non-probability samples from opt-in online panels. Respondents sampled from the state voter files are mailed a postcard inviting them to complete the survey either online or by calling an 800 number to take the survey with an interviewer. VoteCast featured 42,000 probability interviews from the voter files and 87,000 non-probability interviews in its 2020 Presidential election survey. In an effort to improve both response rates and representativeness of the probability sample, VoteCast conducted a series of mail recruitment experiments in 2021. The experiments featured more than 200,000 mailings across multiple states, and they tested both different messaging as well as different types of mailings. NORC randomly assigned registered voters into the different conditions. The results provide insights into the effects of the mailings across various demographic and political characteristics. The study's findings show how both, the type of mailing and the messaging of the invitation, can impact the response rates and sample representativeness of an election survey.

Examining Sources of Bias in the 2020 Cooperative Election Study

Brian Schaffner, *Tufts University*

This presentation examines several potential sources of error that may have produced a modest under-estimate of the Trump vote in the 2020 pre-election release of the Cooperative Election Study. Using voter file validation and analysis of post-election re-contact interviews, the presentation I consider whether likely voter modeling, geographical based patterns of nonresponse, or vote preference misreports were significant sources of bias. I also examine the impact of weighting to past presidential vote in reducing bias.

Non-Response Bias: An Analysis of 2020 Data and Experiments

Daron Shaw, *University of Texas - Austin*

In 2016 and 2020, both probability and non-probability surveys systematically underrepresented support for Republican candidates. A wide array of explanations have been offered for this occurrence, with most focusing on relative response rates among individuals who project as right-of-center politically. This research examines several explanations for unit non-response bias, including educational attainment, income, population density, geographic representation of Trump-winning areas, turnout intention compared to actual turnout by state, volunteering, mode of vote, and timing of vote. I rely on data from the 2020 American National Election Study (pre- and post-election surveys), as well as from experiments conducted within 2020 Fox News national and statewide surveys. Based on the ANES and Fox News Poll data analyses, I find that both compositional and distributional errors were factors in Democratic bias. In particular, compositional error explains roughly half of the pro-Democratic error in 2020. Distributional errors are harder to assess, but the low defection rates among Democrats suggest we may be overlooking errors on that side of the partisan equation. In addition, the data indicate that greater attention should be paid to demographic and attitudinal benchmarks used for probability and (especially) non-probability calibration.

AP VoteCast's Application of Advanced Weighting Techniques to Improve the Reliability of Election Surveys

Benjamin Skalland, *NORC at the University Of Chicago*

Benjamin Skalland, *NORC at the University of Chicago*
Susan Paddock, *NORC at the University of Chicago*
David Sterrett, *NORC at the University of Chicago*
Jennifer Benz, *NORC at the University of Chicago*

AP VoteCast is a survey of the American electorate conducted by NORC at the University of Chicago for The Associated Press and Fox News. Developed by the Associated Press and NORC, VoteCast uses a probability-based sample of registered voters to help calibrate a very large sample from opt-in online panels, combining the accuracy of probability-based surveys with the scale afforded by opt-in panels. VoteCast features a four-step weighting process to mitigate demographic imbalances, calibrate the non-probability and probability samples, and improve the regional estimates within states. Since its debut in the 2018 midterm elections, VoteCast's weighing approach has continued to evolve and featured several advances in the January 2021 Georgia Special Senate Elections and November 2021 Virginia Gubernatorial Election. For example, VoteCast's new calibration methodology helps mitigate nonresponse bias, which occurs when those who do not respond to the survey have different attitudes or preferences from responders. The calibration models combine demographic and past vote information from respondents to derive weighting benchmarks for variables in which there could be potential nonresponse bias. This presentation will highlight these new weighting approaches and how they can improve the accuracy of election surveys.

Accuracy and Honesty in Google Searches for Election Prediction and Message-Testing
Skylar White, *unumAI*

(PAPOR STUDENT AWARD WINNER) Over the past several election cycles, underfunded state-level polls, commonly misreported in the media and suffering insufficiencies in sampling representativeness, have fallen short in accurately capturing Republican support. Google Searches, reflecting millions of voter research behaviors over time, have great potential to supplement survey accuracy. People's social desirability and self-desirability biases are reduced when they interact in the privacy of their computers, so they're 'observed' behavior can be better indicators of what they do than what they say. Google Searches also have the advantage of greater honesty on cultural, racial, conspiratorial, and other issue categories that can suffer from social desirability biases.   This study hopes to show the potential for blending and combining survey responses with Google Search data to harness the advantages of both; long-form message-testing and demographic cross-tabulation from surveys, and better sample sizes, honesty, and inclusivity from this passive data source. This approach helps to bolster outreach to lots of hard-to-reach communities that otherwise wouldn't interact with surveys. That's how Google Searches helped to predict 2020 U.S Senate and U.S. House races twice as accurately as public polls one-month before election. When these results are broken down geographically, Google Searches helped to predict suburban, U.S. House race winners almost 5-times as often as public polls from around the same time.

On a Scale of 1 to X: Scale Design, Development, Direction, & Labeling

In Which Direction Do You Think This Scale Should Go? Scale Direction Preference and Its Impact on Survey Responses
Ting Yan, *Westat*
Florian Keusch, *University of Mannheim*

Survey literature has documented a tendency for respondents to select scale options closer to the start of a scale. Therefore, the direction in which a scale is presented has an important impact on the resultant answers. However, there exist neither agreed-upon best practices nor industry standards when it comes to scale direction. In addition, little is known whether or not survey respondents think there is a preferred order to a response scale and if yes, what their preferred direction is. It is also not clear how respondents' preference affects their answers, especially when they are presented a scale that runs in the opposite direction to what they prefer. To answer these research questions, we conducted two studies that asked respondents directly in which direction they think a response scale should progress. This paper will present the results from these two studies. We will first discuss respondents' preferred direction for two scales – an agreement scale and a frequency scale. We then will examine the impact of the preferred scale direction on

their answers as well as the interaction between the preferred scale direction and the actual scale direction assigned to respondents. Practical implications of the results on questionnaire design and evaluation will be discussed.

Getting Crafty with Scale Labels: A DIY Rating Scale
Valerie Lykes, *J.D. Power*
Jay Meyers, *J.D. Power*

J.D. Power benchmarks customer satisfaction across several industries and over the past ten years, the discrimination of our 10-point scale has declined due to ceiling effects. We hypothesized that discrimination and data normality would improve if we asked respondents to "create their own scale." We compared this approach to using a pre-set 6-point fully labeled scale that has tested well for us in the last year (N of 1,000 per scale type) and found that the DIY scale produced better scale discrimination and a more symmetric distribution.   We decided to replicate this first test and expand it in order to understand whether we saw a short-term priming effect or whether the improvement would be retained over time.   A sample of 3,000 panelists were asked to create their own scale using a "long" or "short" scale creation question battery and to rate two industries.   Two weeks later they were asked to rate two new industries using the scale they created in the first survey.   Results suggest this approach may outperform the 6-point fully labeled scale and that at least over 2 weeks the improvement is generally maintained.

The Good Life: How Improving Quality of Life Measures Can Improve Our Insights in Subgroups
Frances Barlas, *Ipsos*
Randall K. Thomas, *Ipsos*
Ge Tang, *Ipsos*

Measurement of quality of life has been viewed as a critical summary judgment that reflects, at a glance, the evaluative experience of an individual, and is often used in social surveys (from the General Social Survey to European Social Survey). There have been a number of measures that measure quality of life, from overall judgments to evaluations of specific aspects of life. In addition, many have adopted measures that reflect positive emotional evaluations (e.g., the Flourishing Scale by Diener). One difficulty we have identified is that many measures are very bulky, having long items, numerous items, and lengthy response formats. Such properties do not work well in an online survey format as they create problems for self-administration of online surveys on smartphones – the devices of preference for those who are younger and respondents of color. We conducted 2 studies with over 3,000 participants in each, using the KnowledgePanel, a probability-based online panel. Respondents were randomly assigned to either complete the traditional quality of life measures, such as the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) (Endicott et al., 1993) or complete alternative, more mobile-friendly versions (shorter items, fewer items, and shorter response formats). We evaluated the impact on people taking the surveys on different devices and from different subgroups. Generally, we found that the mobile-friendly scales performed equally well in terms of group differentiation, predictors, and correlations with other similar measures. They also took less time to complete and were completed more easily on smartphones. We recommend that these shorter measures be considered for future use to enable better longitudinal and cross-cultural comparisons.

Evaluating the Effectiveness of Visual Analog Scales to Measure Expectations
Tuba Suzer-Gurtekin, *University of Michigan*

The response scales for the measurement of inflation and income expectations have evolved over time and the most recent scale that combines verbal likelihood and numeric responses has been proven to produce valid and reliable survey estimates using the telephone survey mode since mid-1970s. In he parallel web surveys, these response scales produce coarser data conditioned on the location of the underlying expectation distribution as a result of the significant increase in the percentage of "stay the same" responses compared to the telephone surveys. While visual analog scales (VAS) seem inherently present a numeric response scale to measure expectations more precisely, the evidence for data comparability with respect to radio-button scale data is mixed in the literature. This study investigates the effectiveness of VAS in measuring income expectations in a series of randomized experiments in Address Based Sampling (ABS) web surveys. While there will not be randomization by device used, the two versions of response scale will be randomized: 1) combination of verbal likelihood and numeric response, 2) visual analog scale that includes

placing a mark for the response on a numeric scale. The medians, 25 th and 75 th percentiles of income expectations will be compared between the two experimental conditions in addition to estimates from the parallel telephone surveys.

It's All Semantics, or Numerics, or Emojis?: Comparisons of Utility and Function
Randall K. Thomas, *Ipsos Public Affairs*
Megan A. Hendrich, *Ipsos Public Affairs*
Frances M. Barlas, *Ipsos Public Affairs*

Fully-anchored, semantically-labeled scales have somewhat higher levels of validity than end-anchored scales (Krosnick, 1999). However, since most smartphone respondents take online surveys in the portrait orientation, even the horizontal presentation of five categories with full semantic labels can extend off the screen. As an alternative to semantic formats, we developed two alternative formats for smartphone screens: 1) numeric labels as clickable buttons to anchor the responses, with plus or minus indicators and no end semantic labels (e.g. How much do you like doing X? –2 –1 0 +1 +2) and 2) emojis using thumbs up and down to represent direction, with sizes of thumbs and order representing degree.   In our previous research, we found that both numerics and emojis functioned well, having equivalent validity as semantic scales. In the study we report here, we had over 10,000 participants from an opt-in non-probability sample complete the survey. We asked about agreement with 10 different political attitudes, such as reducing the wealth gap or using capital punishment. We randomly assigned to one of the three response label formats for agreement scales (semantic, numeric, emojis). We also randomly assigned respondents to a unipolar or bipolar format, and randomly assigned to 3, 4, or 5 response categories. We found that label type did not lead to any significant differences in mean values. In addition, we found that unipolar formats had lower means than bipolar formats, replicating earlier findings.   We further found that all three label types had equivalent single-variable validity. Then we compared multiple regression models, entering all 10 measures as predictors and political party identification as the dependent variable. We generally found comparable models, although some coefficients were divergent.   We discuss the future applications and some limitations we've encountered in our exploration of both the numeric and emojis formats

Who's Asking? Advances in Research on Interviewer Behavior
Identification of Partial Interviewer Falsification in Panel Surveys
Silvia Schwanhäuser, *Institute for Employment Research*
Jonas Beste, *Institute for Employment Research, Germany*
Lukas Olbrich, *Institute for Employment Research, Germany*
Joseph Sakshaug, *University of Mannheim, University of Munich (LMU), and Institute for Employment Research, Germany*

Interviewer-administered surveys are, in many respects, seen as the gold-standard form of data collection. Interviewers play a vital role in achieving high data quality by contacting, identifying, and recruiting target respondents, answering their queries, and administering standardized interviews. However, some interviewers may be enticed to intentionally deviate from the prescribed interviewing guidelines and fabricate (parts of or entire) interviews. Such fabrication can lead to severe bias, especially in multivariate analyses. Hence, several studies have proposed methods for preventing or identifying complete falsifications. At the same time, current literature mostly neglects two important aspects: (1) How can survey researchers detect partial falsifications? and (2) How can researchers effectively detect different falsification forms in panel survey data? The common notion in panel surveys is that falsifications are easy to detect, since inconsistent or implausible answers between waves could be flagged as suspicious. Nonetheless, information on the concrete implementation of "between waves" checks, as well as evaluations on the effectiveness of such checks for detecting falsifications are missing from the literature. Further, previous literature lacks methods targeted on partial falsifications for both longitudinal and cross-sectional data.   In the present case study, we aim to close these gaps by examining whether we can effectively identify partial falsifications in the German Panel Study "Labour Market and Social Security" (PASS), which included verified cases of interviewer misbehaviour and partial falsifications. First, we assess whether established statistical detection methods and falsification indicators also succeed in identifying partial falsifications. Second, we test the common notion that falsifiers provide inconsistent answers to otherwise time-stable items between different waves of data collection. Altogether, the results of this study inform how survey researchers can improve their quality control procedures for panel surveys.

Do Humor and Relatability Predict Interviewer Performance?

Matt Jans, *ICF*
Don Allen, *iCF*
Zoe Padgett, *ICF*
Dave Roe, *ICF*

Facets of an interviewers' personality and their interpersonal skills can increase their ability to gain cooperation and avoid refusals or hang-ups. Humor and relatability are two important dimensions, that have received relatively little attention in survey methodology, particularly in random digit dial surveys. We wanted to know whether subjective assessments of interviewers' relatability and sense of humor (i.e., how funny they are perceived to be) are correlated with important performance metrics, such as contact, refusals, refusal conversions, and cooperation. Interviewer quality assurance (QA) monitors were asked to nominate interviewers who they thought were the most relatable and who had the best senses of humor using the following questions: Question 1: "Among all the current interviewers you've observed or interacted with, including formal and informal interactions, and QA monitoring sessions, who do you think has the best sense of humor ? Who makes you laugh? You can list as many interviewers as you can think of." Question 2: "Thinking about the same interviewer interactions, who do you think is the most relatable ? Please list interviewers again here if they are also in your answer above (i.e., if you find them funny and relatable)." Nominations were used to create four groups of interviewers: 1) "nominated funny only", 2) "nominated relatable only", 3) "nominated both funny and relatable", and 4) "not nominated". We also captured the number of nominations each interviewer received as a potential metric of the strength of these dimensions (i.e., are they widely known among staff as funny or relatable?). Contact, refusal, refusal conversion, and cooperation rates were compared between each of the four groups. The findings are discussed in the contacts of nonresponse avoidance, and interviewer training and management.

Interviewing the Interviewer: Phase 1 Results from the Interviewer Quality of Life Study

Mariel Leonard, *DeZIM*

Survey interviewers, both those involved in qualitative data collection in the pre-fielding stage, as well as those involved in quantitative data collection in the field, are an essential part of survey research. Numerous studies have investigated how interviewers may influence the quality of the data they collect and thus the overall conclusions of a survey project. Ethics committees and institutional review boards additionally spend tremendous time discussing the effects of survey research participation on respondents. However, very little research has considered how interviewers may be affected by the survey projects they work on. In this paper, I will present results from the first phase of the Interviewer Quality of Life Study (IQLS). The IQLS explores the day-to-day work of survey interviewers through all stages of their interactions with respondents (recruitment, interviewing, refusal conversion, etc.), and assess the effects of interviewer work on the interviewers themselves. In Phase 1, I conducted 20 in-depth interviews with survey interviewers from diverse personal and professional backgrounds. These interviews were analyzed using grounded theory to identify emergent themes of interviewer motivation, causes of stress and burnout, and needs for institutional support.

The Effect of Audio Recordings on Deviant Interviewer Behavior

Lukas Olbrich, *Institute for Employment Research*
Jonas Beste, *Institute for Employment Research*
Silvia Schwanhäuser, *Institute for Employment Research*
Joseph W. Sakshaug, *Institute for Employment Research, LMU Munich, University of Mannheim*

Survey organizations increasingly rely on audio recordings of face-to-face interviews to deter interviewers from deviant behavior and to monitor their work. As the audio recording usually requires respondents' consent at the beginning of the interview, interviewers are aware if the current interview is monitored. In this study, we investigate if the status of consent impacts interviewers' behavior. We use data from a large-scale mixed mode survey in Germany where both telephone and face-to-face interviewers ask respondents for recording consent. In this setting, we exploit that recording should not affect telephone interviewers as supervisors can monitor the interviews live in the telephone studio without recording, while in face-to-face interviews such live monitoring is not possible. Using paradata on the duration of questionnaire sections and quantile difference-in-differences regressions, we find that respondent consent to audio

recordings in face-to-face interviews reduces the prevalence of extremely short and long durations that indicate deviant behavior. We implement several robustness checks to ensure that mode effects or respondent self-selection do not explain the audio recording effect. In summary, the results show that face-to-face interviewers change their behavior when they know the interview is being recorded. Thus, audio recordings deter interviewers from deviant behavior. In addition, the results show that interviewer monitoring strategies mainly based on audio recordings, conditional on respondent consent, will deliver a biased picture of the interviewers' work as non-recorded interviews with more frequent deviations are not evaluated.

What Drives Happiness? The Interviewer's Happiness

Adam Stefkovics, *Harvard University Faculty of Arts & Sciences*
Endre Sik, *Centre for Social Sciences, Hungarian Academy of Sciences Centre of Excellence*

Interviewers in face-to-face surveys can potentially introduce bias both in the recruiting and the measurement phase. One reason behind this is that the measurement of subjective well-being has been found to be associated with social desirability bias. Respondents tend to tailor their responses in the presence of others, for instance by presenting a more positive image of themselves instead of reporting their true attitude. In this study, we investigated the role of interviewers in the measurement of happiness. We were particularly interested in whether the interviewer's happiness correlates with the respondent's happiness. Our data comes from a face-to-face survey conducted in Hungary, which included the attitudes of both respondents and interviewers. The results of the multilevel regression models showed that interviewers account for a significant amount of variance in responses obtained from respondents, even after controlling for a range of characteristics of both respondents, interviewers, and settlements. We also found that respondents were 3.879 times more likely to report a happy personality in the presence of an interviewer with a happy personality. We argue that as long as interviewers are involved in the collection of SWB measures, further training of interviewers on raising awareness on personality traits, self-expression, neutrality, and unjustified positive confirmations is essential.

Novel Sample Design & Weighting Approaches
Developing the Sample Design for the New Annual Integrated Economic Survey

Katherine Thompson, *U.S. Census Bureau*

Over the next several years, the economic directorate of the U.S. Census Bureau will introduce the Annual Integrated Economic Survey (AIES), an economy wide survey that replaces a suite of six independently designed ongoing surveys. The AIES sample design requirements are informed by the user community's longstanding data needs (e.g., national and sub-national tabulations), as well as by extensive respondent research on collection. This paper provides an end-to-end high-level overview of the AIES probability sampling design, including determination of the sampling unit, computation of unit-level inclusion probabilities for the implement probability proportional to size sampling, stratification, allocation, and sample selection and validation. Throughout, we selectively highlight specific challenges of developing a multi-purpose business survey whose collection covers a wide range of economic sectors. The paper concludes with a brief discussion on post-sampling activities designed to improve the quality of the tabulated domain estimates, including calibration weighting and proposed imputation methodology.

An Assembly of Survey Sample Size Planning Parameters - Observed Values of Vaccination Coverage Intracluster Correlation Coefficients from Many Recent Studies

Dale Rhoda, *Biostat Global Consulting*
Becca Robinson, *Biostat Global Consulting*
Caitlin B. Clary, *Biostat Global Consulting*
Mary Kay Trimner, *Biostat Global Consulting*
Shiraz Johnson, *Biostat Global Consulting*
Jennifer Brustrom, *Biostat Global Consulting*

When calculating the sample size for a vaccination coverage cluster survey in low- and middle-income countries, planners begin with numeric targets for the outcome proportion and its precision, and they calculate an effective

sample size that would succeed with a simple random sample.   Then they account for complex sampling using a guesstimated design effect that depends on the unknown spatial heterogeneity of the outcome as parameterized with an intracluster correlation coefficient (ICC).   In this talk we graphically summarize hundreds of observed values of ICC (and other tantalizing planning parameters) from surveys conducted in recent years and stratify the results by survey type and sampling design. We also showcase the generic user-written Stata and R programs that we make available so collaborators around the world may extract parameters from their datasets using consistent calculations.   Summaries from our (still growing) collective dataset will help planners balance their tolerance for risk of fielding too small a survey to meet the stakeholders' precision goals versus one that is larger and more costly (and perhaps of lower quality) than necessary.   Of note: the programs that extract parameters from survey datasets include extensions to Stata's and R's survey estimation commands that could be generically useful for cluster surveys related to other topics.

Advanced Cellular RDD – The Next Generation

Mansour Fahimi, *Marketing Systems Group*
David Malarek, *Marketing Systems Group*

For decades, random digit dialing (RDD) sampling methodology has served as the reliable workhorse for survey and market research applications, surviving the many trials and tribulations brought about by the digital evolution. Setting aside the diminishing rates of response, which have been plaguing all survey modalities rather indiscriminately, proliferation of mobile phones has created several unique challenges for the RDD methodology. Unless addressed effectively, this capable data collection modality is bound to sunset even sooner than what many pundits have surmised. As reliance on cellular numbers for RDD surveys continue to increase while landlines fade away, the mobile nature of cellular numbers has been creating a formidable challenge for survey sampling because such numbers may not be associated with intended geographies. Although a rather moot issue for national surveys, however, subnational surveys that rely on cellular RDD are subject to in- and out-migrations that can introduce unknowable coverage biases. This means for any study geography, there are cellular subscribers with area codes serving the given geography who actually reside outside of that area (out-migration) as well as those residing within the given geography with nonnative area codes (in-migration). Traditional sampling of cellular numbers is subject to both omission of in-migrating and increased cost for screening out out-migrating subscribers. This becomes of particular concern as the study areas reduce from states to smaller geographic parcels. By taking advantage of massive consumer databases, it is now possible to resolve the majority of in- and out-migrated cellular numbers. From this perspective, the authors will introduce a pioneering process for enhancing cellular RDD frames that can effectively accommodate telephone surveys that focus on subnational geographies. This presentation will include unseen coverage reports that are critical in construction of the next generation of cellular RDD sampling frames for surveys that demand uncompromised coverage properties.

Understanding the Impact of Enhanced Weighting Tools on Key Health Survey Indicators in the California Health Interview Survey

Jiangzhou Fu, *UCLA Center for Health Policy Research*
Todd Hughes, *UCLA Center for Health Policy Research*
Royce Park, *UCLA Center for Health Policy Research*
YuChing Yang, *UCLA Center for Health Policy Research*
Ninez Ponce, *UCLA Center for Health Policy Research*
Arina Goyle, *SSRS*

The California Health Interview Survey (CHIS) has implemented several methodological innovations in recent years. Since 2019, CHIS has employed address-based sampling (ABS) frame with a mail push-to-web survey followed by telephone nonresponse follow-up as the respondent recruitment approach, which has been proven efficient to increase response rate and reduce survey cost (Todd Hughes et. al., 2020). Apart from advancements in survey sampling and data collection strategies, CHIS 2020 post data collection adjustment approaches have also been improved, mainly regarding California population projection estimates and weighting calibration variables. To be aligned with CHIS 2020 revised weighting tools, CHIS 2019 estimates were also re-weighted utilizing the updated weighting tools. This paper will first discuss the motivations of enhanced weighting tools and impacts on CHIS 2019-2020 sample compositions, especially focusing on reducing the underrepresentation of Asian subgroups. Additionally, we will explore the

discrepancies between original and re-weighted CHIS 2019 estimates. CHIS 2020 estimates of crucial indicators monitoring health behavior, health care, health insurance will be presented and examined, and differences by key subgroups (age group, race/ethnicity, citizenship, and poverty level) will be analyzed. This paper will also scrutinize multiple indicators with salient variation in 2020 compared with previous years, potentially due to COVID-19 pandemic. Keywords: health surveys, survey re-weighting, minority representation

Don't Use the UWE to Evaluate a Poststratified or Raked Estimator (Use Something Else Instead)
Phillip Kott, *RTI International*

The unequal weighting effect (UWE) purports to measure the increase in variance of an estimated mean due to using unequal weights. It does that when the observations in the sample are independent and identically distributed, which is rarely the case in practice. When weights are unequal due to poststratification or raking, the UWE often does not provide a very good proxy for the relative "increase" in variance due to weighting. The variance may, in fact, decrease. This should not be surprising because these methods, although popular for bias reduction in the face of unit nonresponse or coverage error, have been shown to be effective for variance reduction when variance measures are properly conditioned. The presentation will demonstrate the potential variance-reducing impact of poststratification and raking in some simple examples and describe direct measures of variance that should be used instead of the UWE.

Quality, Calidad, Qualité: No Matter How You Say It, We Want to Improve It

3MC in a Full Quality Perspective
Peter Mohler, *Mannheim University*

Those where the days when we published the results of a „representative" survey with "3% error rate" using probability based statistical analyses of sorts. In terms of Deming's fundamental definition of quality, such a survey was „fit for the intended use", namely making quantitative inferences from a sample survey to the targeted population. That held also for 3MC surveys such as the Eurobarometers or the ISSP. Deming's fitness for use can be said to be today's best available general quality conception for all industrial purposes. Survey production today is undoubtedly governed by industrial production processes. Thus Deming's overarching concept can and should be used as top quality definition especially in the complex production environment of 3MC surveys. However, the AAPOR/WAPOR Force Report on Quality in Comparative Surveys (2021) did not adopt this view. Following the now standard literature on 3MC it stated several, concurring quality concepts such as TSE or Deming's. However, comprehensive fitness for use perspective would allow users to know, how far they can stretch their interpretation, how complex the statistics must be for proper comparisons. This is perspective streamlines that taken up to now in 3MC literature and modernises it The perspective, named here "Full Quality Perspective" also subsumes the many thoughts, guidelines, hints etc. embedded in the now standard literature on 3MC. Neither TSE, translation, sampling, harmonizing, demographics etc. further on can dominate a survey production by maximizing their idiosyncratic directives. In the end it is the paper to be published that defines the quality requirements including ease of use, transparency, replicability and so on. The paper will show how this could be accomplished by breaking down the survey life cycle model (p. 29) into comprehensive survey production processes with well defined quality characteristics for the intended use.

Identifying and Surveying Low Incidence Race and Ethnic Minority Populations by Cell Phone
Brian Robertson, *Market Decisions Research*
David Malarek, *Marketing Systems Group*

One of the challenges in surveying low incidence populations is effectively identifying in cell phone samples households with low incidence populations. This was a challenge we faced in conducting the Oregon Health Insurance Survey. For this project, our goal was a minimum number of completed surveys among African American, Asian, Native American, and Hispanic residents which make up 3.0%, 3.1%, 6.3%, and 13.5% of the Oregon population, respectively. We relied on two different strategies for the 2019 and 2021 administrations. During 2019, sampling used targeted non-probabilistic cell phone samples identified as more likely to include residents that were African American, Asian, Native American, or Hispanic. This sample produced marginally higher incidences of these populations than in the RDD cell phone sample used for the project. We changed our approach in 2021 and used Marketing Systems Group's advanced cellular frame to generate RDD cell phone sample to which demographics were appended when available. This identified records more likely to include African American, Asian, Native American, or Hispanic residents. These records were prioritized during calling with additional call attempts to reach residents and complete a survey. The results in 2021 were dramatically improved over 2019. We were able to exceed our set goals for surveys among these four populations. The incidence among sample records including appended demographics was 19% among African American, 55% among Asian, 19% among Native American, and 56% among Hispanic records flagged for the specific low incidence population. We conclude that the use of the advanced cell frame with appended demographics is an effective method to identify and complete cell phone surveys with low incidence race and ethnic minority populations.

Efforts to Improve Inclusion of the Spanish-Speaking Population in the 2020 Residential Energy Consumption Survey
Kristine Wiant, *RTI International*
Joe Murphy, *RTI International*
Joe McMichael, *RTI International*
Carolyn Hronis, *U.S. Energy Information Administration*

The 2020 Residential Energy Consumption Survey (RECS), conducted by the U.S. Energy Information Administration, sampled over 50,000 sampled households in all 50 US states and the District of Columbia. The survey contacted sampled households by mail, inviting them to participate in either a web or paper survey. Survey invitations, reminders, and the web survey were all available in both English and Spanish. An English-language paper version of the survey was sent to all respondents during Phase 1 of data collection, and respondents were informed that they could call to request a Spanish-language version. Very few such requests were made. Inclusion of Hispanic and racially diverse respondents is important for this representative survey. Yet, Phase 1 results indicated that Spanish-speaking households were responding at lower rates than expected based on American Community Survey (ACS) estimates. To address this, we appended ACS data at the census block group (CBG) level indicating the percentage of occupied households that speak Spanish or Creole and in which no person aged 14 or older speaks English "very well." In Phases 2 and 3, we selected the 1,429 households (3.4% of the sample) in areas with the highest Spanish-speaking concentration to receive both an English- and Spanish-language questionnaire. The households selected to receive the Spanish-language questionnaire were from CBGs in which 17.7% or more of the households speak Spanish (threshold based on a budgeted number of cases). The goal was to proactively provide Spanish-speaking households with a paper survey option that met their requirements. In this presentation, we will discuss the strategy to increase inclusion of Spanish speakers and examine the results of the experiment in terms of data quality, representation, response, and cost. We will conclude with additional thoughts on improvements for inclusion of the Spanish-speaking population in the future for RECS and other similar surveys.

Assessing the Measurement Invariance of the General Health Questionnaire Across Ethnic Groups

Nick Allum, *University of Essex*
Kirby King, *University of Essex*
Paul Stoneman, *Kings College, London*
Alexandru Cernat, *University of Manchester*

Disparities in health outcomes for different racial and ethnic groups are a serious issue for public policy and for individual lives. Robust data are needed in order to assess the nature and extent of such disparities before policies can be developed to mitigate them. In this paper we investigate the extent to which a widely used set of items - the General Health Questionnaire (GHQ) - are valid for comparing mental health outcomes for six ethnic subpopulations in the UK. We use data from the UK Household Longitudinal Study (UKHLS), a long-established probability-based panel survey representative of the the UK population, which oversamples selected minority groups. Using confirmatory factor analysis, we test for configural, metric and scalar invariance of the 12-item version of the GHQ. A sensitivity analysis comparing the more usual summated score and latent means across groups is carried out to assess the validity of the former approach to analysing the GHQ. Finally, revised estimates of scale reliability were derived using two different methods. We show how scale items with differing factor loadings are better assessed by the use of an alternative measure of reliability, McDonald's Omega (ω) rather than the more usual Cronbach's Alpha (α). We find that the GHQ-12 does not display obvious bias in regard to ethnic groups in the UK and that valid comparisons across these groups can be made for the purposes of survey research. We also discuss more broadly the advantages of using ω rather than α to assess scale reliability.

DEI-BR in Multilingual and Multicultural Tests & Surveys - Language Variations and Cultural Considerations

Musab Hayatli, *cApStAn Inc*
Grace DeLee, *cApStAn LQC Inc.*

Awareness of Diversity, Equity and Inclusion in tests, assessment, polls and surveys is growing, becoming increasingly indispensable as it both ensures the fairness of tests and the reliability for any data collection instrument. It is a very positive move! The increased awareness, however, is not without its challenges. And the challenges for multilingual and multinational polls and surveys are naturally exponentially bigger! Not only they need to have a larger and more diverse coverage, but this coverage would need to be in other languages and would need to consider potentially very diverse cultural norms. We have been partnering with organizations who administer surveys and assessments in multiple languages and in multiple cultures. While we are not a DEI-BR compliance organisation per se, we have been advising our partners on the language and cultural dimensions for their multilingual and multinational surveys for years.

Language and language issues are our bread and butter, and we would like to share some of our observations and offer some of the solutions and approaches which we hope would help you ensure better DEI-BR practices in your surveys.

Multicultural Insights Test: Development of a Diagnostic Tool for Identifying Basic Knowledge Needed in the Multicultural Industry
Edward Rincon, *Rincon & Associates LLC*
David Morse, *New American Dimensions*
Mandy Sha, *mandysha.com*
Dexter Purnell, *University of North Texas - Denton*

According to the 2020 Census, the U.S. multicultural population currently numbers 125 million – a presence that has created an increasing demand for a workforce that understands their cultural and linguistic nuances. Our academic institutions, however, have struggled to prepare its graduates with the knowledge needed to meet the demands for talented individuals in the public and private sectors, especially the research industry. Further complicating the availability of a talented workforce is the absence of a measurement tool that can identify individuals with the basic knowledge needed to work effectively with multicultural persons. In the long run, misplaced talent is costly to an organization, leading to employee turnover, job dissatisfaction, poor sales, decreasing market share, and pre-mature termination of multicultural projects. To address this workforce need, a team of multicultural experts will develop and test a measurement tool entitled the Multicultural Insights Test (MIT) with a non-scientific sample of approximately 1,000 industry stakeholders – including students, research practitioners, and other working professionals in the private and public sectors. The MIT will include 100 to 150 multiple choice questions that will measure knowledge in six areas for the Black, Latino and Asian segments: demographic trends and concepts, language behavior, history, lifestyles, beliefs and the relationship of these areas to decision-making skills. The MIT is expected to provide several benefits: (a) Allow individuals to obtain a self-assessment of their multicultural knowledge and define their training needs, (b) provide employers a diagnostic profile of the training needs of their workforce, (c) provide academics a tool to supplement their curriculum, and (d) define the training or information needs of research practitioners that plan to focus on multicultural communities.

Critical Data Needs: Policing, Incarceration, & Gun Violence
Critical Data Needs: Policing, Incarceration, and Gun Violence
Jennifer Benz, *NORC at the University of Chicago*

The United States does not have a systematic and accessible criminal justice data infrastructure which leaves the country's policymakers at the national and local levels without reliable evidence to inform decision making. Findings from several recent surveys reveal the critical insights that public perceptions will contribute to this data infrastructure. This panel will highlight several of these insights and demonstrate the power of pairing surveys with other administrative and policy data to assess outcomes and inform policy. The presentations will include: data from the AP-NORC Center revealing the disparate perceptions of and experiences with the criminal justice system faced by different racial and ethnic groups in the U.S; a demonstration of how local surveys and open data assets can be paired to understand the impacts of policing practices; findings from a large national survey that revealed the widespread and unequal impacts of incarceration of American families; a rationale for how accurate survey data on firearms can improve policy; and, an overview of the data and science needed to evaluate criminal justice reforms.

Attitudes Toward Criminal (In)justice in the United States
Emily Alvarez, *NORC at the University Of Chicago*
Jennifer Benz, *NORC at the University of Chicago*
Marjorie Connelly, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*

The murder of George Floyd at the hands of a white police officer in May 2020 sparked a nationwide dialogue on race and racism that continues today. For many, it was another example in a long history of racism and violence at the hands of police. For others, it was the first exposure to criminal justice injustice. This presentation will synthesize over 5

years of research, that began in the wake of the killing of Michael Brown, conducted by The Associated Press-NORC Center for Public Affairs Research tracking attitudes toward the criminal justice system in the United States. The data illuminate the often wide gaps in perceptions of and experiences with the criminal justice system by different racial and ethnic groups in the United States. For example, data collected at 5 points since 2015 show gaps as wide as 54 percentage points in Black and white perceptions of the seriousness of police violence against the public. A survey conducted in the immediate wake of Floyd's murder found Black Americans to be more than twice as likely as white Americans to say police officers generally respond to protesters with excessive force. A 2021 survey found that 60% of Black Americans believe racism to be an extremely serious problem in the U.S. while just 23% of white Americans agreed. Not all data indicate racial and ethnic gaps in attitudes, however. The same 2021 survey found widespread support for the implementation of several policies aimed at reducing policy violence. For example, 81% of white Americans and 79% of Black Americans support requiring officers to report misconduct by their peers.

Understanding Individual and Neighborhood Level Differences in Experiences With and Perspectives on Policing, a Case Study of Two Counties
Jennifer Benz, *NORC at the University of Chicago*
Nick Hart, *Data Foundation*

In early 2021, NORC conducted a survey on experiences and perceptions of police activity in two large counties - Dallas County, TX and Cook County, IL. Using a population-based sample, we explored variation by race and ethnicity of the respondents with a focus on three groups: white non-Hispanic, Black non-Hispanic, and Hispanic adults. We identified significant differences between groups in both counties on several key outcomes, including satisfaction with police activity, expectations regarding use of force during arrests, and critical perceptions of police. We then combined the survey data with area-level data from the 2019 American Community Survey. Using multilevel regression models, we examined variation in these outcomes after adjusting for individual and neighborhood sociodemographics and explored support for a theoretical model that encompassed multiple aspects of an individual's perspective and experiences with policing in their community. We attempted to incorporate data on arrests, violent crime, and officer-involved fatalities into our regression models but were unable to do so due to limitations with public data. In this presentation, we will share highlights from the results of our study and lessons learned during our efforts to combine open data sources. In line with federal efforts to improve data infrastructure and transparency, we will provide recommendations for local policy makers to encourage data sharing and outline steps needed to create open data sets that are comparable across jurisdictions within counties.

How Public Attitudes Respond to Crime and the Carceral State
Peter Enns, *Cornell University*

This presentation brings several sources of survey data together to understand the causes and consequences of mass incarceration in the United States. To understand the historical relationship between crime and public attitudes, 120 survey questions about concern with crime are combined to generate a time series of public concern for crime in the U.S. from 1965 to 2021. Then 380 survey questions measuring the public's support for tough-on-crime policies are combined to generate a measure of public punitiveness from 1953 to 2020. These series show that concern with crime is only moderately correlated with crime rates, but punitive attitudes closely track crime victimizations and crimes reported to the police. These patterns offer important insights into the effects of crime on public opinion. I then use the Family History of Incarceration Survey (FamHIS) (Enns et al. 2018) to document the prevalence of family incarceration and how personal and family incarceration influence social and political behaviors. FamHIS is a probability-based survey of more than 4,000 adults and found that 45 percent of Americans have had an immediate family member incarcerated in prison or jail for at least a night at some point during that family member's life.

The Role of Data and Evidence in 21st Century Policymaking
John Roman, *NORC at the University Of Chicago*

In 2020, NORC at the University of Chicago convened a panel of 14 experts with diverse backgrounds as researchers, clinicians, policymakers, and practitioners to develop recommendations for an expanded and improved US firearms data infrastructure. The panel issued three reports, including a summary of recommendations in A Blueprint for U.S.

Firearms Data Infrastructure which details changes in policy and practice to improve the public health and criminal justice data necessary to develop more valid and reliable gun violence measures to inform policymaking and improve public safety. Building on these recommendations, in 2021, NORC and Arnold Ventures commissioned seven papers to implement the panel's Blueprint, in Improving Data Infrastructure to Reduce Firearms Violence . The report describes the need for a national nonfatal firearms injury database, the strengths and limitations of existing federal data collections around firearms injury and violence in public health and criminal justice, and, how to support efforts to create better firearms data at the state and local level. Emerging from these papers is the need to collect better survey data to fill gaps in the administrative data that has traditionally formed the foundation for data-driven policymaking. Data linkages between criminal justice data and public health data are extremely limited, and survey data is needed to fill those gaps

How Do We Know What Works? The Role of Data and Evaluation in Criminal Justice Reform
Jennifer Benz, *NORC at the University of Chicago*

This presentation will focus on how data are leveraged for developingcriminal justice reforms and assessing their impacts. The presentation will highlight examples of innovative approaches designed to advance effective reform of policing and public safety strategies that minimize harm and promote the safety and wellbeing of individuals and communities. Example studies include: a project with the DC Metropolitan Police Department to conduct a randomized controlled trial of MPD's body-worn camera program, an evaluation of a training that equips police officers with knowledge of the historical and cultural context in policing, and combining analyses of administrative data with survey experiments to assess civilian attitudinal and behavioral responses to state violence. The session will also focus on the ways that a systematic and accessible criminal justice data infrastructure would facilitate policymaking, reforms, and better understanding of the range of impacts that policing practices have on the communities they serve.

The Current State of Practice in Health Issues Measurement

Measuring Public Attitudes Toward Political and Science-Related Populism in Surveys: A Tale of Two Scales?
Niels Mede, *University of Zurich*
Robert A. Huber, *University of Salzburg, Austria*
Jakob-Moritz Eberl, *Ludwig Maximilian University of Munich, Germany*
Esther Greussing, *TU Braunschweig, Germany*

Opposition of populist citizens against political elites has been intensively investigated, whereas populist criticism of other societal institutions, such as science, has only recently gained scholarly attention. Political and science-related populism are both understood as a set of ideas that revolve around an antagonism between a virtuous common people and an unvirtuous elite, with this antagonism stemming from the elite's illegitimate claims for decision-making authority. But while political populism focuses on political power claims and challenges political elites, science-related populism addresses epistemological truth claims and criticizes academic elites. Conceptually , therefore, both populism variants have overlaps but also differences. Empirically , however, it is unclear how distinct these variants are when manifesting in people's attitudes, i.e., as political and science-related populist attitudes. We address this gap and compare two established survey scales designed for measuring both kinds of populist attitudes, testing the following research questions: First, how do political and science-related populist attitudes (and different dimensions thereof) relate to each other? Second, what predicts differences between these attitudes? Third, how do they differ in predicting related attitudes? Drawing on a cross-sectional analysis of two waves of a large-scale panel survey (N = 1,500 per wave) fielded during the COVID-19 crisis in Austria, we show that political and science-related populism indeed strongly correlate but capture different nuances of populist ideation. We also find that both variants have similar antecedents and differ in their capability to predict related constructs. We then discuss implications for future research using the scales in survey and experimental research, recommending that researchers need to (1) think about which of populism variant is (or whether both are) relevant to their research question; (2) develop different theoretical and statistical models to predict either populism variant; (3) scrutinize which variant is theoretically better suited to explain their dependent variables.

Assessing Population Support for Aspects of Syringe Service Programs in a Rural State

Patrick Habecker, *University of Nebraska-Lincoln*
Ryan Herrschaft, *University of Nebraska-Lincoln*
Bergen Johnston, *University of Nebraska-Lincoln*
Rick A Bevins, *University of Nebraska-Lincoln*

Syringe service programs (SSPs), though well demonstrated to reduce cases of HIV and hepatitis C (HCV) among people who use drugs, are politically contentious operations. In six recent HIV outbreaks in the US between 2016 and 2019, either launching or expanding access to syringe services programs was an early response to new HIV cases in every location. Although the primary purpose of SSPs is to promote access to safer injection resources, these facilities frequently offer additional health care services, such as screenings for HIV and HCV, and linkage to other care.   Even though SSPs are an evidence-based intervention for increasing community health, efforts to establish these facilities are often met with political opposition and community stigma. Negative attitudes about SSPs have not only prevented the establishment of SSPs in many states but have also caused the closure of many operational SSPs that were previously an integral resource in their communities.   In this project, we examine support for SSPs among the general population of Nebraska – a state where SSPs and sterile syringe possession for nonmedical purposes are illegal. We surveyed a representative sample of Nebraskans to determine attitudes toward SSPs and support for various services that may be provided by these facilities. Variation in support is assessed by rurality, previous personal use of illicit drugs, and self-reported stigma towards people who use various types of drugs. Assessing these relationships in Nebraska provides us with a baseline understanding of general support for SSPs in a state where these facilities are currently prohibited. This analysis is important due to rising rates of injection drug use in Nebraska, where significant HIV or HCV outbreaks have not yet been reported but could be prevented with this type of evidence-based program.

A New Perspective on Health Inequity

Orin Puniello, *MWWPR*
Michelle Toennies, *Edelman Dxl*
Joey Krutov, *Edelman Dxl*
Maggie Mclaughlin, *Edelman Dxl*
Emily Wagner, *Edelman Dxl*
Cheresa Wright, *Genentech*

Inequity in our healthcare system is an urgent issue. Previous research on healthcare inequity focuses on census data, healthcare outcome data, and the perspective of HCPs, but the direct voice of medically disenfranchised patients is largely missing. Therefore, we sought to study medically disenfranchised individuals and reveal how this long-standing inequity impacts their relationships within the healthcare system. While many studies have probed into the complex causes and effects of health inequity, few are grounded in the lived experience of patients. In support of Genentech's ongoing commitment to advancing Health Equity, the company partnered with Edelman DXI (on behalf of Genentech) to develop a multiyear study of more than 2,200 patients in the United States yearly: 1000+ members of the general population and 1,200+ who qualified as medically disenfranchised from key communities (Black, Latinx, LGBTQ+, and low socioeconomic status). This past year we included 300 medically disenfranchised individuals from the Asian American community. Our core findings revealed a crisis of trust in our healthcare system among those who need it most. Medically disenfranchised patients believe, quite simply, that the healthcare system treats patients unequally and unfairly, and is even stacked against them. While half (49%) of the U.S. general population feels all patients are treated fairly and equally, based on our study, only 27% Black, 27% Latinx, 34% LGBTQ+, 23% low SES agree that all patients are treated fairly. Furthermore, medically disenfranchised patients are delaying and discontinuing routine care because they do not feel understood. The implication of this research is clear that we must build bridges to medically disenfranchised patients to make them feel valued, respected, and understood. The results of this research have informed key discussions around health equity, and are available to health care professionals/advocates, governments, and researchers to support them when developing policy around healthcare inequities.

Primary Care Physicians, State Governments, and COVID-19 Responsibility and Response

Kirby Goidel, *Texas A&M University*
Timothy Callaghan, *Texas A&M University*

David Washburn, *Texas A&M University*
Tasmiah Nuzhath, *Texas A&M University*
Abigail Spiegelman, *Texas A&M University*
Julia Scobee, *Texas A&M University*

Emerging health crises, by their nature, challenge and overwhelm federal political systems. Within the context of the COVID-19, states and governors were called on take charge in the absence of coordinated federal response (Bowman & McKenzie, 2020; Weissert, Uttermark, Mackie, & Artiles, 2021). The result was uneven policy responses both in terms of efforts to control the spread of the virus (economic lockdowns, masks mandates, and social distances) and the health-related and economic consequences of the pandemic (Warner & Zhang, 2021; Zhang & Warner, 2020). Birkland et. al. (2021) refer to this patchwork of fragmented policy responses as "variable-speed, go-it-alone" kaleidoscope federalism. While existing work has explored public evaluations of state government policies aimed at minimizing the public health and economic consequences of the pandemic, we explore how primary care physicians evaluated state responsibility for the pandemic and the efficacy of state government policies. Using an original survey of 625 primary care physicians, we find that, like the general public, physician attributions of responsibility and evaluations of state government response were driven more by partisan affiliation and ideology than by the effectiveness of state policies.

Frames & Bias: Measuring Coverage & Error in Household Surveys
Untangling Representation Error in Address-based Sample Surveys
Robyn Rapoport, *SSRS*
Cameron McPhee, *SSRS*
Robert Manley, *SSRS*

As household surveys continue to shift away from random-digit dial sample frames and interviewer-administered data collection modes towards address-based sample frames and self-administered modes, representation error seems to be changing as well. One notable example of this is gender. Historically, with dual frame RDD surveys, landlines have tended to underrepresent males, since women are more likely to answer the phone in households while we often obtain disproportionately more interviews with men on cellphones . Interestingly, it seems that some ABS studies using a push-to-web or mixed-mode strategy provide a near-Census split of males and females, while other ABS studies are underrepresenting males in their responding samples. It is well documented that quasi-random within-household selection methods, such as the last birthday method, are not always complied with (e.g., Olson, Stange, and Smyth, 2014, Olson and Smyth, 2017), leading to one source of potential bias. In ABS studies this bias may not be systematic and may be directly related to the survey topic. For example, out of 10 ABS push-to-web studies administered by SSRS in 2021, the unweighted bias by gender ranged from near zero to as high as 17 percentage points, with the largest discrepancy appearing with surveys addressing health and health-related topics. This paper seeks to better understand the drivers of these biases. We will examine a set of national and regional ABS surveys administered within the previous three years to determine if the respondent gender bias can be explained by study design factors including survey topic, mailing method, within-household selection protocol, as well as potential interactions between these study characteristics. By disentangling some of the complexities of this type of error, we will be able to point to some important levers for improving representativeness in ABS studies.

Comparing a Probability and Non-Probability Within-Household Respondent Selection Method for an Address-Based Web-Push Survey
Jesus Arrue, *Westat*
David Cantor, *Westat*
Ting Yan, *Westat*
Jieru Chen, *Centers for Disease Control and Prevention*
Sharon G. Smith, *Centers for Disease Control and Prevention*

This presentation reports on the results of an experiment comparing a probability and non-probability method to select a respondent for a web-push survey using an address-based sample (ABS). The tradeoffs between probability and

non-probability methods have been investigated in the context of telephone and mail paper surveys (Marlar, et al., 2018; Yan et al., 2015), but there has not been as much research for ABS web-push surveys. The experiment was part of a research project for the Centers for Disease Control and Prevention. The experiment involves administering a survey that covers topics related to health and violence victimization. The experiment sent requests to sampled households to complete a screening survey on the web to select a respondent for the main survey. One method asked the screener respondent to follow the Rizzo-Brick-Park procedure (Rizzo, et al, 2004), which is considered a probability method. The non-probability method used the Youngest Male Oldest Female (YMOF) approach. This randomly divides the sampled households into four groups based on who should be asked to be the respondent. The groups are based on sex (male, female) and age (oldest, youngest). One advantage of the YMOF method is that it targets groups that are traditionally under-represented in ABS surveys (young people, males). The disadvantage of this method is that it is not a full probability method. For example, it does not give everyone in all households a non-zero chance of selection. Overall, the analysis did not find significant differences between the probability and non-probability method. There was no significant difference in the response rate when using the two methods (33.8 vs. 32.4). There were no significant differences in the distribution of key demographics (e.g., age, sex). With one exception, there were no significant differences with respect to topline estimates of violence victimization.

Reinforcing Predictions: Replicating Analytic Models Across Samples and Modes

Madeline Gosz, *Ipsos*
Randall K. Thomas, *Ipsos Public Affairs*
Frances M. Barlas, *Ipsos Public Affairs*

There are generally two main types of online samples used in polling – probability-based and opt-in non-probability samples. Though online samples are often used for general population national studies, there has been increasing interest in their use for smaller groups or geographies. In addition, there are increasing attempts to develop alternative methodologies that can target specific areas, including redirected inbound calling sample (RICS), which uses IVR technology to present and receive answers that can identify cases relevant to a study. To compare these various sample sources, we conducted a political poll in Georgia between August 13 th 2020 and September 8 th 2020 using three separate sample sources – 1) probability-based KnowledgePanel (KP), 2) an opt-in non-probability online sample, and 3) a RICS sample methodology. For KP, opt-in, and RICS samples, we had 528, 1,331, and 1,185 participants, respectively. In this study, we were interested in examining how predictive multiple regression models could obtain similar results with the different samples/modes.   We had three dependent variables of interest: Party ID, Presidential candidate choice, and Senate candidate choice. We used a number of political attitudinal items (e.g., right-wrong direction, presidential approval) as predictors. We found that, while the models identified generally common strong predictors that had similar strength and direction of effect, some weaker predictors were more variable in their influence across the dependent variables and in the different samples. We discuss some of the limitations of using these different sample sources for use in surveys for targeted geographies and their methodological and analytic challenges.

Migrating from Phone to Online as Part of a Probability-based Multinational Project

Alexandra Castillo, *Pew Research Center*
Patrick Moynihan, *Pew Research Center*

Pew Research Center has fielded the annual, cross-national Global Attitudes Survey for two decades, surveying publics in around 20 to 40 countries per year via phone and face-to-face methods. As the marketplace for survey research has evolved over time internationally in response to respondent accessibility, data quality, in-country capacity, cost concerns and – more recently – the global pandemic, online probability-based panels have become increasingly available. To assess the viability of this mode of data collection outside of the United States, including its comparability to long-term designs and estimates using interviewer-administered surveying, the Center surveyed via panels in parallel to our 2020 and 2021 waves of the Global Attitudes Survey. Operationally, our standard random-digit dial (RDD) computer-assisted telephone-interview (CATI) surveys were fielded alongside nationally representative, probability-based online panels in Australia and the United Kingdom. The RDD questionnaire for each country was adapted to field online and allow for comparisons of sample demographics and attitudinal measures. This paper will provide insight into the cross-comparability of results from these independent panels, considering the need to standardize (among other

issues): questionnaire platform design to minimize differences in measurement; sampling methods to reduce biases in nonresponse; and weighting procedures to consistently correct for known panel deficiencies. Additionally, we will consider how a cross-national mode change could affect within-country trends: the differences in measurement across modes, such as scaled responses and item nonresponse; the comparability of sample composition across the phone and panel surveys; and, ultimately, the viability of continuing time-series estimates on specific measures using a different mode. Finally, we discuss possible strategies to preserve cross-national comparisons for future rounds of this international project with face-to-face, phone and online data collection as options.

How Can We Reach You? Strategies for Mixed-Mode Surveys

Some Say Branding Is Everything, But Is It? Experimenting With Branded Survey Materials

Kyle Endres, *University of Northern Iowa*
Erin O. Heiden, *University of Northern Iowa*
Ki Park, *University of Northern Iowa*
Mary E. Losch, *University of Northern Iowa*

Conventional wisdom recommends including government and University logos on the envelopes of mailed survey materials to improve response rates. However, some recent studies suggest omitting the sponsor's logo from envelopes improves response rates, presumably by increasing the likelihood envelopes are opened. We evaluate how branded survey materials affect response rates for parallel statewide surveys fielded in late 2021 using a probability-based sample of dual-frame random digit dial (DF-RDD) and an address-based sample (ABS). We sent each sampled address up to three mailings beginning with a Push-to-Web (PTW) letter, followed by a self-administered mail back (SAM) packet with a PTW option, and a final PTW letter. We randomized whether we sent each address mailings in envelopes branded with the logo of the University conducting the survey or no logo (return address only). We also randomized whether or not the shortened, unique URLs (used for the PTW option) were customized with the acronym of the University research center or with the name of the company providing the free URL shortening service. Preliminary analyses reveal a significantly higher response rate for addresses randomly assigned to receive envelopes without the University's logo (and only the research center's return address). Customizing the URLs, on the other hand, had minimal effect on response rates.

Screening Households to Select a Sample: Evaluating a Multi-Mailing Contact Strategy

Jessica Graber, *National Center for Health Statistics*
Jay Clark, *Westat*
Kevin Wilson, *Westat*
Allan Uribe, *National Center for Health Statistics*
Juliana McAllister, *National Center for Health Statistics*

The National Health and Nutrition Examination Survey (NHANES) is a nationally representative survey that, since 1999, has used in-person household screening to identify sampled persons. In-person surveys require trained interviewers to conduct them, which can be an expensive and inefficient use of resources. These considerations – coupled with the safety concerns raised by the COVID-19 pandemic, and the general survey environment of declining response rates – led NHANES to implement a multi-mode household screener in the 2021-2022 survey cycle. Sampled dwelling units received four mailings inviting them to respond via a secure website with a unique pass code. All mailings also included a toll-free number that the respondent could call to complete the screener by telephone. The mailings spanned six weeks and included: (1) an initial invitation letter with a study-branded magnet; (2) a reminder postcard; (3) a FedEx package that included a letter and a paper version of the questionnaire with a postage-paid return envelope; and (4) a final letter that, in addition to the website and phone response options, informed the recipient of an upcoming visit by a field interviewer. Responding addresses were removed from subsequent mailings. Nonresponding addresses were visited by a field interviewer who attempted to administer the screener in person. In this presentation, we examine completed screener cases to identify which contact prompted the successful response and identify any variation in mode of response, respondent or household characteristics, including household size, demographic characteristics of the respondent, and number of people sampled. We will also report on available instrument paradata by mode, such as

time to complete, device type and time of day of completion. Finally, we compare responses by mode to identify any systematic mode effect.

Evaluating the Effectiveness of Pressure-Sealed Mailing Materials in a Web-Push Design

Jiashan Cui, *American Institutes for Research*
Mahi Megra, *American Institutes for Research*
Tzu-Jou Wan, *American Institutes for Research*
Rebecca Medway, *American Institutes for Research*

The results of the 2017 National Household Education Survey web test (NHES:2017) suggested a pressure-sealed envelope could be a promising option as a survey reminder (Medway et al. 2018). However, the NHES:2017 administration was web-only and thus not entirely comparable to a mixed mode web-push design. In addition, it was not a randomized experiment, in which a reminder postcard and pressure-sealed envelope are tested simultaneously within the same data collection. The NHES:2019 administration predominately used a web-push design and experimented with the use of a pressure-sealed envelope as a survey reminder. For this experiment, sample members in the baseline group received a regular reminder postcard, while the sample members in the treatment group received a pressure-sealed envelope. Both types of survey reminders were sent a week after the initial screener package. The NHES:2019 pressure-sealed envelope included the web survey URL and the household's unique web login credentials, but the regular reminder postcard did not (because the postcard format did not allow for sufficient protection of this information). It was hypothesized that the pressure-sealed envelope's ability to include the survey URL and the household's web login credentials would increase internet response (Census, 2018). We will present, within the context of a web-push design, whether a pressure-sealed envelope reminder motivated survey participation. The results show that it is more effective to send a pressure-sealed envelope (a 4-percentage-point difference). We also evaluated the effect on response rates by selected household characteristics. There were a few subgroups for which sending a pressure-sealed envelope led to a somewhat larger-than-average increase in response relative to the reminder postcard. Our presentation will provide insights into the effectiveness of pressure-sealed mailing materials in a web-push design.

Investigating Timing and Frequency of Email Reminders in an Establishment Survey

Joshua Langeland, *U.S. Bureau of Labor Statistics*

The Occupational Employment and Wage Statistics program, at the Bureau of Labor Statistics, conducted an experiment designed to measure the impact of alternatively timed Email nonresponse follow-up (NRFU) regimes. A subset of establishments sampled for the survey with Email addresses available were randomized into one of four treatment groups. The treatments varied whether an establishment received an Email NRFU at one week after invitation, two weeks after invitation, or both one and two weeks after the initial invitation. A fourth cohort acted as a control group and did not receive any Email NRFU reminders. All cohorts received the survey invitation through Email and all units received a postal mail follow-up after a five-week period. The fully crossed experiment design allowed investigators to measure the impact of both the timing of Email nonresponse follow-up contacts and the number of follow-up contacts. Analysis presented will focus on response rates and time to response. Results found that increasing the number of reminders increased response (two reminders had the highest response rate, zero reminders had the lowest) and sending reminders earlier encouraged faster response (sending the Email reminder after one-week increased timeliness of response).

Transition from Telephone Mode to Mixed-mode Surveys: A Case Study

Chintan Turakhia, *SSRS*
Jennifer Su, *SSRS*
Cameron McPhee, *SSRS*
Jennifer Agiesta, *CNN*
Ariel Edwards Levy, *CNN*

While traditional telephone surveys have provided valid and reliable data for decades, they are faced with many challenges including declining response rates over time, risk of social desirability bias, interviewer effects, house effects,

and increasing costs. A combination of these challenges can threaten the validity of a telephone poll and data trends. These challenges with the telephone mode coupled with internet usage at all time high (Pew Research Center Reports internet usage in 2021 at 93%, highest level since they started measuring it in 2000) and increased smartphone usage have accelerated the adoption of online and mixed-mode data collection methodologies in recent times. People have started spending more time online in the past year and half than ever before, smartphone adoption and usage are at record levels, and internet access increases as mobile providers continue to provide larger and more affordable data plans. An AAPOR task force report on transitioning from telephone to mixed mode acknowledges that with this change in environment, an increasing number of surveys are transitioning from telephone only to self-administered modes such as online. However, the empirical literature is just beginning to provide guidance to survey researchers about best-practices for these transitions and there remain many unknowns.   We wish to broaden this knowledgebase by presenting a case study on transitioning from an RDD-based telephone methodology to an ABS-based mixed-mode (mail push to web with phone) methodology in the context of political polling. We provide a demographic comparison of the RDD poll and ABS poll, discuss strategies to reduce non-response bias, particularly among hard-to-reach groups, and discuss innovative use of auxiliary data from the ABS frame to target under-represented populations. Finally, we examine questionnaire-related issues and adjustments that may be needed when transitioning from telephone only to mixed-mode

Anybody Home? In-Person Recruitment in a Multimode Design on the Medicare Current Beneficiary Survey (MCBS)

Rachel Carnahan, *NORC at the University Of Chicago*
Alanah Raykovich, *NORC at the University of Chicago*
Holly Hagerty, *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. For nearly 30 years, the MCBS interview has been conducted in-person via computer-assisted personal interviewing. In response to the coronavirus (COVID-19) pandemic, data collection for the MCBS rapidly shifted from in-person to phone interviewing in March 2020. This led to a revised design incorporating both phone and in-person outreach and shifting to a multimode data collection design. As a continuous, longitudinal survey of a nationally representative sample of the Medicare population, a new panel of beneficiaries is selected to join the survey every fall. In Fall 2021, the MCBS began to gradually reintroduce in-person data collection activities, starting with in-person recruitment for the purpose of non-response follow-up. This return to field activities presented an opportunity to analyze the outcomes of in-person outreach as a companion for phone recruitment. This work can inform future decisions about how and when to focus resources on in-person versus phone outreach, particularly for the recruitment of new panel members. This presentation will describe the protocol used and discuss the outcomes associated with in-person outreach. The analysis will consider differences in final case status or contacting patterns as a result of respondent material packet drop-offs versus short in-person conversations, as well as variation by demographic and other health status subgroups. While limited in size and geographic area by the ongoing pandemic, analysis of this return to field can illuminate the challenges and implications of a new approach to panel recruitment for the MCBS and beyond

In It for the Long Haul: Changing Modes on Longitudinal Surveys
Modeling Mode Effects for a Panel Survey in Transition from In-Person to Web-Main Mixed Mode

Dan Liao, *RTI International*
Brian J. Burke, *RTI International*
Paul P. Biemer, *RTI International; The University of North Carolina at Chapel Hill*
Kathleen Mullan Harris, *The University of North Carolina at Chapel Hill*
Carolyn Tucker Halpern, *The University of North Carolina at Chapel Hill*

Over the last two decades, in-person interviewing costs continued to increase while the data quality advantages traditionally identified with this data collection mode have faded. Consequently, some longitudinal surveys have begun transitioning from in-person to web data collection despite potential risks to data quality and longitudinal comparability. This was the case for the National Longitudinal Study of Adolescent to Adult Health (Add Health). Funding

reductions combined with increasing data collection costs motivated the Add Health study to transition from its traditional approach of in-person interviewing to a more cost-effective approach at Wave V. The new approach includes the mail/web data collection mode in Phase 1 and in-person interviewing with a random sample of nonrespondents in Phase 2. In addition, a small, random, in-person subsample was selected in parallel to serve as the control for a mode comparison study. As will be shown in this presentation, with this parallel design, a Markov Latent Class model can be fitted to Waves III, IV, and V to separately estimate the measurement error effects of the mail/web and in-person modes. Our results showed that concerns about reduced data quality due to the redesign effort were unfounded. In several important respects, the new two-phase, mixed-mode design outperformed the traditional design with greater measurement accuracy, improved weighting adjustments for mitigating the risk of nonresponse bias, reduced residual nonresponse bias, and substantially reduced total mean squared errors of the estimates. However, the study results suggest that significant differences can occur in longitudinal change estimates purely as an artifact of the redesign. This begs the question: how, then, should a data analyst interpret significant findings in a longitudinal analysis in the presence of mode effects? This presentation will summarize the analysis results and attempt to address this question.

Mode Assignment Effects on Response Rates and the Validity and Reliability of Measures of Cognitive and Physical Health

Eric Grodsky, *University of Wisconsin-Madison*
Jennifer Dykema, *University of Wisconsin-Madison*
Shana Brown, *NORC*
John Stevenson, *University of Wisconsin- Madison*
John Robert Warren, *University of Minnesota*

In order to stay within budget and maximize response rates, the High School and Beyond panel study (HS&B) elected to pursue a mixed-mode design (web and CATI) for the current wave of data collection. We evaluate the effects of initial mode assignment (CATI vs. web) on response rates and on the validity and reliability of measures of physical and cognitive health by employing a randomized blocked design to assign initial mode of collection to a subset of sample members. We restrict our attention to the ~13,700 (of ~25,500) sample members who (1) were believed to be alive at the start of the 2021 field period, (2) were not hostile refusals in the prior wave of data collection, and (3) for whom we believed we had good contact information. We estimate the effects of mode assignment (intent to treat) and mode of participation (treatment on the treated) on response patterns and response rates. The mode assignment results do not require additional assumptions as they are based on experimental manipulation. The participation results rely on adjustments for the propensity to respond, but based on an extensive set of controls extending back to adolescence for respondents approaching 60 years of age. To evaluate the validity and reliability of items that constitute scales of physical and cognitive health, we employ the measurement framework proposed by Bollen (1989) and employed in Cernat, Couper, and Ofstedal (2016).

Sample Design and Implementation for the Fourth Round of the National Social Life, Health and Aging Project: Remote Data Collection During COVID-19

Ned English, *NORC at the University Of Chicago*
Colm O'Muircheartaigh, *NORC at the University of Chicago*
Anna Wiencrot, *NORC at the University of Chicago*
Becki Curtis, *NORC at the University of Chicago*
Katie O'Doherty, *NORC at the University of Chicago*
L. Philip Schumm, *The University of Chicago*
Stephen Smith, *NORC at 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, 2021). Currently in its fourth round, NSHAP is designed to follow 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. The current round represents a departure from the past in that we adapted a primarily face-to-face data collection approach to include self-administration for a sample of prior-round respondents. While

such a change was eventually necessitated by the COVID-19 pandemic, NSHAP had already been planning to implement remote data collection for some respondents. In this paper, we describe the factors taken into account in assigning respondents to start either remotely or face-to-face. These factors include past participation in NSHAP, availability of contact information, refusal history, etc. We also discuss our protocols for reinterviewing both cohorts, methodological results for the main study thus far as well as a COVID-19 sub-study administered in 2019-20; and, analytical opportunities for data users.. Our results show the promise of collecting detailed information in remote survey designs for lengthy questionnaires, in our case approximately 60 minutes, along with biomeasure collection. Our paper will be useful to survey practitioners given the need for remote study adaptation is likely to increase in importance after the COVID-19 pandemic for all age groups and study domains.

Mode Effect Adjustment in a Longitudinal Study

Xiaoshu Zhu, *Westat*
Ting Yan, *Westat*

The outbreak of COVID-19 led to unexpected changes to survey data collection. Many in-person surveys were forced to pause data collection or to switch to another mode of data collection when face-to-face contact was not allowed or feasible during the pandemic. Mode effect occurs when estimates differ by the mode the data is collected. Mode effect is a major concern in mixed mode surveys and research is needed to evaluate approaches to adjust for mode effect statistically. One approach of mode effect adjustment is to regress survey outcomes on mode and other relevant variables. Another approach is to weight respondents by propensity scores to adjust for mode effect. This paper evaluates the performance of these two approaches using a synthetic dataset generated from a longitudinal survey. Longitudinal surveys provide rich information on survey outcomes and auxiliary variables over time, all of which can be utilized to improve estimates and adjustment of mode effect. Initial data exploration indicated potential age difference on mode effect. This paper intends to compare the performance of the adjustment in three age groups, i.e., youth(12-17 years old), young adults (18-24 years old) and adults (25 years and older). External surveys conducted at the similar time period are selected as benchmarks to assess the recovery of response distributions from the two approaches.

Data Science Techniques to Efficiently Process Text Data

Can I Have Your Name? Classification of Names for Case Prioritization in Household CAPI Surveys

Xin (Rosalynn) Yang, *Westat*
Anil Battalahalli Sreenath, *University of Michigan*
Ting Yan, *Westat*

In household CAPI surveys, interviewers are instructed to collect sampled respondents' names among other contact information as part of the screener. Depending on the case, interviewers may be able to get respondents' full names, first names, or last names. When interviewers are not provided with any information, they tend to assign generic labels based on sampled respondents' characteristics (e.g., husband, daughter) and follow up later. This type of information is highly predictive of interview response propensity – in a prior study, we found that respondents who gave full names were significantly more likely to complete the interview. For household CAPI surveys where case is prioritized based on response propensity, it is of interest to improve the propensity model with as many correlates as possible, including the name information. To quickly process a large amount of name data in a short time, we proposed the use of natural language processing and machine learning methods. We trained a bi-directional character level LSTM model to classify names into whether it is a full name, a first name, a last name, or a generic label. In this paper, we will describe our classification model and examine its accuracy and efficiency when compared with traditional methods (e.g., dictionary based) and human coding. Deployed as a REST API, our classification model can be easily incorporated into any propensity model pipelines in surveys involving case prioritization efforts.

Using Natural Language Processing to Help Develop a Frame of Energy Suppliers

Cindy Good, *Westat*
Meghan Martin, *Westat*
Francisco Cifuentes, *Energy Information Administration (EIA)*
Michelle Amsbary, *Westat*

Applications of natural language processing (NLP) don't have to be confined to complex sentences or paragraphs; natural language comes in small chunks too – such as energy supplier names as reported by household respondents in open text fields. The frame for the Residential Energy Consumption Survey (RECS) Energy Supplier Survey (ESS) is developed by identifying the correct name of the energy suppliers reported by respondents in the RECS household survey. For the most recent cycle of the study, approximately 19,000 household respondents provided nearly 30,000 entries of electricity, natural gas, fuel oil, and propane supplier names in open text fields of web and paper survey instruments. A single energy supplier might be reported dozens of different ways, due to abbreviations, acronyms, nicknames, misspellings, and typos. In prior RECS ESS cycles, human coders manually compared these name variations against a reference list of known energy suppliers to clean up and deduplicate entries. This was a labor-intensive and time-consuming task. To reduce time and costs and increase efficiency, Westat turned to NLP. This presentation will describe how we programmed and utilized NLP to match energy supplier names reported in the household survey to those on our reference list of suppliers. We discuss developing and fine-tuning the matching algorithm using an iterative design process to improve outcomes, our quality control review on the output, and key results. Additional evaluation of the matching after data collection will also be examined, as well as recommendations for using similar NLP matching for frame development in future ESS cycles.

Using Supervised Classification for Categorizing Answers to an Open-Ended Question on Survey Participation

Anna-Carolina Haensch, *Ludwig Maximilians University Munich*
Bernd Weiß, *GESIS Leibniz Institute for the Social Sciences*

Our research illustrates how supervised learning models can be used to categorize a considerable number of answers to an open-ended question on survey participation motivation. This question on survey participation motivation has been asked annually in the GESIS Panel since 2014. The GESIS Panel is a German probability-based mixed-mode access panel that has around 5,400 panelists. Over time, around 25,000 answers based on six panel waves have been collected. The respondents' answers to the open-ended answers are of high interest since they offer us more insights into panel participation, the underlying motives of panelists in particular, and, hence, offer possibilities for better panel management in the future. The coding scheme, which is utilized to categorize the survey participation, is based on similar coding schemes in the literature but has been adapted to fit the needs of the GESIS Panel. The number of answers (25,000) would make coding by hand very time-intensive. We, therefore, opted to code only a much smaller sample of several thousand answers by two coders and used a supervised classification algorithm to classify the remaining answers. The answers to the open-ended questions in the GESIS Panel are particularly suited for automated classification since they are, for the most part, one-dimensional. 10-fold evaluation results show excellent performance; macro measures are around 0.85 for precision, recall, and the F measure. Also, for most of the categories, micro measures are good. Significant parts of the panelists are motivated by interest, curiosity, incentives, and the need for recognition, but also by the wish to help the scientific community, political leaders, or even society in general.

Expanding Your Toolkit With Web-Scraping and Content Extraction

Angelina KewalRamani, *American Institutes for Research*

Due to the COVID-19 pandemic, elementary and secondary schools and post-secondary institutions were in dire need of support from the federal government. The Coronavirus Aid, Relief, and Economic Security Act, or CARES Act, was passed by Congress in March 2020 and provided funding to U.S. schools and institutions of higher education. Federal government agencies needed to distribute funding rapidly and subsequently collect important indicators from post-secondary institutions that received funding in very short timeframes. However, the majority of institutions were reporting data in unstructured PDF reports and posting the reports on their websites. Identifying the reports from nearly 5,000 institutions and extracting the relevant data using conventional techniques would have proved time consuming. We employed innovative and efficient methods to identify the reports and extract the data. This research will describe the process of developing code for a data pipeline that performed web-scraping and content extraction for the PDF documents.   We supported the U.S. Department of Education in determining the uses of funds authorized to address challenges facing institutions of higher education resulting from the pandemic. Our team used an automated process to access the student and institutional portions of the CARES Act data reported through almost 5,000 websites. We used web scraping techniques to identify and collect information from CARES Act reports from approximately 4,900

postsecondary institutions across the country. Our team used specific search terms to identify relevant sections from unstructured reports. We performed text analysis and content extraction methods to identify answers for various text and numeric questions within the report. We conducted multiple rounds of quality control to fine-tune the content extraction algorithms and validate the accuracy of the results. Our web scraping program was developed in an open-source programming language. The findings will help researchers evaluate different methods of finding and extracting data

Using Social Listening to Explore Effect of Work from Home on Loneliness
Edward Johnson, *Harris Poll*

While our society might be more connected than ever we are still seeing a considerable amount of loneliness in society (Cacioppo et al, 2009). As there is a connection between social media activity and loneliness levels (Park et al, 2016), we wanted to use social listening to uncover any potential relationship between working from home and loneliness.   We started by using a social listening platform called Talkwalker. We review the sources of information and isolate those who we believe are reliable (e.g. not spam bot). We monitor key words such as "loneliness" and "work from home" and look at the volume of these levels over time. We also use basic sentiment analysis to see how "work from home" might be perceived over time. While we would not recommend using this to draw inferences to the entire US population due to the nature of the sample, attendees will learn: 1.   Some of the basic outputs from social listening platforms 2.   How to effectively curate the data coming from the social listening platforms 3.   Draw actionable insights from the results of the data scraping exercise   Cacioppo, J. T., Fowler, J. H., & Christakis, N. A. (2009). Alone in the crowd: the structure and spread of loneliness in a large social network. Journal of personality and social psychology , 97 (6), 977. Park, S., Kim, I., Lee, S. W., Yoo, J., Jeong, B., & Cha, M. (2015, February). Manifestation of depression and loneliness on social networks: a case study of young adults on Facebook. In Proceedings of the 18th ACM conference on computer supported cooperative work & social computing (pp. 557-570)

Improving Response Rates & Evaluating Nonresponse Bias

The Relation Between Nonresponse Rates and Nonresponse Bias. An Update and Extension of Groves and Peytcheva (2008)
Peter Lugtig, *Utrecht University*
Bella Struminskaya, *Utrecht University*
Annemarie Timmers, *Utrecht University*
Celine Henneveldt, *Utrecht University*

Groves and Peytcheva (2008) showed that the relation between nonresponse rates and nonresponse bias is weakly positive. Their meta-analysis of almost 50 studies showed that higher nonresponse rates do lead to higher nonresponse bias, but higher response rates aren't always better when it comes to the extent of nonresponse bias. As the response rate is still often used as a heuristic for the risk of nonresponse bias, the findings of Groves and Peytcheva have proven important. The study of Groves and Peytcheva relies on studies conducted mostly in the 1980s and 1990s conducted via face-to-face and telephone (RDD) surveys. The purpose of the current study is twofold: 1. To update the meta-analysis with about 100 new studies conducted in the period 2000-2020, and study whether the relation between nonresponse rates and nonresponse bias has become stronger or weaker. Anecdotal evidence suggests that the steady decline in response rates has often not led to much higher nonresponse bias, implying that the relation between the two is perhaps even weaker as before. But is this really true? 2. Because of the large number of added studies, we can include many more covariates that help explain when the relation between nonresponse rates and nonresponse is weak or strong. We coded a wide range of survey design characteristics on modes, contact protocols, and key outcomes in every study. Using this covariate information, we can study design characteristics such as the effect of reminders on both nonresponse rates and nonresponse bias, mixed mode studies, and other measures that are generally taken to increase response rates. During the conference, results will mostly presented graphically.

Response Rates -- Traditional Mailed Surveys vs. Mailed Invitation to Online
Parker Quinn, *Polco*

Response rates to resident surveys have been declining for all modes of data collection, increasing data collection costs. For the last year, researchers from the National Research Center at Polco have been conducting split sample data collection modes for our community surveys. One portion of the sample is assigned to a "traditional mailed survey" methodology with three contacts per randomly selected household: a prenotification postcard, a survey packet with cover letter and hard copy survey with postage-paid return envelope, and a second wave survey packet with reminder letter, hard copy survey and postage-paid return envelope. The other portion of the sample is assigned to a "mailed invitation to online survey" methodology, in which the randomly selected households are mailed an initial and reminder postcard invitation with a URL to the online survey. The results will be analyzed to answer four research questions: How do the response rates for each methodology compare? What differences are there in the unweighted demographic profiles for each? What differences are there in the ratings of local government services, ratings of community quality of life, and in rates of participation in community and civic activities? What is the cost per complete for each group, on average?

Increasing Response Rates: How Different Mailing Materials and Protocols Affect Survey Response Rates

Carol Cosenza, *University of Massachusetts - Boston*
Kathryn Bell, *Center for Survey Research - UMass Boston*

How to increase response rates is an important topic within survey research. Researchers constantly are creating new approaches to encourage survey participation and increase overall response rates. This project evaluates how different mailing methods, such as varying envelope size and labels, order of mailed materials, use of postcards or push-to-web letters, or number of mailings affect the response rates of an address-based sample. This survey focused on respondents' experiences during Hurricane Harvey and COVID-19. A sample of residents from 88 different neighborhoods in Houston received mailings between November 2020-June 2021. There were 8 mailing groups created from a sample total of 11,773. The original sample consisted of 4400 addresses. To help increase the number of returns, additional sample was added during the 8-month field period. As more sample was added, new mailing groups were created. All groups received $2 with their first contact. Respondents then received an additional $10 for completing the survey either by mail or web. Some mailings included a 'Hurricane Harvey Experiences Survey' label on either the bottom left of the envelope or on the address label. All groups received a second contact with a postcard that contained the survey weblink. Not all groups received a third contact. Each mailing group got a variation of the following mailing materials: 1.   10x13 envelope   2.   #9 envelope   3.   6x9 envelope   4.   Survey packet with letter, factsheet, paper survey, and return envelope   5.   Postcard with weblink   6.   Push-to-web letter with no paper survey   7. 'Hurricane Harvey Experiences Survey' label on either the envelope or address label   8.   No 'Hurricane Harvey Experiences Survey' label We will report on response rate variation by group and which mailing protocol or mailing order was best. In addition, we will report on how push-to-web letters affected the returns via web and the overall response rates.

COVID-19's Effects on RDD Sample Productivity: Shock and On-Going Effects

Matt Jans, *ICF*
James Dayton, *ICF*
Randy ZuWallack, *ICF*
Don Allen, *ICF*
Andy Dyer, *ICF*
Thomas Brassell, *ICF*
Samantha Collins, *ICF*
Traci Creller, ICF
Zoe Padgett, ICF

COVID-19 has had a complex impact on survey productivity and response rates, but presents a type of "natural experiment" through which we can learn about a pandemic's effect of sample productivity. During the first year of COVID, we observed possible shock effects (i.e., large increases in March 2020) on RDD contact response, and refusal conversions, but had difficulty identifying ongoing effects. By the end of 2020, these sample productivity metrics had largely returned to expected levels, although there was still large between-surveys variability on most metrics. The current presentation extends our past COVID effects evaluation (presented at AAPOR 2021) by adding sample

productivity from 2021 for the same 12 surveys evaluated previously. In addition, it adds statistical tests of shock and on-going effects (i.e., tests of differences in productivity rates between months or groups of months, and tests of between-survey variability). Findings are discussed in the context of COVID-19's impact on data collection and the general difficulty of disentangling effects from natural experiments. Results should help survey organizations understand COVID effects and plan for future situations, large and small.

Observing and Addressing the Relationship Between Vaccination Rates and Nonresponse Bias

Eran Ben-Porath, *SSRS*
Ashley Kirzinger, *KFF*
Liz Hamel, *KFF*
Jazmyne Sutton, *SSRS*
Akilah Evans-Pigford, *SSRS*

Over the course of the Covid-19 pandemic, tracking polls (cross-sectional surveys asking the same question over time) have provided an ongoing assessment of the progress and impact of the unprecedented vaccination effort in the U.S. A challenge to this polling effort has been the likelihood that self-reported vaccine uptake will be positively correlated with willingness to participate in surveys. This possibility of response bias suggested a risk of over-reporting vaccination rates, and it could also limit our ability to hear from the vaccine-hesitant. An additional challenge facing researchers is that some of the groups that lagged behind in vaccination rates early on in the vaccination effort, such as communities of color and low-income adults, are also typically underrepresented in survey research. In this study, we report on a multi-level effort to address nonresponse in a year-long monthly tracking study (with more than 16,000 respondents so far). The efforts we report took place in planning the study, and in the course of the field period, as we regularly analyzed nonresponse patterns. The steps taken to address nonresponse included those taken in advance of each wave of data collection (i.e. in sample design), and after the fact (in statistical nonresponse adjustments). This process has been ongoing as vaccine attitudes, and hence the potential interplay of these attitudes with survey participation, changed over the course of the year. In this presentation we will detail the various forms of nonresponse we encountered, and the steps taken along the way to address them. We will demonstrate how this allowed for the study's strong explanatory power, and its usefulness in understanding the diversity of factors underlying vaccine hesitancy, and possible ways to address them.

Measurement Error in Issue Polls: Reality & Perception

What Election Polling Errors Mean for Issue Polling

Scott Keeter, *Pew Research Center*
Nick Hatley, *Pew Research Center*

Most preelection polls in 2020 overstated Joe Biden's lead over Donald Trump in the national vote for president, and in some states incorrectly indicated that Biden would likely win or that the race would be close when it was not. These problems led critics to claim that "polling is broken" raising questions about the trustworthiness of all kinds of polls. To examine the potential consequences of election polling errors for polls about issues, we used weighting to produce two versions of several waves of a large, nationally representative panel. One version balanced the sample to match the national popular vote margin for president (4.4-point Biden advantage), while the other created a biased version with a 12-point Biden advantage (party affiliation among nonvoters was manipulated in the same fashion). Across 48 poll questions about a range of issues, the difference between the biased and unbiased versions was less than one percentage point, varying from 0.5 to 3 points. Errors of this magnitude would not alter any substantive interpretations of opinion about the issues in these polls. We will describe the methodology and results of this simulation. Various postmortems on polling in the 2020 election have converged on the conclusion that Biden supporters were more likely than Trump supporters to cooperate with polls. We will address the question of whether this type of simulation – simply boosting the weight of Trump voters in the poll – fully corrects for the hypothesized bias. Are there types of Trump voters in polls who are different from those who don't participate or is the underrepresentation of Trump voters in polls simply the workings of a stochastic process by which otherwise similar individuals have different propensities of participating in polls. More generally, this exercise examines the standards of accuracy applied to election polling and issue polling.

Why Do Ballot Measure Outcomes Differ From Issue Polling Estimates? Survey Experiments on Status Quo Bias and Issue Preferences

Jonathan Robinson, *Catalist*
John Sides, *Vanderbilt University*
Christopher Warshaw, *The George Washington University*

This paper examines how and why state ballot measure outcomes differ from survey estimates of opinions on the same topics. First, drawing on data on hundreds of ballot initiatives and issue polls, we find significant discrepancies. In particular, the issues most popular in polls tend to underperform in ballot initiatives, relative to issues on which opinions were more evenly divided. This was true for both liberal and conservative issues. Second, using a series of original survey experiments, we investigate the reasons why ballot initiative outcomes differ from issue polls. One reason is status quo bias, or the hypothesis that the public is less likely to support actual changes to policy or law. Another reason is that ballot initiatives contain language and ideas, such as fiscal impact statements, that are not part of typical issue polls. Our survey experiments enable us to identify the causal effect of both priming respondents to think about changing policy from the status quo and providing respondents with actual ballot measure language. The issues in these experiments include background checks for gun purchases, parental notification for abortions, Medicaid expansion, marijuana legalization, government healthcare, paid family leave, and limits on carbon emissions. Our results speak not only to the factors that affect voter behavior in direct democracy, but also to how and why public opinion polls about important issues may misestimate the choices that Americans make when voting on those issues at the ballot box.

The Polarized Pandemic: An Analysis of How Political Polarization Shaped American Public Opinion Towards the COVID-19 Pandemic

Ryan Tully, *Ipsos*
Frances M. Barlas, *Ipsos*
Randall K. Thomas, *Ipsos*

Public opinion researchers have become increasingly interested in the apparent increase in political polarization within the United States. While there is ongoing debate over the extent and nature of this polarization, previous research has focused on its impact on voting behavior and views on longstanding political issues. The current body of polarization research has less often examined how political polarization influences Americans views on, and responses to, major domestic or foreign policy crises. The COVID-19 pandemic began at a contentious period in American politics. In early 2020, the first impeachment of then-president Donald J. Trump was being conducted and the primary process for the 2020 U.S. presidential election was underway. This unique confluence of events primed the American political landscape at the start of what would become arguably the world's most acute public health crisis since the 1918 influenza pandemic. Starting in March 2019, Ipsos and Axios have conducted a weekly public opinion tracking study about COVID-19 using KnowledgePanel®, the largest probability-based online panel in the U.S. These surveys have captured Americans views on a host of topics related to the pandemic, ranging from evaluations of governmental responses to individual behaviors used to mitigate contracting the virus. This paper will utilize this robust dataset to demonstrate the pervasive influence that partisan political polarization had on Americans' public opinion throughout the pandemic on things like approval ratings of the government's handling of the crisis and attitudes towards vaccination. We further show the influence of partisanship on individual behaviors such as social distancing, socializing with friends and family, mask wearing, and even likelihood to get vaccinated. These longitudinal analyses will confirm that American public opinion has been largely crystallized since the start of the pandemic and that real-world events during this time had only minimal effects in influencing public opinion.

Understanding Partisan Selective Exposure in Election Campaigns: A Motivational Approach

Heysung Lee, *University of Wisconsin - Madison*

The presidential election, especially in 2016, was highly divided in part because of the rise of Donald Trump and extensive partisan media coverage on political campaigns. On the basis of a motivational approach (Kunda, 1990; Nir, 2011), this paper investigates how evaluative motivation, need to evaluate , could be related to a relationship between partisanship and partisan selective exposure for the presidential campaign in 2016 by pursuing a directional goal in

information processing. The analysis of ANES 2016 data indicates that need to evaluate is positively related to conservative media use. Also, partisans show a different media consumption pattern based on their level of the need to evaluate; Democrats with a high level of need to evaluate tend to be exposed to liberal media but stay away from conservative news outlets, whereas Republicans with strong evaluative motivations are likely to seek out conservative media and less likely to be exposed to liberal media for election campaigns. The findings provide nuanced explanations for a relationship of cognitive motivation and a widened partisan gap in partisan news consumption.

Political Awareness and the Identity-to-Politics Link in Public Opinion
Phil Jones, *University of Delaware*

Social identities like those rooted in religion, race, gender, and sexuality are frequently associated with distinctive political attitudes. For example, LGBT Americans and people of color hold generally liberal views, while evangelical Protestants more conservative and Republican ones. This suggests a straightforward "identity- to-politics" link in public opinion. But making that link between your identity and political views requires knowledge and understanding of politics, which not all citizens possess to the same extent. Cues from elites might help voters make the connection -- but only attentive voters are likely to receive them, leaving the less-engaged still out of the loop. As a result, this paper argues, the identity-to-politics link is conditional on political awareness. Across multiple datasets, I show that awareness moderates the relationship between social identities and partisanship, ideology, policy views, and vote choice. Among the least engaged, identity is barely tied to political views, and there are only muted differences between groups. As awareness increases, so does the extent to which individuals align their group memberships and political views, and the magnitude of various identity gaps grows. Greater awareness is associated with particularly large increases in the liberalism of Black, Jewish, secular, and LGBT voters, and in the conservatism of evangelical Protestants and military veterans. In contrast, there is no evidence that Asian or Hispanic respondents' views vary with engagement, a finding that fits with previous findings of a weaker effects of identities for these groups. Overall, the results show that the identity-to-politics link for many groups, and the attitudinal differences that are usually attributed to such demographic characteristics, are highly conditional on political awarenes

Impollster Syndrome
Impollster Syndrome
Molly Fisch-Friedman, *Climate Nexus*

It's Not You, It's Me: Effects of Interviewer Characteristics & Behavior
Interviewer and Respondent Skin Tone: Further Evidence for Contextual Race
Angelica Phillips, *University of Nebraska-Lincoln*

Interviewer assessments of respondent race and skin color are often used to understand the process of racialization. However, interviewers systematically vary in their assessments of skin tone, meaning that evaluations of racialization using interviewer-assessed respondent skin tone may be subject to biased estimates. While interviewer assessments of respondent characteristics are well-studied, little research has investigated similarities between interviewer and respondent assessments of respondent skin tone. In this paper we first assess (1) whether interviewers vary in their assessments of respondent skin tone, and (2) whether the degree of similarity between interviewer-assessment and respondent's assessment of their own skin tone is also subject to interviewer variance. Social context such as the respondent's actual skin tone and geographic region where the interview occurred may partially explain interviewer variance in our outcomes. Racialized physical spaces resulting from past colonial encounters may influence how people understand race and consequently how an interviewer approaches racializing a respondent. Therefore, we additionally test whether (3a) interviewers' assessment of respondent skin tone and (3b) the degree of similarity between interviewer-assessed and respondent self-reported skin tone can be explained by the respondent's objective skin tone. We also test whether variance in our two outcomes (4a and 4b, respectively) can be explained by the geographic context of the interview. We use data from the 2018 American National Elections Study (ANES), which includes measures of interviewer-assessed and respondent self-reported respondent skin tone. The 2018 ANES is a mixed-mode face-to-face and web survey, and we restrict our analyses to the post-election face-to-face data (AAPOR RR1=50%, n=1,058) to examine interviewer assessments of respondent skin tone in-person. Preliminary results indicate

that interviewers significantly vary in their assessments of respondent skin tone (p<0.001) and in the degree of similarity between interviewer- and respondent-assessed respondent skin tone (p=0.0009). We conclude with implications for interviewer training.

Interviewer Gender and Age Effects in Face-to-Face Surveys

Neha Sahgal, *Pew Research Center*
Kelsey Jo Starr, *Pew Research Center*
Manolo Corichi Gomez, *Pew Research Center*
Ariana Monique Salazar, *Pew Research Center*
Jonathan Evans, *Pew Research Center*
Adam Wolsky, *Pew Research Center*

Face-to-face surveys remain the preferred mode of conducting survey research in most middle- and low-income countries. This mode has many advantages for researchers – for example, response rates tend to be higher than phone or web surveys, and interviews can be considerably longer, allowing researchers to gather richer data. But face-to-face surveys can also suffer from disadvantages – one of these being that the demographic characteristics of interviewers (such as gender and age) can influence how respondents answer questions. Men, for example, may give more progressive answers on gender issues when speaking with women interviewers. And younger respondents may express more conservative views when interviewed by an older interviewer. In other words, respondents may give responses that they perceive as more socially acceptable or palatable to the demographic characteristics of interviewers. These intervening variables make it imperative for researchers to better understand interviewer effects in face-to-face surveys. This paper uses data from Pew Research Center's 2019 nationally representative survey conducted in India to understand the effects of interviewer age and gender on a variety of interview topics included in the study: national identity, tolerance and gender attitudes. Roughly 400 interviewers carried out face-to-face interviews with about 30,000 respondents across the country. Half of the interviews were conducted by female interviewers, and the age of the interviewers varied between 19 and 52. The analysis uses regression modeling to examine the relationship of interviewer gender and age to survey responses grouped by topic area. The resulting analysis shows which topics are most suspectable to interviewer effects, and the direction of the potential bias that can be expected. The paper also explores the uniqueness of the Indian context where attitudes on gender and identity as especially sensitive and often politically charged

Interviewer Effects on Self-Reported Height

Lukas Olbrich, *Institute for Employment Research*
Yuliya Kosyakova, *Institute for Employment Research, University of Bamberg*
Joseph W. Sakshaug, *Institute for Employment Research, LMU Munich, University of Mannheim*

Surveys serve as an important source of information on key anthropometric characteristics such as body height or weight in the population. Such data are often obtained by directly asking respondents to report those values. Numerous studies have examined measurement errors in this context by comparing reported to measured values. However, little is known on the role of interviewers on the prevalence of irregularities in anthropometric survey data. In this study, we explore such interviewer effects in two ways. First, we use data from NHANES III to evaluate whether differences between reported and measured values are clustered within interviewers. Second, we investigate changes in adult self-reported height over survey waves in two German large-scale panel surveys. Here, we exploit that height should be constant over time for the majority of adult age groups. In both analyses, we use multilevel location-scale models to identify interviewers who enhance reporting errors and interviewers for whom unlikely height changes over waves occur frequently. Our results reveal that interviewers can play a prominent role in differences between reported and measured height values. Furthermore, we find that changes in reported height over survey waves are clustered within interviewers and identify several interviewers who are particularly susceptible to frequent and unlikely height changes. We proceed by providing an analysis of the consequences of height misreporting on regression coefficients where we especially focus on the role of interviewers who reinforce reporting errors and unlikely height changes. Our findings contribute to the literatures on measurement error in anthropometric data, methods of estimating interviewer effects, and methods to identify error-prone interviewers in face-to-face surveys.

Moving Beyond Interviewer and Respondent: Rapport Measurement With Multidimensional Factors

Melike Sarac, *Hacettepe University Institute of Population Studies*

A. Sinan Türkyılmaz, *Hacettepe University Institute of Population Studies Department of Social Research Methodology*

The operationalization of rapport between interviewer and respondent in social surveys is rarely studied, although remarkable efforts were spent in the last decade (Davis et al., 2010; Foucault, 2013; Sun et al., 2021). Apart from nonverbal behaviors of respondents, rapport rating scales pointed by interviewer or respondent allow to make analyses within the scope of quality of data. From quantitative view, this study aims to extend rapport definition especially in sensitive survey settings where such scales are not available. A household survey in Turkey, 2014 Research on Domestic Violence against Women in Turkey (VAW Research), constitutes main data source of the study, along with field staff data. Relevant variables which have potential to broaden rapport concept were selected according to literature as well as data availability. Exploratory factor analysis results suggested that rapport between interviewer and respondent could be formulated with three different factors: 1) timing and frequency of visits (12.5 percent) 2) interviewer characteristics, and socio-demographic matching between interviewer and respondent (10.7 percent) 3) interview settings (9.9 percent). Time of the interview, interview length, and field stage are found to be quite powerful determinants of rapport, in addition to fixed interviewer characteristics such as whether those are student or not and their background. These findings point out certain practical implications in survey process such as interviewer recruitment and interviewer training. In the next stage, analyses will be conducted to find clues about interviews with high rapport.

Starting From Scratch: Research on Questionnaire Development & Design

Measuring Health Care Deferral Among Medicare Beneficiaries During the COVID-19 Pandemic

Elena Moroz, *Mathematica*

Nancy Duda, *Mathematica*

Karen Bogen, *Mathematica*

Charlene Kemmerer, *Mathematica*

Kenneth Lim, *Mathematica*

Efforts to reduce COVID-19 transmission, such as lockdown policies, individuals deciding to avoid public places, providers limiting hours and services, and shifts to virtual care that necessitate access to and comfort with technology can lead to patients deferring care (Czeisler et al. 2020). Capturing delayed care is important for understanding patients' experiences in the context of the pandemic. However, given the novelty of the pandemic, few tools exist for measuring whether patients are delaying health care due to COVID-19. Mathematica completed surveys with over 14,000 Medicare beneficiaries to measure patient experience. This work supports the evaluation of Comprehensive Primary Care Plus, an initiative of the Centers for Medicare & Medicaid Services designed to improve primary care. Between November 2020 and January 2021, we developed two survey questions to measure if beneficiaries delayed or avoided getting medical care from their primary care doctor's office due to COVID-19. Mathematica tested these questions in three pre-test rounds with cognitive testing and full survey debriefs. In this presentation, we discuss the pre-test process, findings, and limitations, and the final wording for the questions. Our pre-tests revealed that respondents had a multifaceted understanding of this concept and their perspectives differed based on the type of care described and whether the delay in care was due to patient choice or health care provider limitations.

Are "Nones" a Survey Artifact? Evaluating an Explanation of the Increasing Percentage of Non-Religious Americans

Philip Brenner, *University of Massachusetts - Boston*

Tracy L. Reed, *University of Massachusetts Boston*

The percentage of Americans reporting no religious affiliation on surveys has increased over the past five decades, growing from around 5 percent in the 1970s to about 30 percent today. Research has argued that some of this growth is caused by poor reliability of answers to filter questions. A survey respondent with a nominal or weak religious affiliation may select "no" to religious affiliation filter question ("Do you currently have a religious affiliation?") before a follow-up question ("What is your current religious affiliation?") is asked that includes a list of religious affiliations as response options. In contrast, the respondent would validly choose their religious affiliation if these two questions were essentially merged: including "none" as an option in a list of religious affiliations included with the question "What is your religious

affiliation?" This paper reports on the findings of a question order experiment testing question reliability. We ask two religious affiliation questions of all respondents: (1) a filter-type question that asks if the respondent has a current religious affiliation (yes or no), and (2) a question asking for current affiliation that presents a list of affiliations including "nothing in particular" as a possible response. The order of these two questions is randomly varied across respondents. We hypothesize a lack of significant question order differences for the filter question and affiliation question with a full set of affiliations as response options. This finding would contradict the extant research and fail to support presuppositions that a survey artifact is bolstering the number of non-religious Americans. Data collection is currently underway using an online nonprobability panel and will be completed in December 2021

Not All Decisions Are Equal: The Need for Continued Cognitive Testing in Decision Quality Research

Carol Cosenza, *University of Massachusetts - Boston*

Shared decision making is considered a cornerstone of good medical care. In order to measure decision quality, survey questions focus on the key components of shared decision making -   defining the decision, e nsuring that patients understand the risks and benefits of each treatment option, and finding out if the medical decision considers patients' values and preferences.   In some medical situations, like surgical interventions, the decision point is clear.   For example, when considering the quality of a decision about whether or not to have hip replacement surgery, we have created and tested questions focusing on the pros and cons of the options (having or not having surgery), what knowledge the patient has about the surgery (for example, risks associated with surgery and time needed for recovery), and what is important to the patient (in relation to hip problems and surgical intervention).   A recent project examined several medical decisions that were not as clear-cut as surgery, including managing chronic conditions, such as depression and ADHD, and asking about decisions related to screening tests, such as mammograms and colorectal cancer screening.   We cognitively tested questions with 10 respondents asking about 20 different medical decisions they had made.   While the format of the new questions was identical to what we had previously tested for surgical decisions, how some respondents talked about their choices or described the decision they were answering about was unexpected. Once again, we found that learning how people were understanding the questions and how they formed their answers, was essential to finding out whether or not a set of questions was actually measuring what we wanted to measure and whether the answers meant what we assumed they meant.

It's Not as Bad as It Looks: A New Way to Measure Community College Outcomes

Jason Kosakow, *Federal Reserve Bank of Richmond*
Laura Dawson Ullrich, *Federal Reserve Bank of Richmond*

Community colleges play a broad role in the American education system. Unlike four-year institutions that have a clear primary goal of awarding bachelor's degrees, community colleges have a diverse set of objectives. These include, but are not limited to, awarding associate degrees, providing vocational training, providing courses for high school students, and offering continuing education for professionals. Graduation and retention rates collected by the federal government and published via the Integrated Postsecondary Education Data System (IPEDS) may not be a good measure of outcomes at community colleges, especially if the community college has a large cohort of non-first time college students, non-credit students, or dually enrolled high school students. Many community colleges have created internal non-reported metrics beyond those required by IPEDS that they feel better reflect student outcomes. These outcome measures are not uniform, making it hard to compare community colleges across states, regions, and the nation. Due to this inconsistency, it is difficult for researchers, legislators, and prospective students to better understand how community colleges are serving their students and impacting their communities. Our research sets out to create uniform outcome measures for community colleges by focusing on measures that are applicable to community colleges; specifically, if a student received a degree or credential, transferred to a 4-year institution, or continue their enrollment. In addition to credit-seeking students, we report outcome measures for non-credit and dually enrolled high school students. For many schools, the outcome measures that better reflect the goals of community colleges resulted in improved outcome measures, especially when compared to data reported on IPEDS. The initial results focus on a survey pilot of ten community colleges in the Fifth Federal Reserve District, with an expectation to expand District-wide in 2022.

Developing Items on Health-related Social Needs—Going Beyond Screening

Charlene Kemmerer, *Mathematica*
Elena Moroz, *Mathematica*
Kenneth Lim, *Mathematica*
Karen Bogen, *Mathematica*
Nancy Duda, *Mathematica*

Social determinants of health (SDOH)—environmental and social conditions that can impact health, such as housing and transportation—have emerged as important considerations in primary care (Hughes 2016). There is a growing trend among primary care practices to screen for health-related social needs (HRSN), but some practices are starting to go further, addressing those needs by, for example, connecting patients to community resources. In this methodological brief, we share our experience developing questionnaire items to capture these emerging primary care practice activities. Comprehensive Primary Care Plus (CPC+) is the largest primary care transformation model in the United States, sponsored by the Centers for Medicare & Medicaid Services, and some participating practices are required to screen patients for HRSN, connect patients in need to community resources, and follow up on those connections. This methodological brief will describe how we developed questionnaire items for a CPC+ Medicare beneficiary survey to measure the social needs of respondents and capture whether they experienced any of these activities through their primary care practice. We will describe the challenges of integrating this sensitive content into an already long questionnaire, engaging a range of beneficiaries and recruiting pretest participants who have been economically/socially marginalized, evolving the questions through three rounds of pre-testing, and developing a final set of questions that we used to collect data from more than 14,000 respondents in fall 2021. As CPC+ is leading primary care transformation, incorporating SDOH into primary care—and not just screening for HRSN—might represent the future. Our brief on question development could inform other patient surveys, as more primary care practices begin to address HRSN beyond screening. References Hughes, Lauren S. "Social Determinants of Health and Primary Care: Intentionality Is Key to the Data We Collect and the Interventions We Pursue." JABFM , vol. 29, no. 3, May 2016. Link

Comparing Probability & Nonprobability Samples

Evaluating the Consistency and Accuracy of Probability-Based Web-Panels and Online Opt-in Sample Sources

Andrew Mercer, *Pew Research Center*
Arnold Lau, *Pew Research Center*
Nick Hatley, *Pew Research Center*

Numerous studies have compared the accuracy of benchmark estimates from different online, opt-in sample sources, often comparing them to a estimates from a single sample obtained from a probability-based online panel. We were unable to find any published studies in which the same kinds of benchmark comparisons were made across multiple samples from different probability-based panels. Although most probability-based online panels currently rely on address-based sampling (ABS) methods for recruitment, other methodological differences in recruitment protocols, panel maintenance, sampling, weighting and other survey operations could result in important differences in sample composition. It remains unclear how well findings based on one ABS panel will generalize to others. Here, we present the results of a Pew Research Center Study in which the same questionnaire was administered in English and Spanish to six large (n»5,000), online samples of U.S. adults obtained from three different probability-based panels and three different online, opt-in sample providers. For both full-sample estimates and key demographic subgroups, we evaluate the consistency of attitudinal estimates across samples and compare their accuracy on a series of measures for which gold-standard, benchmark values are available. We review the study's broader implications for the administration of future online surveys and discuss potential avenues for improving the accuracy of estimates from ABS and opt-in sample sources alike.

Comparison of Probability vs. Nonprobability Designs for Questionnaire Evaluation

Reanne Townsend, *Westat*
David Cantor, *Westat*
Darby Steiger, *Westat*

Many studies that evaluate survey instruments use data from nonprobability sources, assuming that conclusions are generalizable to probability designs. Nonprobability samples have been used to investigate a range of topics in questionnaire design, from the relationship between response latency and error in phone surveys (Draisma & Dijkstra 2004), to the use of progress indicators in web surveys (Yan, Conrad, Tourangeau & Couper 2011). There is a growing body of research that finds substantive point estimates differ between nonprobability and probability sample designs (Callegaro et al 2014, Yeager et al 2011). However, questions remain about whether other factors such as data quality measures and mode effects also differ significantly between these two designs. In this study, we examine the effect of sample design on results from quantitative questionnaire evaluation methods. Data come from a Bureau of Justice Statistics study testing alternative measures of rape and sexual assault to support the redesign of the National Crime Victimization Survey. The study includes a probability design (n=5,777) using address based sampling and random digit dialing of adult women, as well as a nonprobability design using a convenience sample of 2,116 female volunteers recruited from Craigslist. In both the probability and nonprobability designs, approximately half of the interviews were randomly assigned to be administered in-person via Audio Computer Assisted Self-Interview (ACASI), and half administered by an interviewer using Computer Assisted Telephone Interviewing (CATI). Questionnaire evaluation measures such as rates of item missing data, survey timing, test-retest reliability, respondent debriefing data, and behavior coding results will be compared between the probability and nonprobability designs. Respondent characteristics such as demographics and victimization rates will also be compared across designs.

Track Me but Not Really: Tracking Undercoverage in Metered Data Collection

Oriol Bosch, *The London School of Economics and Political Science*
Jouni Kuha, *The London School of Economics and Political Science*
Patrick Sturgis, *The London School of Economics and Political Science*

Metered data, also called "web-tracking data", has been proposed as a potentially useful way of measuring online behaviours, since it allows observation of web browsing unobtrusively and without relying on fallible self-reports. Metered data is generally collected from a sample of respondents who willingly install or configure, into their devices, tracking technologies that track digital traces left when people go online (e.g. URLs visited). To track the complete online behaviour of an individual, tracking technologies must be installed and track all devices, browsers/apps and/or networks (from now on targets) used to go online. When only a subset of targets is tracked, an incomplete record of online behaviours is observed. This undercoverage can negatively affect metered data quality, producing potentially large biases in population estimates. Although little is currently known about this type of undercoverage, past research indicates that a range of factors can prevent researchers from tracking all targets that participants use (Bosch and Revilla, 2021), and that a high proportion of individuals participating in metered studies are undercovered (Pew Research Center, 2020). To assess the impact of this type of undercoverage on the quality of metered data estimates, we collected metered survey and paradata in Spain, Portugal and Italy. Combining paradata and self-reports of participants' tracked and used targets, we show the prevalence and characteristics of undercoverage. Besides, using metered data from the subsample of fully covered individuals and through the use of simulations, we estimate the extent and mechanisms in which undercoverage biases metered data estimates, both univariate (e.g. means and proportions) and bivariate (e.g. correlation and regression coefficients). These estimates are computed for different levels of (% of participants affected by) device, browser and app undercoverage.

Why Random and Opt-in Online Panels Overestimate Harmful Gambling: A Multi-Survey Study

Patrick Sturgis, *The London School of Economics and Political Science*
Jouni Kuha, *London School of Economics*

Recent general population surveys have produced highly variable estimates of the extent of problem gambling in Great Britain, ranging from as low as 0.4% to as high at 2.7% of adults. In this paper, we assess the extent to which differences in approaches to sampling and measurement between surveys contribute to variability in estimates of problem gambling. We compare estimates of problem gambling using the Problem Gambling Severity Index (PGSI) across eight different surveys conducted at approximately the same time but which use different sampling and measurement strategies. Our findings show that surveys conducted online produce substantially higher estimates of problem gambling compared to in person interview surveys. This is because online surveys, whether using probability or non-probability sampling, over-represent people who are more likely to gamble online and to gamble frequently, relative to

the proportions of these groups in the general population. We show that this happens, at least in part, because attrition from online panels progressively selects for more online and tech-savvy individuals.

History of Hispanic Survey Research in the US

History of Hispanic Survey Research in the U.S.

Ana Gonzalez-Barrera, *Pew Research Center*

This panel will delve into the history of survey research of Hispanics in the U.S., from the very early days of Hispanic research, before the term Hispanic was even coined and used widely in survey research to today when Hispanics are often captured in many nationally-representative samples as a subsample of the U.S. population, but are also often targeted as a special hard-to-reach population for research of all types. Participants in the panel will draw from their own experience working in the field of Hispanic survey research explaining their own involvement in different surveys over the years and the challenges they have faced targeting this population and accommodating for its particular characteristics. Some of the main surveys covered by the presenters work will include the very first representative sample of Mexican descent Hispanics 1977; Census Bureau's decennial census itself; and AAPOR's 2018 Inclusive Voices awardee the National Survey of Latinos.

The Early Years of Hispanic Survey Research

Robert Santos, *US Census Bureau*

Back in the 1970s, something very special occurred at the Universtiy of Michigan Survey Research Center (SRC): the National Chicano Survey -- the first ever national area probability face-to-face sample survey of people of Mexican descent. The survey, itself, was conceived and designed by Chicanos, and it was fielded, processed and analyzed by Chicano researchers. Apart from the PI, a researcher at SRC named Carlos Arce who was awarded a an NIMH grant to conduct the study, most of the principals on the research team, including myself, were Chicano graduate students at the University of Michigan and used this project as a vehicle to learn how to conduct survey research on the Mexican-American population. As pioneers in Hispanic/Latino survey research, we addressed a variety of challenges from sampling to questionnaire construction and translation, interviewer training, fielding, and processing and analyzing the data. This presentation will touch on some of the methodological challenges we faced surveying a Latinx population and discuss lessons learned that continue to be relevant to this day

Measuring Hispanic Racial Identity: Using Alternative Measures of Racial Identity in Surveys of Latinos

Ana Gonzalez-Barrera, *Pew Research Center*
Luis Noe-Bustamante, *Pew Research Center*
Mark Hugo Lopez, *Pew Research Center*

Racial identity among U.S. Hispanics has many layers and many dimensions. Racial identity also may impact experiences with discrimination—in a recent Pew Research survey those who identified their skin color as darker were more likely to experience discrimination than those who identified their skin color as lighter. Traditional survey questions for race, like those done by the Census Bureau, often find that many Hispanics respond "other race" rather than black, white or indigenous of the Americas. For example, in the 2020 decennial census 42% of Hispanics selected "Some other race," 33% selected two or more racial groups, and 20% selected White as their race. But traditional race questions do not always correlate with experiences of discrimination or how Hispanics view their identity. This presentation will compare the responses of Latino adults to a variety of questions about racial identity, highlighting differences among methods and exploring how Latinos differ from other Americans on some measures. Among the alternative measures of racial identity are a measure of how Hispanic survey respondents say people would describe them if they walked past them on the street; self-assessed skin tone using version of the Yadon-Ostfeld skin-color scale, an open-ended question about race identification and direct questions about being Afro-Latino, of indigenous roots, and mestizo or mulatto (i.e. mixed race). All these measures will be compared to how Hispanics respond to traditional racial identity questions.

A History of the Development of the Hispanic Ethnicity Category at the Census Bureau

Anna Sandoval, *U.S. Census Bureau*

This presentation discusses the history of the development of the Hispanic ethnicity category at the United States Census Bureau. The Hispanic or Latino population, which includes people of any race, was 62.1 million in 2020. The Hispanic or Latino population grew 23%, while the population that was not of Hispanic or Latino origin grew 4.3% since 2010. The first major attempt by the United States Census to estimate the Hispanic population was during the 1970 Census. Since then, evaluations and changes to the data collection of race and ethnicity have been made every decade. The presentation details this history using Census documents including census forms and reports to provide a comprehensive picture of the research and process to develop and improve the Hispanic category. In addition, the author will discuss her own experience researching Hispanics during her career in Hispanic research in the U.S., from academia, to the private sector and at the U.S. Census Bureau where she is now part of the Integrated Partnerships and Communications Program, which communicates the importance of participating in the 2020 Census to the entire population. Keywords: Hispanic research, Latino research, Census Bureau, history, Ethnicity, Hispanic, Latino/

The National Survey of Latinos and Its Contributions to the Study of Latinos in the U.S.
Ana Gonzalez-Barrera, *Pew Research Center*

The Hispanic population in the U.S. has increased dramatically in the last four decades, growing from an estimated 14.5 million in 1980 to 62.1 million in 2020. Today, U.S. Hispanics make up 19% of the total population in the U.S and it is the largest minority group in the U.S. As the Hispanic population in the U.S. became the largest minority group sometime in the early 2000s, it evidenced the need for a public opinion research instrument that provided insights into their attitudes and opinions about their changing country and their place in it. The National Survey of Latinos is a nationally representative survey of Latino adults conducted annually by the Pew Research Center. It was first fielded in 2002 with the express goal of comprehensively explore the attitudes and experiences of Latinos in a wide variety of topics. The survey was designed to capture the diversity of the Latino population and allowing comparisons among key Latino subgroups in addition to describing Latinos overall. The NSL explores the attitudes and opinions of the nation's fast growing Latino population on topics ranging from identity to politics to immigration policy to education to religion and health care, among others. In 2022, the survey will turn 20 years and continues to be one of the leading instruments for the study Latinos and their attitudes. This presentation will make a recount of some of the main findings that the survey has had in the last two decades, will remark on the significance of having such an instrument dedicated to this specific population and will also discuss the methodological design it has had and some of the challenges it has overcome.

Discoveries from the Inaugural Research of the Gallup Center on Black Voices
Discoveries From the Inaugural Research of the Gallup Center on Black Voices
Lydia Saad, *Gallup*
Jenny Marlar, *Gallup*
Camille Lloyd, *Gallup*
Whitney DuPree, *Gallup*
Megan Brenan, *Gallup*

This session is built around four papers that review public opinion and methodological insights drawn from Gallup surveys conducted since the founding of the Gallup Center on Black Voices in August 2020. This includes three papers reviewing public opinion attitudes – one reviewing Black Americans' attitudes about policing and the criminal justice system; one focused on Black employees' workplace experiences with discrimination and their attitudes about workplace equity and inclusion more broadly; and one focused on U.S. public reaction to policy proposals for officially acknowledging and making reparations for the harms done by slavery and racial discrimination. All three papers are based on web surveys conducted at least partially using the probability-based Gallup Panel. One survey, the fall 2020 Workplace survey, also utilized a non-probability panel to supplement the sample and try to improve coverage of certain under-represented demographic groups within the Black population. The methodology involved in fielding,

weighting and combining the two samples will be discussed in the fourth paper, which will also report on the quality of the resulting sample and how the accuracy of the data resulting from the two samples, individually and combined, was assessed. The moderator's discussion of the papers will focus on what is new in these findings, what connections can be made across the three topical papers, and how the methodological innovation of the blended sample supports these discoveries.

Understanding U.S. Black Employees' Perceptions of Equity and Inclusion Where They Work

Lydia Saad, *Gallup*
Whitney DuPree, *Gallup*
Emily Wetherell, *Gallup*
Ellyn Maese, *Gallup*

This paper reviews the major findings on workplace equity and inclusion from two Gallup surveys focused on employee experiences in the workplace. Both surveys include large oversamples of Black and Hispanic workers to allow for robust analysis by race and ethnicity. The surveys were conducted in November 2020 and June 2021 by web using the Gallup Panel, a nationally representative probability-based sampling frame. The paper first reports on the workplace circumstances and experiences of Black, White, Hispanic and multi-racial workers, both overall and, as sample size permits, by gender and age within each racial/ethnic group. The items reviewed include their type of employer and job position, racial diversity in their workplace (including in workplace leadership), and their experiences with career development, advancement and discrimination. The paper then reviews the relationship between these experiences and workers' attitudes on numerous aspects of their workplace and career opportunities, controlling for age, gender, education and type of job position. At a time when many U.S. employers seek to better understand their Black employees' work life and create an inclusive culture where all employees feel welcome and respected, the paper discusses survey findings on the extent to which employees have participated in training sessions or workplace townhalls on the subject of diversity, equity, inclusion or race. Additionally, it reviews how comfortable workers in managerial roles say they are about having meaningful conversations about race and equity with those they supervise.

Black Americans' Perceptions of Criminal Justice and Personal Safety After George Floyd's Murder

Megan Thee-Brenan, *Gallup*
Justin McCarthy, *Gallup*

The murder of George Floyd at the hands of Minneapolis police in 2020 reignited a national dialogue about racism, law enforcement and police reform that continues today. As such, Black Americans' experience with the legal and criminal justice systems in the U.S. is one of the key research pillars of the Gallup Center on Black Voices. The Center leverages Gallup's historical trends as well as new insights from recent and continuing Gallup Panel surveys. This paper will analyze trended data on a variety of survey questions on the subject and examine how views vary across racial groups. These include trust in law enforcement, public expectations for police, evaluations of interactions with police, support for various policing reforms, and bias and procedural fairness in the criminal justice system. Additionally, the paper will address differences in racial groups' sense of personal safety in their communities and in society in general

U.S. Attitudes Toward Efforts to Repair Effects of Racial Discrimination and Slavery

Jeffrey Jones, *Gallup*
Camille Lloyd, *Gallup*

This paper will report the results of a Gallup Center for Black Voices Survey that measures U.S. adults' attitudes toward official efforts that acknowledge and repair the harms of slavery and racial discrimination. The survey includes a series of question wording experiments to gauge agreement on the impact of "racial discrimination" versus "the history of slavery" on Black Americans today and whether the government should "make amends," "repair the impact" or "make reparations" to Black Americans. Additionally, the survey measures support for a variety of programs to help improve the situation of Black Americans, who should benefit from such programs and whether lessons on racism should be taught in the U.S.

Combining Probability and Non-Probability Panel Samples to Improve Representation of Racial and Ethnic Minorities

Jenny Marlar, *Gallup*
Emily Wetherell, *Gallup*
Ying Han, *Gallup*
Manas Chattopadhyay, *Gallup*
Dato Tsabutashvili, *Gallup*

The Gallup Center on Black Voices (GCBV) is committed to studying the lives of Black Americans. It is critical to the center's objectives that the research represents and gives a voice to all Americans and not just those who are easiest to reach using traditional survey methods. GCBV studies utilize the probability-based Gallup Panel as a primary source of data collection. However, the Panel alone does not have enough members to fully support the required sample sizes. Further, the research team recognized possible under-representation of certain demographic groups within the Black population, such as younger individuals and those with lower education levels. Therefore, the team turned to supplementing the sample with non-probability sample as a possible solution. This presentation will examine results of a GCBV survey conducted in fall 2020. Approximately 7,000 responses were collected via the Gallup Panel and nearly 8,000 from the Dynata non-probability panel, including approximately 7,500 completes from Black and Hispanic respondents. The survey focused on workplace issues and included several questions pulled from Federal statistical systems. The federal statistical system questions serve as a gold-standard benchmarks to evaluate the accuracy of the GCBV estimates. After data collection, Gallup weighted both samples using post-stratification weighting and calibrated the non-probability sample to the probability sample using propensity score matching. The presentation will: 1) describe the data collection methodology and techniques used for weighting and calibration, 2) analyze the demographic composition of the samples, and 3) will compare results to the gold standard benchmarks. More specifically, we explore how the probability and non-probability samples performed individually and combined. We consider if and how the weighting approaches reduced bias. We also explore whether the addition of the non-probability sample helped to improve the inclusion of typically hard to reach demographic subgroups.

Trust & Attitudes Toward Privacy & Confidentiality

The Privacy Paradox and Secondary Disclosure in Government Data Collection

Aleia Clark Fobia, *US Census Bureau*
Casey Eggleston, *US Census Bureau*
Jennifer Hunter Childs, *US Census Bureau*
Kelly Mathews, *US Census Bureau*
Julia Coombs, *US Census Bureau*

The privacy paradox refers to research that demonstrates discrepancies between privacy attitudes and actual behaviors. While respondents report in surveys that privacy is important, they still disclose information. However, critics of the privacy paradox argue that the role of secondary disclosure of information –the release of information to other marketers—has been missed when assessing the discrepancy between attitudes and behaviors.   Much of this literature on the privacy paradox and secondary disclosure addresses the context of social media and online consumer behavior. The context of government data collection and use provides a different focal point by which to understand the potential privacy paradox. In this context, the use of administrative records for statistical purposes is akin to secondary disclosure of information as conceptualized in market research. While previous research has shown that there is some support for the privacy paradox in government data collection, the concept of secondary disclosure has not been adequately explored.   In this presentation, we analyze data from a large national survey of respondent privacy concerns to investigate the privacy paradox and the role of administrative records as a form of secondary data disclosure. The survey data include measures of privacy concerns, including opinions about the use of administrative records, and concerns about individual data items. Privacy-related behaviors include actions that might be taken to protect privacy (e.g., signing up for the national Do Not Call registry, changing internet browser settings, etc.). We first investigate the privacy paradox by exploring how well self-reported privacy-related behaviors are predicted by respondent privacy concerns. Next, we explore the relationship between opinions on the use of administrative records (as a proxy for concerns about secondary disclosure) and privacy-related behaviors.

A New Approach to Study Trust in Media: Moral Values and Attitudes toward News

Jennifer Benz, *NORC at the University of Chicago*
David Sterrett, *NORC at the University of Chicago*
Mariana Hernandez, *NORC at the University of Chicago*
Tom Rosenstiel, *NORC at the University of Chicago*

With long-term declines in trust in news, there has been a great deal of research examining how factors such as objectivity, transparency, and balance relate to people's news attitudes. However, there has been less focus on how Americans' deeply held values might impact their views toward the media. This study from The Media Insight Project, a collaboration between the American Press Institute and The AP-NORC Center, tests whether Moral Foundations Theory (MFT) could provide insights into trust in news. The theory identifies five innate moral values which impact how a person sees and interacts with the world. This study features two nationally representative surveys. The first survey was conducted in October/November 2019 with 2,124 adults and featured interviews from both the AmeriSpeak Panel®, NORC's probability-based panel, and an online non-probability panel. The survey results show that people's moral values are closely tied to their news behaviors and attitudes. The second survey was conducted in August 2020 and re-interviewed 1,155 AmeriSpeak panelists who completed the first survey. The second survey featured an experiment to test whether incorporating additional moral values into three short news stories would change people's attitudes toward those articles. For each story, respondents were randomly assigned to see either the standard version of the article or a revised version of the article designed to appeal to additional moral values. The results show that broadening story framing and headlines to include more moral values increases aggregate audience interest and trust without losing the current audience. We will discuss these and other findings from the project that highlight how moral values impact trust in news.

Cybersecurity Culture in Small and Medium-Sized Businesses

Brian Evans, *Abt Associates Inc*
Grant Gibson, *CIBR Ready*
Ji Sun Lee, *101 Research*
Allison Smith, *101 Research*

Businesses regularly experience cyberattacks, revealing weaknesses within these organizations. Understanding what cybersecurity practices organizations deploy, especially in small to medium-sized companies, is key to knowing how best to defend against these attacks. However, what's not well documented is how well employees of small to medium-sized companies understand and employ cybersecurity best practices. CIBR Ready, in conjunction with 101 Research LLC, conducted a survey of professionals in small to medium-sized companies to better understand respondents' perspectives on their organization's cybersecurity culture, practices, and barriers. The results of this survey highlight areas for improvement in cybersecurity policies and practices that should be considered by small and medium-sized businesses. The survey asked about key topical areas, including overall cybersecurity knowledge and awareness, cybersecurity training, password use and requirements, and potential barriers to cybersecurity. This study suggests there are significant areas for improvement in the awareness and training of cybersecurity practices in small and medium-sized businesses. Results are analyzed by the entire respondent population, those who work in IT jobs versus those that do not, managers versus non-managers, and company size. In this paper we identify specific areas that small to medium-sized companies should address to improve their overall cybersecurity. Lastly, we explore the potential a study like this could have if administered with a different design. Although it provides key findings relevant to understanding cybersecurity at small to medium-sized businesses, this study's use of a non-probability sample limits its generalizability. We discuss challenges in studying this population and potential solutions that will advance high-quality research in this area.

Unpacking Trust: Understanding How General Measures of Trust in Government are Related to Component Measures in Specific Policy Areas

Naila Rahman, *Princeton University*
Rachael Johnson, *Princeton University*
Brian Lee, *Princeton University*
Osama Safeer, *Princeton University*

Edward Freeland, *Princeton University*

Surveys have indicated for decades that Americans' trust in all levels of government has been declining. Since 1960, Americans' trust in the federal government to handle domestic problems has fallen from roughly 75 % to about 20% (Pew Research Center, 2021). However, the questions used to poll Americans' trust in government have focused primarily on how much individuals broadly trust the government. This survey seeks to provide a deeper understanding of Americans' trust in government by decomposing trust into four distinct components: (1) appropriateness of involvement; (2) transparency of what the government does; (3) expertise or competence; (4) reliability of performance. The purpose of the survey is to provide a more in-depth and nuanced understanding of which dimensions of trust are declining and how they are inter-related.   We examine these components across five policy areas: bank regulation, immigration control, environmental protection, COVID-19 response, and health insurance accessibility. Preliminary results indicate that perceptions of government performance or reliability are more strongly related to measures of overall trust than appropriateness of involvement.

Can't Get Enough Trust in Government -- What Can Survey Data Tell Us?
Parker Quinn, *Polco*
Erin Caldwell, *Polco*

While levels of federal government trust are at an all time low, trust at the local level has remained steady and is slightly improving in a number of communities across the nation. This may be due to the fact that the pandemic and its myriad policies made residents more aware of the distinction between the levels of government -- federal, state and local. In this presentation, researchers from the National Research Center at Polco will present our survey data from more than 850 communities over the last 20 years to look at trends in local government trust and help identify the factors of community livability and government service provision that influence trust. We also will show disparities in trust by resident socio-demographics, community characteristics and other public data on government performance.. The dataset analyzed is based on the National Community SurveyTM, a templatized benchmark that survey local governments use across the nation to measure quality of life, community livability, local government performance and civic participation. The database includes the surveys from more than 850 probability samples and 100+ non-probability paired samples

Designing Questions to Increase Data Quality
Effect of Inclusion of Test Questions That Precede Survey Questions on Respondents' Answers
Christopher Antoun, *University of Maryland*
Brian Kim, *University of Maryland*

Academic survey researchers commonly embed knowledge and test questions (measuring political interest, numeracy, etc.) into survey questionnaires. However, little attention has been paid to whether asking these questions might inadvertently influence answers to subsequent survey questions. To address this issue, we conducted a question-order experiment in a web survey of 975 doctoral students at 10 institutions in the U.S. conducted from April-June 2021. We focus on three questions from Lawson's Test of Scientific Reasoning (LTSR). This test was placed either immediately before or immediately after two survey questions that asked respondents to rate their overall performance as a PhD student and the performance of their advisor. Our analysis focused on whether respondents reported lower/higher performance ratings when they performed poorly/well on the preceding test. We found no differences between the experimental groups in how they answered the survey questions: average self ratings were 2.86 for the rating-then-test group and 2.75 for the test-then-rating group ($p=.06$), and average advisor ratings were 2.91 and 2.83 respectively ($p=.31$). Correlations between the test scores and self ratings were -.07 (95% CI -.16 to .03) for the rating-then-test group and -.05 (95% CI -.14 to .05) for the test-then-rating group, and correlations between the test scores and advisor ratings were .08 (95% CI -.02 to .17) and -.03 (95% CI -.12 to .07) respectively. A possible reason for these null findings is that respondents were not aware of their test results and so did not use this information to form subsequent judgements. A future study could examine whether bias does emerge when it is the case that respondents are aware of their test scores

Effect of Contextual Information on Survey Responses

Erik Green, *HarrisX*
Jason Sclar, *HarrisX*
Erik Green, *HarrisX*
Katherine Bradshaw, *HarrisX*
Dritan Nesho, *HarrisX*

Online surveys are an effective way to collect data about a market. One persistent source of bias comes from the way the question is delivered in the survey, specifically, it is unclear as to how much context or background information should be provided to the respondent before they see and answer the question of interest. Providing too much context or misleading information may influence the survey taker's decision, while providing too little may obscure the intent of the question and create noise. According to Tourangeau et al. (2000) people go through a four-step process when answering survey questions: comprehension, retrieval, judgement, and response. The decision by the researcher to provide context or not will directly impact comprehension and retrieval which will lead to changes in judgement and ultimately response. Presenting the information simulates a high information environment how respondents will react as the question is more discussed while declining to present the information creates a more accurate snapshot of what is in the respondent's head at the time of interview. The decision to include information is not value neutral and is imperative that survey designers understand the implications of their decision in order to collect the best data for the purpose of their study. Our study discusses the effect of providing pre-question information on survey takers' responses by presenting results across a range of issues, spanning both commercial and political topics, that vary in both salience and respondent awareness. Through a series of large survey experiments, we show the range of effects and provide valuable information for helping decide the appropriate amount of context.

Controlling for Bias in Survey Input Methods

Jason Sclar, *HarrisX*
Erik Green, *HarrisX*
Woojae Chung, *HarrisX*
Katherine Bradshaw, *HarrisX*
Dritan Nesho, *HarrisX*

Researchers have various methods of collecting data in online surveys, and deciding which method to use in surveys is a common point of concern in survey design. In particular, radio buttons and slider scales are two widely used methodologies that may yield different results. This research builds on Ries (AAPOR 2020), which found that slider scales yielded statistically lower means than radio buttons when the slider was initially positioned at the midpoint of the scale. This research reproduces those findings and extends it across an array of input methods on both computers and cell phones. Using a series of randomized control trials on both large and small screens, we show the differences in responses using all of the most popular input types and the importance of the decision about the respondent inputs their result, with an emphasis on eliminating bias between cell phone and computer based respondents

Calibrating Continuous Ratings with Proxy Respondents in Telephone Surveys

Yi-Chun Chang, *National Taiwan Normal University*
Jui-Chung Allen Li, *Quanthon Corporation*

The respondents who answer and respond to the household telephone survey, without within-household sampling, tends not to be representative of the general population. In this paper, we describe a proxy-respondent method to minimize such sample selection biases. It is a general approach for any continuous rating scales.   Specifically, we ask the respondent to report their own rating. We then ask them to provide information about the number of household members, followed by how many of these household members will give a higher, an equal, or a lower rating than theirs. The ratings of other household members are thus calibrated at either the upper or the lower bound of the original scale and the proxy respondent's reported rating, which can then be estimated using a Tobit regression model. Compared to traditional methods, this approach enables efficient estimation with a given sample size and presumably minimal nonresponse bias.   We report two examples to illustrate the implementation of the method: (a) how pro-independence

versus pro-reunification is the Mayor of Taipei City, and (b) the overall satisfaction with the performance of the Mayor of Yunlin County. We discuss the limitations of our method—e.g., lack of detailed sociodemographic information of other household members not interviewed—and potential solutions for future developments.

Examining Item Nonresponse for Open-ended Question Characteristics in Online Survey Panels
Dorene Asare-Marfo, *Pew Research Center*

The literature on item nonresponse—or question refusal—points to a major conclusion: closed-ended questions result in lower item nonresponse rates than open-ends. This is true regardless of survey mode, such as online versus phone. On Pew Research Center's online American Trends Panel (ATP), item nonresponse rates are 2% on average for closed-ended questions, but average 18% for open-ends. Rates for open-ends also tend to range widely from as low as 3% to over 50%. With such a wide gap, we sought to understand what drives these differential rates. Specifically, we examined the nature of open-ended questions to see if certain features yield higher or lower levels of item nonresponse, and we explored whether differences exist by device type—mobile and desktop. First, we identified close to 100 open-ended questions asked across more than 40 ATP surveys. Then, we computed the item nonresponse rate for each question. We then identified relevant question characteristics – which we also call "dimensions" – that may influence whether people answer an open-ended question: requested answer length , level of cognitive burden and political focus . Generally, we find that for all three dimensions, item nonresponse can vary widely, depending on whether a readily formed opinion exists and can easily be reported. We also find that the cognitive burden associated with a question has the strongest bearing on item nonresponse, regardless of the device used to take the survey. Politically sensitive questions are not susceptible to high item nonresponse compared with apolitical ones, but we do see higher rates for political questions when administered on mobile devices. These results are broadly in line with prior work on this topic and may help researchers better anticipate when the nonresponse rate to open-ends with certain characteristics may be quite high.

WHICH Word stands OUT?
Rachel Horwitz, *U.S. Census Bureau*
Elizabeth May Nichols, *U.S. Census Bureau*

In some instances, survey questions are complex and contain a lot of information. This can lead survey designers to emphasize specific parts of the question to make sure respondents take notice (requesting the "best" description or a time frame of "before" or "after a specific point"). There are many ways to emphasize text. The most commonly used are capitalization, bolding, and underlining. The method of emphasis can depend on the format of the question stem and instructions. For example, if the question stem is bolded then bold cannot be used for emphasis as it would not differentiate. It can also depend on survey mode; underlining is often used in paper questionnaires but not used in web surveys because it can be confused with a link. While survey designers regularly emphasize text in their questions, we do not know the most impactful method to use, nor are we sure it works at all for different types of questions. In this study, we examine the impact of four types of emphasis in a web survey: capitalization with a bold question stem, capitalization and bolding with a nonbold question stem, bolding with a nonbold question stem, and no emphasis with a bold question stem. Respondents were randomly assigned to one of the four groups and received up to seven questions that emphasize different parts of the text, such as what to include/exclude, time frames and to consider the "best" or "main" response. By using follow up probes to measure accuracy, this study will inform the best way to emphasize text, or whether it is necessary at all.

Collecting Nonbinary Gender: Use of Text Entry Prompts for a Third Option
Brian Simmons, *University of Georgia*
Darrell Robinson, M.S., *University of Georgia*
Timothy Collings, *University of Georgia*
Katie Davis, *University of Georgia*

Survey demographics commonly include the respondent's gender; however, there is increasing discussion regarding how to capture nonbinary gender options. The current project tested the use of experimental prompt language associated with an open entry text box when collecting respondents' gender. Survey data were collected via an online Qualtrics survey of Georgia residents. Survey respondents were assigned to one of two experimental conditions. Both

groups were asked the same question, "What is your gender?" Response options differed between the two groups. One group saw the options: "Female," "Male," and "Prefer to self-describe: _ _ _ _ _ _ _ _ _ _;" while the other group saw the options: "Female," "Male," and "_ _ _ _ _ _ _ _ _ _" (without the "Prefer to self-describe" prompt language). Comparison of responses to the gender question will be conducted, particularly the use of the self-description text box. Analyzes will examine any potential differences in the ability to identify nonbinary respondents as well as incidences where the text entry field is used to provide a response not directly related to respondent gender (such as expressing disapproval of a nonbinary option). Data collection started in October 2021 and will conclude by early December, with data analyses anticipated by early spring 2022.5

Let's Try It Another Way: Modern Modes of Data Collection

Can You See My Screen? Testing Multiple Video Platforms in the Collection of Health Care Expenditure Data

Darby Steiger, *Westat*
Jesus Arrue, *Westat*
Lena Centeno, *Westat*
Cindy Good, *Westat*
Rick Dulaney, *Westat*

In October, 2021, Westat staff began testing Computer Assisted Video Interviews (CAVI) as a potential methodology for the Medical Expenditure Panel Survey-Household Component (MEPS-HC). Specifically, we set out to explore the feasibility of offering a video interview for interviews in households in later panel rounds that might otherwise be conducted by phone. In later panel rounds of MEPS, respondents are sometimes wary of continuing to invite the interviewer into their home, and prefer to complete the interview by phone. Benefits of conducting these later interviews by CAVI could include enhanced ability to gain and maintain rapport with respondents and increase respondent attention during the interview. Interviewers could more easily detect problems such as confusion, boredom, discomfort, and distraction through nonverbal cues and the ability to see the surrounding environment than they can over the telephone. Interviewers would be able to more easily convey patience while respondents collect records and smooth awkward moments of silence that may occur while they enter information in comments. Okon and colleagues (2021) found that respondents who preferred live video interviewing to telephone interviewing cited the ability to hear or see, and social connections with the interviewer as reasons for their preference. Respondents who preferred live video over in-person interviewing cited convenience, privacy, and health or safety. This paper will present findings from multiple waves of CAVI testing with mock and actual MEPS respondents using Zoom, Microsoft Teams, and Google Meet, and the challenges of training a field interviewing team on CAVI. Findings and practical recommendations will focus on the technological capabilities, benefits, and limitations of each platform including the use of showcards, interviewer debriefings, respondent feedback, and a review of the scope of content necessary to train a field staff on how to implement CAVI interviewing.

Response Rates and Representativeness in Two Different Types of Text Message Surveys

Kevin Collins, *Analyst Institute*
Dan Castleman, *Clarity Campaign Labs*
Alexander Dawahare, *Survey 160*

Text messaging, or Short Message Service (SMS) communication, is an increasingly common tool in survey researchers' toolkit, and research to date has shown that the surveys that use SMS can be as accurate as phone-only surveys. However, much as there is methodological diversity within phone surveys or online surveys, the same is true for text message surveys. SMS surveys can refer to at least two types of interview methods: text-to-web surveys (T2W), in which respondents are initially contacted by text message but then are provided with a URL to a web survey, with the bulk of the interview self-administered online, and live interviewer SMS surveys, in which the whole of the interview happens back and forth over text message, with a live interviewer coding responses. In this paper, we report the results of two survey experiments to compare these approaches to text message surveys. Respondents, initially contacted by text message, are asked to complete the survey over text message, are asked to complete the survey online, or are given a

choice of modes. One survey covers political topics, while the other asked questions about vaccination against COVID-19. In the latter survey, we also compare text message conditions to phone interviews to the cell-phone component of the phone sample. Across these conditions, we compare (1) survey completion rates, (2) the representativeness of the resulting samples, and (3) the resulting estimates of Presidential approval and vaccination rates from these different approaches.

Speaking or Typing? Comparing Voice and Text Answers to Open Questions on Sensitive Topics in Smartphone Surveys
Jan Karem Höhne, *University of Duisburg-Essen*
Konstantin Gavras, *University of Mannheim*

The increase of smartphone usages in web surveys, coupled with technological developments, provide novel opportunities for measuring respondents' attitudes and opinions. For example, smartphones allow the collection of voice instead of text answers by using the built-in microphone. This may facilitate answering questions with open answer formats and may result in richer information and higher data quality. New advancements in Automatic Speech Recognition (ASR) and text-as-data methods also enable a proper handling of open answers from large-scale web surveys. However, so far, there is almost no research investigating voice and text answers to open questions. In this study, we therefore compare the linguistic and content characteristics of voice and text answers to four open questions on sensitive topics (e.g., deportation of delinquent refugees). For this purpose, we conducted an experiment in a smartphone survey (N ~ 1,000) in Germany in July and August 2021 and randomly assigned respondents to an answer format condition (voice or text). Voice answers were collected using the open-source SurveyVoice (SVoice) tool that resembles the voice function of popular Instant-Messaging Services, such as WhatsApp and WeChat. We initially transcribed voice answers into text using Google's Cloud Speech-to-Text API. In a first step, we compare missing data (i.e., dropouts and item-nonresponse) between the voice and text conditions. We then analyze the length (i.e., number of words), lexical structure (i.e., lexical richness, lexical diversity, and readability), sentiments, and topics of voice and text answers. Our study provides new insights into the linguistic and content characteristics of voice and text answers. Furthermore, it helps to evaluate the usefulness and usability of voice answers for future smartphone surveys.

Demographic and Response Comparisons of Participants Selecting to Use Different Device Types for Survey Completion
Ronna Turner, *University of Arkansas*
Kristen Jozkowski, *Indiana University*
Wen-juo Lo, *University of Arkansas*
Brandon Crawford, *Indiana University*

Smartphones are increasingly being used for survey administration, and experimental studies have investigated potential differences in responses based on mode of administration (e.g., Antoun, Couper & Conrad, 2017; Tourangeau et al., 2017). Although certain types of questions have been discussed as being more challenging for smartphone users, the mode effects across devices appear small in many circumstances. However, there are also studies that have identified significant differences in characteristics such as response rates, response times, and omissions for open-ended items (Vicente et al., 2009; Wells, Bailey, & Link, 2014). As research on the impacts of devices on survey responses continue, it may also be informative to understand who tends to use the different types of devices when completing surveys. In this study, we investigate responses from two web-based surveys administered using two national panels (2020 IPSOS KnowledgePanel [N=919] and 2021 NORC Amerispeak [N=1025]), allowing participants to select the type of device used for survey completion (computer, smartphone, tablet). Characteristics such as demographics and background variables, item omissions, and relationship to survey construct outcomes are compared for participants choosing different types of devices. Results from the IPSOS dataset indicate that smartphone users tend to be younger with lower educational levels than those using a computer. Democrats and participants with no political affiliation were also more likely than Republicans to use smartphones. Item omission levels were also higher for those using a smartphone. However, device used was not a significant predictor of construct outcomes investigated. A comparison of results across datasets will be made along with a discussion of additional characteristics investigated that were not found to be significant.

Respondent-Driven Sampling: Design Considerations & Applications
Using Auxiliary Data to Enhance Surveys & Sampling Frames

Creating Big Data Classifiers for Survey Sampling

David Dutwin, *NORC*
Patrick Coyle, *NORC*
Ipek Bilgen, *NORC*
Ned English, *NORC*

A potentially significant use of Big Data, which has not been explored to date within the survey research literature, is to more effectively target populations for surveys, especially those that are rare or hard to reach. Potentially, Big Data could be used not only to significantly improve the targeting of groups but also to identify groups for which historical sampling techniques have not at all been able to target. This presentation reports on a multi-year effort to develop Big Data classifiers from which to target a wide range of survey populations. We report on a number of testing outcomes of different decision points in the model building process to explore which strategies result in the most effective models. These include choices with regard to a) which Big Data source to use b) which analytic method is most effective, c) how choice in data aggregation from the individual to the household level impacts model efficacy, and d) the impact of differential cut points in classification. To address these questions, we develop classifiers on a range of different individual attributes. We then assess the effectiveness of our final models by applying them to a recent survey effort. Where applicable, we compare our results to the performance of available vendor flags. Our results underscore the substantial utility in the use of Big Data classifiers, with many lessons learned in order to generate classifiers that maximize the critical metrics of accuracy (incidence) and completeness (coverage).

Reevaluating Medicaid Coverage in the Current Population Survey Annual Social and Economic Supplement: Evidence from Linked Enrollment Records

Laryssa Mykyta, *U.S. Census Bureau*
Katherine Keisler-Starkey, *U.S. Census Bureau*

Although research suggests that there has been more stability in insurance coverage and less Medicaid churning since the implementation of the ACA (Vistnes and Cohen 2018; Goldman and Sommers 2020), most of these studies have relied on household survey data for estimates of coverage and transitions in coverage, including Medicaid coverage. Yet, Medicaid underreporting has been well documented in household surveys, including the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) (Davern et al. 2009; Klerman et al 2009; Noon and Fernandez 2013). Fortunately, the redesign of the CPS ASEC questionnaire, together with recent updates to the CPS ASEC processing system and imputation methods, improved the quality of health insurance coverage estimates overall, and increased estimates of Medicaid coverage (Berchick and Jackson 2021; Berchick and Jackson 2019). The updated CPS ASEC processing system also enables researchers to consider monthly level health coverage data collected in the CPS ASEC to explore health insurance dynamics, including transitions to and from Medicaid coverage. In this paper, we leverage Medicaid enrollment records linked to the CPS ASEC to evaluate whether the CPS redesign improved the capture of Medicaid coverage. We estimate logit models predicting concordance between administrative and survey data to determine characteristics associated with matching records of Medicaid coverage. We further use linked Medicaid enrollment-CPS ASEC data to compare transitions in Medicaid coverage in the CPS ASEC with transitions reported in the enrollment data.

Data Fusion for Enhanced Survey Sampling

Mansour Fahimi, *Marketing Systems Group*
David Malarek, *Marketing Systems Group*

With the explosion of information technology in the Digital Age, now surveys stand to benefit from an unprecedented wealth of ancillary data that can be appended to traditional sampling frames to enhance their design and analytical possibilities. The available sources that make such data fusions possible include a myriad of publicly available statistics from the Census and other government entities, as well as a growing number of commercial databases that are licensable from private organizations. Addition of such data can not only allow fairly nuanced targeting and stratification options, but also provide an assortment of "free" covariates that can enrich inferential capabilities of surveys. Moreover, fusion of ancillary data can have huge possibilities for market research applications. Oftentimes, the tidbits companies have about their customers are too scant to support important business decisions. Enhancing

customer databases can produce rich empirical foundations upon which actionable intelligence could be generated to improve profitability – way beyond what basic descriptive statistics could otherwise produce from meager customer data. By geocoding every single delivery point in the US to a unique coordinate, it will be possible to marry all Census geographic delineations to each address, starting from blocks to counties and beyond. This process bridges the existing divide between addresses and the plethora of data from the government sources. On the other hand, rapidly evolving commercial data streams can provide an assortment of data unavailable from public sources. The authors will demystify the process of data fusion and discuss possibilities for sampling frame and database enhancements for survey and market research applications. Specifically, we will describe what ancillary data items are available for fusion for address and telephone sampling frames. Further, results from case studies will be reviewed to illustrate the many practical utilities of this data fusion methodology.

2020 Census Off-Campus Administrative Records Submission Program

Andrew Keller, *U.S. Census Bureau*
Vincent T. Mule, *US Census Bureau*
Sarah Brady Clark, *US Census Bureau*
Andrea Johnson, *US Census Bureau*

In March 2020, because of COVID-19 concerns, many colleges and universities discontinued in-person classes and activities, closed residential facilities, and as a result, students residing on-campus and potentially those residing off-campus completed the spring semester in a virtual learning environment from an alternate location. The closure of in-person learning and on-campus housing by most colleges and universities impacted the Census Bureau's plans for enumerating both on and off-campus students.   For on-campus students the Census Bureau adjusted some of the Group Quarters methodologies and timeframe to support these changes. However, adjusting the enumeration of off-campus housing students was in some ways more challenging. During the summer of 2020, the U.S. Census Bureau contacted colleges and universities to determine if they could provide information about students who were living in off-campus housing units. The schools were contacted to see if they could provide the student's name and date of birth, local off-campus address and an alternative permanent address if available. The Census Bureau then executed two main processing steps after receipt of the off-campus data. First, geographic information was assigned to the local and alternative addresses. Second, the Census Bureau combined this information with other administrative records to enumerate the off-campus households. This presentation discusses the processes and challenges associated with using this data for enumeration.

Advancing Equity Through Vaccination, Modeling, and Experimentation

Masahiko Aida, *Project Applecart*

As the COVID-19 pandemic upends our society and impacts the living of millions of people, it also provided social scientists ample opportunities to apply our tools for the betterment of humanity. Our partner, a grassroots mobilization organization, ran COVID-19 vaccination appointment efforts over the phone. Because the state did not provide vaccine status due to HIPPA concern, the partner was calling in blind, resulting in a very low success rate. Applecart was hired to create a predictive model for targeting in order to improve vaccine appointment efforts, using auxiliary data available such as voter data, Census data, aggregated COVID vaccine administration data, and COVID positivity statistics. The modeling aimed to predict several outcomes, successful appointments, refusals (vaccine hesitancy) and already vaccinated. The predictive model represented the likelihood of one's willingness to book a vaccine appointment (v.s. being vaccine-hesitant or already vaccinated). While the model was to improve vaccination efforts, using the state's budget more efficiently for the public's interest, we had a concern that targeting effort may systematically reduce attempts to reach out to the subgroup of the population, thus making the endeavor less equitable. The second phase of the project focused on exploring and testing different message framing and scripts to improve vaccine appointments among the group with higher vaccine hesitancy. Aided by the insights of public opinion survey and survey experiment-based message testing results, we designed a randomized cluster design seeking the messages to identify the most effective message themes for different subgroups. We will report the result of both research phases, in addition to the estimated improvement on the intervention.

All Politics are Local: Influence of Geography & Race on Political ID
How Neighborhoods Form Political Identity: Evidence from Young Movers

Martin Koenen, *Harvard Kennedy School*
Jacob Brown, *Harvard University*
Enrico Cantoni, *University of Bologna*
Sahil Chinoy, *Harvard University*
Vincent Pons, *Harvard Business School*

This paper asks whether the neighborhood where an individual grows up affects their political identity and behavior in adult life. We combine a nationwide voter file with detailed address histories to track families who relocate across states and counties. Using a movers' design, we then assess the extent to which individuals' propensity to vote and their party affiliation are shaped by their childhood neighborhoods. Specifically, we compare children who have very similar own and parental characteristics and whose parents moved between the same places, but who differ with respect to the exact age at which their parents moved. Our preliminary results show that the political environment of the neighborhoods that children grow up in substantially influences their future party affiliation, even after accounting for the political orientation of their parents. The effect of the location where one grew up is particularly large during adolescent years, in line with the impressionable years hypothesis

Political Attitudes and the Measurement of Racial and Ethnic Identity

Alex Von Hagen-Jamar, *Global Strategy Group*
Rosa Mendoza, *Global Strategy Group*

Common approaches to measuring and weighting on race and ethnicity can obscure political and attitudinal diversity. Drawing from research by academics and practitioners we use Global Strategy Group's unique database of thousands of interviews of registered and likely voters to evaluate the relationship between political attitudes and the measurement of racial and ethnic identity, including multiracial identities. We show that the distribution of political attitudes within and across racial and ethnic groups depends on how identity is measured and used for weighting, with broad implications for the representativeness and accuracy of political polls. We conclude with a discussion of current best practices and future avenues of research.

Racial Differences in Affective Polarization: Survey Evidence of Relative Inter-Partisan Harmony Among Hispanic Americans

Samara Klar, *University of Arizona*
Rongbo Jin, *University of Arizona*
Fabian Neuner, *Arizona State University*
Mark Ramirez, *Arizona State University*
Chris Weber, *University of Arizona*

Affective polarization—that is, personal dislike and distrust between Democrats and Republicans—is shown to arise, at least in part, from a shrinking number of cross-cutting identities that are shared by both Democrats and Republicans. We argue that this phenomenon might be specifically relevant to non-Hispanic white Americans, but much less so to Hispanic Americans who form a politically diverse group with strong social ties that unite them. Across multiple surveys, we indeed find lower animosity toward out-partisans among Hispanics than among non-Hispanic whites. With a large and representative survey of Hispanic and non-Hispanic Arizonans, we attempt to uncover the source of inter-partisan harmony among Hispanics. We find that, among both Hispanics and Whites, a strong partisan identity increases animosity toward out-party members but, among Hispanics, a strong racial identity increases animosity only toward out-partisans from outside one's own ethnic group. We argue that affective polarization might be less prevalent among the non-white groups who make up the fastest-growing proportion of the American electorate.

Minority Candidates, Affirmative Action, and Majority Group Voters' Support for Democratic Norms

Beyza Buyuker, *University of Illinois - Chicago*

Recent developments in many democratic societies have introduced a new question that goes beyond the electability of minority politicians. Would majority group members (especially prejudiced ones) go beyond voting down a minority politician and especially one who advocates for minority rights to supporting restrictions on their civil and political liberties? This question arises because some democratic societies have experienced a decline in the salience of democratic norms (Levitsky and Ziblatt 2018). In such contexts, reactionary elites have encouraged not simply electoral opposition to minority politicians, but restrictions on their civil liberties on the grounds that their minority status and/or their pro-minority agenda makes them illegitimate political representatives of the nation (Rogers and Fandos 2019). A long line of research testifies to the context-dependence of public attitudes about democratic norms (e.g., Marcus et al. 1995; Nelson, Clawson, and Oxley 1997). Recent scholarship shows that majority group members with strong negative outgroup priors, embrace anti-democratic norms (Bartels 2020; Miller and Davis 2021). Taken together, these studies suggest that when democratic norms are weakening, majority group members and especially those with strong negative outgroup priors may go beyond not voting for a minority politician to endorsing anti-democratic penalties against them. We tested our expectations using two identical, pre-registered survey experiments fielded concurrently online in the United States and Switzerland with majority group members (non-Hispanic white Americans and Swiss Germans). Our results suggest support for anti-democratic penalties against the minority politician and the proposer of a policy that benefits minorities increases as respondents' negative attitudes towards immigrants go up both in the U.S and in Switzerland.

Improving Data on Black Public Opinion by Improving Data on Black News Media Consumption

Caroline Egan, *University of Michigan*
Mara Ostfeld, *University of Michigan*
Elisabeth Gerber, *University of Michigan*
Jeffrey Morenoff, *University of Michigan*
Lydia Wileden, *University of Michgan*

Public opinion scholars have long noted the importance of news media in how people form their views on political issues. Yet comprehensive studies of news media consumption patterns have almost exclusively relied on samples that are mostly White, leaving much to learn about the news media sources with the greatest impact on public opinion among African Americans. This racial bias in the measurement of news media consumption patterns has downstream consequences on both the way we structure media exposure questions (i.e. program list techniques in which respondents are asked to indicate their news media sources from pre-populated lists), as well as the factors considered when measuring news media effects (e.g. messengers, content, modes).   In this study, we draw on a survey of about 1,500 African American adults to offer new insights into the most common news media sources relied upon by African Americans. Importantly, we consider both the most common modes relied upon for news media among African Americans, as well as the most common programs. Our findings highlight the importance of including a much broader range of media sources when relying on program list techniques. In addition, our findings draw attention to the growing importance of exploring how relatively new platforms for news media (e.g. instagram), which are particularly common among African American audiences, may be affecting how African Americans form their political views.

The Shy Voter: Evaluating Selection Bias & Nonresponse in Political Polls

Evaluating An Alternative Raking Strategy for A National Household Survey Program

Tzu-Jou Wan, *American Institutes for Research*
Jiashan Cui, *American Institutes for Research*
Rachel Hanson, *American Institutes for Research*
Danielle Battle, *American Institutes for Research*
Michael Jackson, *SSRS*

Public-use datasets such as the American Community Survey (ACS) are often used to create population benchmarks for raking. However, researchers often do not consider how differences, even seemingly subtle, in the measurement of weighting variables can affect the accuracy of the resulting weights. For example, the National Household Education Surveys Program of 2019 (NHES:2019) overestimated the number of kindergarteners and underestimated the number of 4-year-olds in the child population due to the misalignment in reference periods between the survey being raked (i.e.

NHES:2019) and the benchmark survey (i.e. ACS). An alternative raking strategy was developed to improve the precision of population estimates of children in the NHES:2019. The alternative raking strategy stacks the two child surveys in NHES:2019 together for the raking process when developing the weights. This strategy was expected to address the estimation issue by avoiding reliance on variables that are sensitive to the differing reference periods. This study compared the population estimates of children in the NHES between the current and alternative raking strategies for the overall child population and by select demographic characteristics. Additionally, key survey estimates and associated standard errors in the surveys were evaluated between the two raking schemes for statistically significant differences. The results suggest that the alternative raking strategy mitigated the estimation issues in both surveys, with minimal disruptions to the cross-year trends of survey estimates. The study demonstrates the importance of consistently defined populations and weighting variables between the surveys being raked and the benchmark survey, and cautions on the use of demographic characteristics that could lead to extreme weighting adjustments for raking

Respondent Driven Sampling Design Considerations

Ai Rene Ong, *University of Michigan*
Michael R. Elliott, *University of Michigan*
Sunghee Lee, *University of Michigan*

Respondent Driven Sampling (RDS) has been used as a method to sample hard-to-sample populations, leveraging the social networks of the initial respondents, typically selected through convenience sampling, to reach more people from the target population. Respondents are asked to invite their eligible peers to participate in the study, and this process continues until the sample size is reached. Although there have been some general recommendations for RDS best practices (e.g., conducting formative studies, a small number of seed respondents), efforts to study the contributions of these design decisions on the productivity of RDS peer recruitment have been hindered by incomplete reporting of RDS methodology in the literature. This study presents an exploratory analysis of the associations of various RDS design decisions on peer recruitment productivity. The data used is from a survey of researchers who have published an article using RDS or have grants funded for research using RDS from 2009 to 2020. These researchers were sampled from a database that represents a census of RDS researchers. A hundred and twenty-one researchers completed the survey which asked about the design of their RDS data collection. Preliminary results indicate that fielding an RDS survey on the web is associated with better productivity, and this effect is moderated by the type of population the RDS study is targeting. Giving more than one form of instructions for peer recruitment appeared to help with peer recruitment productivity. However, conducting formative research prior to data collection was not associated with peer recruitment productivity

Sample Size Estimation in Respondent-Driven Sampling

Yibo Wang, *University of Michigan*
Michael R. Elliott, *University of Michigan*
Sunghee Lee, *University of Michigan*

Sample size is a necessary feature of survey design, as it defines the level of precision of the resulting statistical inference. In contrast to most sample surveys, respondent-driven sampling (RDS) studies do not start with a sample of a fixed size drawn from a sampling frame; rather, they depend on recruitment chains branching from a small number of "seeds". Although researchers may have a target sample, the peer recruitment of RDS brings uncertainty to the resulting sample size, essentially making it a random variable over which researchers have little control. When peer recruitment is more successful than assumed, the recruitment chains may become too thick, and one may need to truncate the chains at an early stage of data collection to stay on the budget. When less successful, the chains may become short leading to too small a sample size, and this may necessitate an extension of the study with adding more seeds. Thus, there is a need for theoretical guidance and computation algorithms to estimate distributions of sample sizes. This study aims to calculate distributions of sample sizes by modeling individual recruitment through a Bernoulli distribution and decomposing the size recursively, under a reasonable set of parameters including number of seeds and coupons and probabilities of peer recruitment. Given our analytic results, the distribution of the sample size is formulated in the form of an exhaustive enumeration of all possible scenarios of peer recruitment that may be observed in the field. To expedite the computational process, we propose an approximation algorithm which estimates the distribution of the sample size by trimming off unlikely scenarios with negligibly small probabilities of occurrence. Simulation studies and

application to real RDS data sets demonstrate the validity and usefulness of the proposed methodology for planning or designing RDS studies

Better or Worse? Investigating the Impact of Requesting Fewer Service User Referrals in the Respondent Driven Sample for New York's Health Information Exchange Customer Usage & Satisfaction Survey

Holly Hagerty, *NORC at the University Of Chicago*
Vicki Pineau, *NORC at the University of Chicago*
David Gleicher, *NORC at the University of Chicago*
Nicolas Fernandez, *NORC at the University of Chicago*
Rina Dhopeshwarkar, *NORC at the University of Chicago*
Krysta Heaney-Huls, *NORC at the University of Chicago*

The Statewide Health Information Network for New York (SHIN-NY) Customer Usage and Satisfaction Survey assesses service usage and users' experiences with health information exchanges (HIE) services provided by multiple Qualified Entities (QE) in New York State. Organizations are sampled from strata formed by QE and organization type, which can include hospitals, physician practices, and other regulated health care entities (OREs). Each sampled organization is associated with a QE-nominated gatekeeper respondent, who responds to the survey as the organization-level representative. At the end of their survey, the gatekeeper respondent is asked to refer a certain number of knowledgeable respondents from their organization to also complete a shorter version of the survey. We analyze knowledgeable respondent referral rates over four rounds of the SHIN-NY Customer Usage and Satisfaction Survey. In particular, we analyze the effect on knowledgeable respondent referral outcomes of requesting a maximum of five referrals in the fourth round of the survey, compared to requesting a maximum of 15 referrals in the first three rounds of the survey. In the first three rounds of the survey, gatekeeper respondents referred an average of 1-2 knowledgeable respondents per completed survey, with a median referral of 0 and a maximum referral of 15 knowledgeable respondents. We investigate whether lowering the number of maximum number of referrals requested has any impact on the number of knowledgeable respondents referred, the number of referrals completing the survey, and whether the change is correlated with an increase in response rates in the fourth round of the survey.

Qualitative Research with Underrepresented Groups

Understanding rare populations using focus groups: Pew Research Center's 2021 Asian American Focus Group Study

Neil Ruiz, *Pew Research Center*
Sunny Shao, *Pew Research Center*

The U.S. Asian population is diverse. A record 22 million Asian Americans trace their roots to more than 20 countries in East and Southeast Asian and the Indian subcontinent, each with unique languages, histories, cultures, and other characteristics. Asian Americans also make up 7% of the U.S. population and the majority are foreign-born. This presentation will focus on how Pew Research Center designed a major qualitative focus group study to capture the diversity of voices and experiences of the U.S. Asian population by conducting 66 online focus groups of the largest 18 Asian origin groups during the summer and fall of 2021. The focus groups were conducted in English for U.S. born population and in-language with simultaneous translations (both in-language transcripts and English translated transcripts) for the foreign-born Asian population. This presentation will focus on how the focus group study was designed, conducted, and analyzed the focus groups during the COVID-19 pandemic, where it became challenging to do in-person focus groups. Additionally, online focus groups enabled the research team to recruit from a nationwide sample to reach audiences from across the U.S., including rural areas, and engage with key audiences in their homes or residences. These online focus groups enabled the research team to develop a comprehensive understanding of the issues surrounding immigration, economic mobility, identity, representation, and discrimination among Asians living in the United States. The qualitative nature allows us to capture rich data, such as non-verbal cues and changes in tone, as participants respond to topics raised during discussions. This presentation investigates the value of focus group research in studying rare populations such as Asian Americans, and how it can better inform researchers on survey questionnaire development such as multifaceted racial/ethnic identity, as well as being reported as qualitative research findings on their own

Results from a Cognitive Interview Study Evaluating Questions on Adverse Childhood Experiences for the Youth Risk Behavior Surveillance System
Meredith Massey, *NCHS*

The Youth Risk Behavior Surveillance System (YRBSS) monitors high priority health-risk behaviors among high school students. The system includes a national survey that is conducted by the Centers for Disease Control and Prevention (CDC), as well as state and local surveys that are coordinated by state and local education and public health agencies in collaboration with other partners. Adverse Childhood Experiences, or ACEs, are potentially traumatic events that occur in childhood (0-17 years) such as neglect, experiencing or witnessing violence in the home, and other aspects of the child's environment that may undermine their sense of safety and stability (e.g., household substance use, instability due to parental separation or incarceration). The Collaborating Center for Questionnaire Design and Evaluation Research at the National Center for Health Statistics conducted a cognitive interviewing study in order to examine a set of questions on childhood experiences for use in an optional module of the Youth Risk Surveillance Survey (Youth ages 12-17). Cognitive interviews were conducted with 30 youth. This presentation will explore the results of this cognitive interview study and discuss methodological considerations for asking sensitive questions to youth. 5

Who Isn't Counted? Evaluating an Age of Disability Onset Question Aiming to Capture Prevalence of Adults with Intellectual and Developmental Disabilities
Ann MacFadyen, *National Center for Health Statistics*

There are not adequate national-level data capturing the prevalence of adults with Intellectual or Developmental Disabilities (IDD). Existing population health surveys with adults do not sufficiently assess defining aspects of IDD such as independent living, learning difficulties, and the age of onset of those difficulties. To address these concerns, the Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER) is conducting a question evaluation study assessing questions related to key identifiers of IDD. The evaluation employs a mixed methods approach, using qualitative data from cognitive interviews as well as quantitative data from the National Center for Health Statistics' Research and Development Survey (RANDS) – which uses a probability-sampled commercial survey panel. This approach allows us to both examine the constructs captured by the evaluated questions and to inform our understanding of measurement error. This presentation will focus on how an age of disability onset question performed in relation to identifying people with IDD on social surveys. The onset question achieves this by determining that the age of a person's disability onset was before 22, meaning onset occurred during that person's development. Following several 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. Findings from both cognitive interviewing and RANDS show that many people with IDD were not counted by the age of onset question. This presentation examines instances of false negatives to this question, meaning someone with IDD answered 'No,' – leading to a potential undercount of those with IDD.

Using Equity-Focused Strategies to Promote Participant Engagement
Natalie Morrissey, *Fors Marsh Group*
Alicia Garza, *Fors Marsh Group LLC*
Maria Porta, *Fors Marsh Group LLC*

Equity-focused research illuminates the voices, experiences, and perspectives of historically underrepresented and socially undervalued populations. This approach requires researchers to show empathy and respect for participants' worldviews, experiences, and knowledge while acknowledging and suspending any of their own biases and assumptions. The benefits of this research approach are improving participant-researcher interactions, developing stronger relationships with underrepresented populations, and increasing access for research participation. In this paper, we outline strategies researchers can follow to promote equity and create solutions to participant barriers when planning, executing, and concluding user experience, exploratory, and market research. We highlight specific examples of how we have leveraged equity-focused strategies to promote meaningful engagement with Spanish dominant

participants and discuss the benefits of the overall approach yielding deep data, accurately representing the diverse worldviews of underrepresented populations

Into the Wine, Not the Label: Explorations in Asking About SOGI
Results of Cognitive Interview Study of a 2-Step Gender Identity Measure
Kristen Miller, *National Center for Health Statistics*

This presentation will describe a study, conducted by the Collaborating Center for Question Design and Evaluation Research (CCQDER) at the National Center for Health Statistics, to investigate the performance of a gender identity measure for federal surveys. Specifically, the study examines a 2-step measure, whereby respondents are first asked sex assigned at birth followed by current gender identity. The primary focus of the study was two-fold: 1) to determine whether respondents perceived the two questions as asking about distinct constructs, as opposed to seeing them as repetitive, and 2) to identify the ways in which respondents defined or conceptualized those constructs when formulating answers. As such, the study sought to investigate construct validity as well as to provide insight into potential response error. Despite the small sample size, a range of respondent interpretations of the measure was identified—though not necessarily problematic. Many respondents saw clear, but varied distinctions between the two questions; a few respondents were either confused or did not perceive a difference. While no cases of response error were identified, some evidence suggests a possible theory for false positive responses among non-minority respondents who do not discern a difference between the two questions. Future cognitive interviewing studies with larger, more diverse samples are required to confirm and/or expand upon these interpretive findings. Quantitative studies, specifically methodological surveys with embedded construct and error probes, are required to understand the extent to which identified interpretative and error patterns would occur in actual surveys. Utilizing a mixed-method, data-driven approach to question design can lead to an optimal gender identity measure that produces the least error.

Beyond the Binary: Experiences with Gender Items in Population-Based and Non-Probability Surveys
Lisa de Vries, *Universitaet Bielefeld*
Mirjam Fischer, *University of Cologne*
David Kasprowski, *DIW Berlin*

Recent legal changes, social acceptance, and the social environment of gender minority people (e.g., transgender or intersex people) in many western countries increased the demand for reliable data about this population. Until now, there are only a few population-based datasets that allow identifying gender minority people, and lived experiences remain underresearched. Further, researchers struggle with different possibilities for measuring gender identity and the integration into existing surveys. Based on experiences with integrating a boost sample of sexual and gender minority people in a population-based German household survey (socio-economic panel) and conducting two large online surveys with the target group of sexual and gender minority people (N=7.000), we present our experiences with different items to measure gender. We will discuss the advantages and disadvantages of open and closed answer options and our results of integrating a two-step approach to identify transgender people. In addition, we will discuss the analytical potential of our measurements and methodological challenges for populations-based and non-probability samples (e.g., a small number of observations, weighting, and comparability of different measures).

Comparing Measures of Gender Expression Among U.S. Adolescents: Findings From a Split-Ballot Experiment in the Nationally-Representative Gender Identity and Sexual Health (GISH) Survey
Christopher Hansen, *NORC at the University of Chicago*
Melissa Heim Viox, *NORC at the University of Chicago*
Erin Fordyce, *NORC at the University of Chicago*
Sabrina Avripas, *NORC at the University of Chicago*
Stuart Michaels, *NORC at the University of Chicago*
Sarahjane Dube, *University of Vermont*
Masha Ivanova, *University of Vermont*
Alexandra Potter, University of Vermont

Measurement of gender expression is an evolving area of sexual orientation and gender identity research. Extant research has assessed bipolar and unipolar measures of femininity and masculinity among adults; however, no known research has evaluated their use among adolescents, who are developing their gender identity and expression. This presentation shares results of a split-ballot experiment that compared bipolar and unipolar responses to measure femininity and masculinity and assessed for differences in explanatory power across key outcomes of interest. The University of Vermont contracted NORC to conduct the Gender Identity and Sexual Health (GISH) Survey of Parents and Adolescents. A total of 279 adolescents age 13-17 were surveyed in April 2021 using AmeriSpeak®, NORC's probability-based survey panel. Youth were asked, "How would you describe yourself?" at two points in the survey. Response scales were either a bipolar, seven-point scale from "very feminine" to "very masculine" or two (masculine and feminine) unipolar, seven-point scales from "not at all" to "very". Order (bipolar or unipolar first) was randomly assigned. The sample included 49.9% cisgender males, 44.4% cisgender females, and 5.5% transgender and/or gender nonbinary adolescents. For all genders, unipolar scales captured more diversity of gender expression than the corresponding bipolar scale regardless of order. For cisgender males, 22.1% reported both feminine and masculine expression with unipolar scales compared to 5.7% using the bipolar scale. Those rates were 46.8% to 10.6% for cisgender females and 51.5% to 27.9% for transgender and nonbinary respondents. Unipolar measures predicted attention problems and parental acceptance, whereas bipolar measures predicted only parental acceptance. Findings indicate that unipolar measures capture more diversity in gender expression than bipolar measures. Regression analyses with outcomes of interest suggest this diversity is meaningful, beyond just methodological variance. Future research may consider using unipolar measures to capture greater diversity in femininity and masculinity among adolescents.

A Pathway to Measuring Sexual Harassment in the Science and Engineering Enterprise

Jennifer Beck, *National Science Foundation*
Anna Scolese, *Fors Marsh Group*
Rachel Walker-Kulzick, *Fors Marsh Group*
Jennifer Sinibaldi, *National Center for Science and Engineering Statistics*

In 2020, the National Center for Science and Engineering Statistics (NCSES) began exploratory work aimed at developing approaches to measuring the incidence and effect of sexual harassment in the science, technology, engineering, and mathematics (STEM) workforce and educational systems. Although the number of women in STEM has increased, they and sexual and gender minorities remain underrepresented. While underrepresentation stems from multiple factors, sexual harassment and gender-based discrimination contribute. With no current high-quality data set to estimate the effects of sexual harassment and gender-based discrimination on participation in STEM, NCSES initiated this investigative work to understand the experiences of women and gender minorities in STEM harassment. This presentation will discuss the qualitative work NCSES has undertaken to explore sexual harassment and gender discrimination in STEM education and the workplace and present the results of focus groups and in-depth interviews with women, men, and sexual and gender minorities. The results will help NCSES and the federal statistical system develop a pathway to measuring sexual harassment and discrimination and its effects on the STEM.

New Research on Filter & Open-Ended Questions
Toward Respondent Self-Coding
Erica Yu, *U.S. Bureau of Labor Statistics*

Survey programs that rely on open-ended responses from respondents must categorize those text responses in order to created coded data for quantitative analysis. Coding can be costly, requiring time, skilled coders, and even the development of automated coding models. And yet, even with advanced systems of coding, measurement error may still remain due to the intrinsic open-ended nature of the question: the format often does not provide guidance to respondents on what level of detail they should include or how exhaustive their responses should be. To make the shift toward respondent self-coding, there are at least five possible strategies, ranging from converting the open-end question to a closed-end question to building a real-time search algorithm for respondents to type in text and select a code from a list of search results. Each strategy has advantages and challenges that may affect whether it is suitable for a particular survey. In this talk, I will describe our approaches to the development of self-coding protocols for multiple surveys. I will discuss the use cases for each (complex code structures, potentially sensitive code labels, and mismatch with diary survey format) and the strategies being explored by each survey program.

Identifying Data Quality Challenges in Online Non-Probability Opt-in Panels Using Cognitive Interviews in English and Spanish

Yazmin Garcia Trejo, *U.S. Census Bureau*
Mikelyn Meyers, *US Census Bureau*
Mandi Martinez, *US Census Bureau*
Patricia Goerman, *US Census Bureau*
Angela O'Brien, *US Census Bureau*
Betsarí Otero Class, *US Census Bureau*

In this paper, we evaluate how the analysis of open-ended probes in an online cognitive interview can serve as a metric to identify cases that should be excluded due to disingenuous responses by ineligible respondents.   We analyze data collected via an online non-probability opt-in panel in English and Spanish to pretest a public opinion questionnaire in 2019 (n=265 in English and 199 in Spanish). We find that analyzing open-ended probes allowed us to flag cases completed by respondents who demonstrated problematic behaviors (e.g., answering many probes with repetitive textual patterns, by typing random characters, etc.), as well as to identify cases completed by ineligible respondents posing as eligible respondents (i.e., non-Spanish-speakers posing as Spanish-speakers). These findings indicate that data collected for multilingual pretesting research using online non-probability opt-in panels likely require additional evaluations of data quality.   Finally, we find that open-ended probes can help determine which cases should be replaced when conducting pretesting using non-probability opt-in panels. We argue that open-ended probes in online cognitive interviews, while more time consuming and expensive to analyze than close-ended questions, serve as a valuable method of verifying data quality and respondent eligibility, particularly for researchers conducting multilingual surveys with online opt-in panels.

Measuring Cultural Bias in NPS Verbatim Responses

Mika Jugovich, *IBM*
Jason Telner, *IBM*

Workplace effectiveness surveys are an effective tool used by organizations to assess the satisfaction and concerns of their employees, work environments, and the applications used. Many organizations are relying on Net Promoter Score (NPS) as a metric to benchmark internal and external applications, products, and services and on the NPS verbatim for further insights and how to make improvements. For large global enterprises, consisting of employees working around the globe with diverse cultural backgrounds, careful analysis of survey data that accounts for potential cultural differences in responses to survey questions is necessary to determine the proper insights and recommendations from the data. While prior research has shown that NPS has cultural biases in NPS score results, there has been little analysis on cultural differences to NPS verbatim responses. In this presentation, we will present our findings on cultural bias within NPS verbatim responses. Utilizing a large dataset of NPS verbatim responses collected from over 40 internal applications over the past 5 years, our team conducted an item-level non-response analysis. After detecting differences in non-response based on geographic location we then applied a variety of text analysis models to label the responses across a number of dimensions (sentiment, themes, key points, etc.) and then analyzed the cultural differences across each dimension.   We will discuss the underpinnings of the models as well as the cultural insights obtained from the data and the appropriate recommendations that were made based on the model's findings.

Strategic Misreporting to Avoid Follow-up Questions: The Roles of Topic Interest and Response Fatigue

Matthew Berent, *GHF9R4*
Jon A. Krosnick, *Stanford University*
Matthew DeBell, *Stanford University*

Many prominent surveys have used filter questions to avoid asking irrelevant questions to respondents. For example, all respondents in the 2019 National Crime Victimization Survey (NCVS) reported if something belonging to them was stolen in the last six months. Only respondents who offered an affirmative answer were asked follow-up questions about the number of times something was stolen and details about those events. Despite the apparent efficiency of using filter questions, such questions may motivate misreporting among some respondents. Recent research suggests that

respondents strategically misreport when they learn how much response burden will be eased by misreporting. The current research investigated two factors that might moderate strategic misreporting to filter questions in order to ease response burden. First, respondents might be more likely to strategically misreport in order avoid answering questions on topics in which they have little interest. Second, strategic misreporting might become more prevalent as response fatigue increases. A series of nine surveys using online questionnaires found support for the moderating influence of response fatigue on misreporting, but no evidence that topic interest played a role. When a filter question was used late in a questionnaire and respondents were forewarned that an affirmative answer would result in additional questions on the topic, respondents were more likely to avoid selecting the affirmative answer than respondents who were not forewarned. However, when filter questions appeared early in the questionnaire, forewarning respondents about additional questions did not alter answers to the filter question. Encouragingly, the results did not find a general tendency to strategically misreport answer to filters in all situations. However, they do suggests questionnaire designers should use filter questions sparingly, and when such questions are necessary they should appear early in a questionnaire.

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

Aaron Sedley, *Google Inc*
Yongwei Yang, *Google Inc.*

Survey communities have regularly discussed optimal questionnaire design for attitude measurement. Specifically for consumer satisfaction, which has historically been treated as a bipolar construct (Thurstone, 1931; Likert, 1932), some argue it is actually two separate unipolar constructs, which may yield signals with separable and interactive dynamics (Cacioppo & Berntson, 1994). Earlier 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, about the intensity of the selected direction (Krosnick & Berent, 1993). The current experiment evaluated differences across a variety of question designs for in-product contextual satisfaction surveys (Sedley & Müller, 2016). Specifically, we randomly assigned respondents into the following designs: Traditional 5-point bipolar satisfaction scale (fully labeled) Branched: a directional question (satisfied, neither satisfied nor dissatisfied, dissatisfied), followed by a unipolar question on intensity (5-point scale from "not at all" to "extremely," fully labeled) Unipolar satisfaction scale, followed by a unipolar dissatisfaction scale (both use 5-point scale from "not at all" to "extremely," fully labeled) Unipolar dissatisfaction scale, followed by a unipolar satisfaction scale both use 5-point scale from "not at all" to "extremely," fully labeled)   The experiment adds to the attitude question design literature by evaluating designs based on criterion validity evidence; namely the relationship with user behaviors linked to survey responses. Results show that no format clearly outperformed the 'traditional' bipolar scale format, for the criteria included. Separate unipolar scales performed poorly, and may be awkward or annoying for respondents. Branching, while performing similarly as the traditional bipolar design, showed no gain in validity. Thus, it is also not desirable because it requires two questions instead of one, increasing respondent burden.

Identification and Interpretation of Substantive Trends Following Disruptions

John Eltinge, *United States Census Bureau*

Many statistical organizations, and their stakeholders, have a longstanding interest in exploring possible "trends" in the underlying substantive phenomena, e.g., economic, public health, demographic, and agricultural conditions that we measure. Over the past two years, disruptions induced by the pandemic, and related societal responses, have heightened interest in these potential substantive trends - including some that may have predated the pandemic, and others that may have arisen more recently.   Even under ideal conditions, "trend detection", and related interpretation of empirical results, can involve complex conceptual and methodological issues. These issues became even more complicated after pandemic conditions disrupted some standard data collection and estimation processes, and also changed stakeholder priorities among multiple dimensions of data quality, e.g., timeliness, comparability, interpretability, relevance, and accuracy. In some cases, additional complications arose from implementation of design changes that were planned and initiated before the start of the pandemic.   This session explores concepts, methods, and case studies arising from the abovementioned issues. The first paper develops and applies models to improve the accuracy, timeliness, and temporal comparability of estimates from the Dutch Labor Force Survey, which uses a

rotation-sample panel design and multiple modes of data collection. The second paper considers time-trend analyses of data collected from a series of independent cross-sectional sample surveys. The third paper considers analysis of trends in health outcomes based on data from the U.S. National Health Interview Survey, with emphasis on exploring effects of questionnaire and weighting redesign, as well as the impact of the pandemic on field procedures. The fourth paper presents an in-depth case study of methodology for the analysis of within-year and across-year trends in corn yield and production predictions, based on multiple measurement approaches. The discussant will integrate, compare and contrast the concepts, methods, empirical results and communication issues considered in these four papers.

Identification and Interpretation of Substantive Trends Following Disruptions

Improving Accuracy, Timeliness, and Temporal Comparability of the Dutch Labor Force
Jan van den Brakel, *Central Bureau of Statistics*

Official monthly statistics about the Dutch labor force are based on the Dutch Labor Force Survey (LFS), which is designed as a rotating panel. The sample is designed to produce accurate estimates at a quarterly frequency using the general regression (GREG) estimator. Recognizing that besides variance, timeliness and comparability over time are also important quality indicators, Statistics Netherlands implemented a model-based inference method based on structural time series models (STM) to produce single-month labor force figures. This model solves multiple problems. First, the model is used as a form of small area estimation to produce sufficiently accurate model-based estimates on a monthly frequency to improve timeliness. Second, the model accommodates typical panel effects like rotation group bias and serial correlation as a result of partial sample overlap. Third, the model accounts and corrects for discontinuities that are the result of survey redesigns. Redesigns generally change measurement and selection bias in the data, resulting in sudden shocks in the outcomes of the survey; so-called discontinuities. Despite the occurrence of such discontinuities, the STM allow to estimate uninterrupted time series. Finally, due to the COVID-19 pandemic, face-to-face interviewing (CAPI) stopped. This mode is used in a sequential mixed mode design in combination with web interviewing and telephone interviewing. The sudden loss of CAPI respondents had a systematic effect on the outcomes of the LFS. At the same time, the lockdown marked a sharp turning point in the evolution of these series. The time series model is used to accommodate the loss of CAPI respondents and avoids confounding of the sudden change in mode effects with real period-to-period changes in the labor force figures. The paper explains how the aforementioned model-based inference method is used to improve timeliness and temporal comparability without compromising accuracy, in the production of official labor force figures.

Time-Trend Analysis With Cross-Sectional Survey Data
Dan Liao, *RTI International*
Rachel Harter, *RTI International*

Time-trend analysis has been practiced widely to study changes over time in public opinions and behaviors with survey data, especially surveys that are collected periodically using similar questionnaires, sampling, and survey administration procedures. Nonetheless, trend analysis can define "trend" differently and be performed using different tools in different research fields. Whether a statistically significant trend discovered with the weighted sample data truly indicates a trend in the population is often questionable. Furthermore, accommodating survey redesigns over time can be challenging. To review past experiences in trend analysis and better understand how trend analysis should be utilized, the authors will discuss methods in practice, first in market research from years past and second in modern official statistics. In this presentation, we will first summarize different definitions of "trend" and review a novel approach to trend analysis that was prototyped for market research. The method utilities a Kalman filter to smooth the data over time to capture both short-term and long-term trends. Then, the authors will provide a review on the current trend analysis practices in U.S. federal data collections. Although alternative methods are used for different studies, open questions remain about inclusion of survey design features into complex regression and time series models, the sensitivity of analyses to outliers, moving time windows, and the availability of new data to the current year. To explore some of these intricate issues, simulations were conducted to compare the statistical properties (e.g., classification error rates) of the most common trend testing approaches as identified through the review. The overarching purpose of this presentation is to facilitate the discussion on how trend analysis should be performed and interpreted in modern survey statistics.

Environmental Jolts and Methodological Changes: Assessing the Impacts of a Pandemic and a Redesign on Trends in the National Health Interview Survey

Jim Dahlhamer, *National Center for Health Statistics*
Stephen Blumberg, *National Center for Health Statistics*
Matt Bramlett, *National Center for Health Statistics*
Emily Terlizzi, *National Center for Health Statistics*

The impact of COVID-19 on face-to-face data collection with the National Health Interview Survey (NHIS) was substantial as field work moved exclusively to telephone in quarter two of 2020. Though in-person visits resumed in July 2020 in some areas and in September 2020 in all areas throughout the country, telephone remained the primary mode of recruitment in the latter half of 2020. The impacts of the recruitment mode switch on the realized sample were immediate, with participating households trending older, more educated, and more affluent. With respect to trends, the altered sample composition imposes challenges in disentangling mode effects from real change wrought by the pandemic. To complicate matters, the NHIS redesigned both its questionnaire and weighting procedures in 2019, increasing the potential for confounding of methodological changes with true change.   In this presentation, we examine the potential impacts of each of these three environmental and methodological changes on trends of health estimates, separately, using three different approaches. First, to examine possible disruptions in trends due to the questionnaire redesign, we review the results from a bridge sample fielded in quarter 4 of 2018 in which both the old and new designs were fielded concurrently, allowing comparisons of common content across the two designs. Second, we examine the impact of new weighting procedures on trends. To facilitate these analyses, 2016-2018 data were reweighted using the new method, while the old method was applied to 2019 and 2020 data, enabling comparisons of the two methods. Finally, we look for mode-induced biases in health estimates and possible disruptions in trends due to the pandemic by comparing 2020 estimates with various benchmarks assumed to have remained unchanged. Discussion will focus on the challenge combining these results to interpret trends from 2018 to 2021.

Understanding the Trends in U.S. Corn Production: A Case Study From the Pandemic

Linda Young, *USDA NATIONAL AGRICULTURAL STATISTICS SERVICE*
Barbara R. Rater, *USDA National Agricultural Statistics Service*

Each year the United States Department of Agriculture (USDA) produces corn supply and demand estimates for the Nation and the world. The National Agricultural Statistics Service (NASS) generates monthly yield and production predictions for corn beginning in August and extending through November for the U.S. and the highest producing U.S. corn states. Stakeholders are intensely interested in the trends in these predictions within and across years. Both grower-reported yields and objective-measurement modeled yields contain a measurable forecast error based on the historic difference between these survey estimates and the final end-of-season yield. With the adaptations made in data collection processes during the pandemic, providing these predictions and identifying the trends became more complex. In this paper, the changes made in data collection processes during the pandemic will be reviewed. The impacts of these on the survey-based predictions will be described. The utility of non-survey data in the prediction process will be explored. The final impacts on data quality will be highlighted. The anticipated long-term revisions in the process of setting these predictions will be discussed.

May 12, 2022 | 2:45 pm – 3:45 pm | Poster Session #2

COVID-19 Perspectives Across Generations

COVID-19 Perspectives Across Generations
Marjorie Connelly, *NORC at the University Of Chicago*

People of all ages are affected in some way by the COVID-19 pandemic. While older Americans have been more likely to suffer severe health consequences, the youngest adults may have other hurdles including job loss, attending college online, and social isolation. Among women of different ages, the pandemic has added to their stress of working in lower wage jobs. Stress from the pandemic has led to sleeping issues, weight gain, and problems with their skin and hair. Semilla Stripp, from the Public Affairs and Media Research department at NORC, will present data from five surveys conducted by The AP-NORC Center with funding from The SCAN Foundation. The surveys, conducted with Americans age 50 and older, look at aging issues within the context of the pandemic. During a time of heightened stress from the pandemic, many women placed a renewed emphasis on their well-being and self-care. Angela Houghton from AARP will provide the results from their third annual Mirror/Mirror Survey. The survey, which included large oversamples by generation, race, and ethnicity, examined how different women reacted to circumstances surrounding the pandemic. Betsy Broaddus from the Public Affairs and Media Research department at NORC will compare the attitudes of Gen Z with Millennials and Gen X using the results of a survey of Americans ages 13 to 56 conducted by MTV and The AP-NORC Center. A longitudinal study conducted by NORC with funding from the National Science Foundation looked at the beliefs, mental health, and outlook of Americans before, during, and after the coronavirus outbreak. Shifts in loneliness, stress, and other negative effects of the pandemic will be presented by Claire Krummenacher from the Public Affairs and Media Research department at NORC.   The panel will be moderated by Emily Swanson from The Associated Press.

Aging in America: Examining the Impact of COVID-19 on Older Generations
Semilla Stripp, *NORC at the University Of Chicago*
Dan Malato, *NORC*
Jennifer Benz, *NORC*
Trevor Tompson, *NORC*

COVID-19 has severely disrupted the lives of many Americans, and its impact has been particularly acute for older generations. Over the course of 2020 and 2021, The Associated Press-NORC Center for Public Affairs Research conducted a series of studies on aging issues within the context of the pandemic. While the pandemic has forced older adults to adapt new realities – including an increased reliance on technology to access health care and to socialize with loved ones – many key attitudes about issues relating to aging and expectations for the future have remain unchanged. Large majorities of Americans would prefer to receive any needed ongoing living assistance at home or with loved ones, even as concerns about the safety of nursing homes declined over the course of the pandemic. And, despite the impact of COVID-19 on older adults, few Americans are prepared for their own aging and potential needs for care. The studies were conducted with funding from the SCAN Foundation, and used both the AmeriSpeak Panel, the probability-based panel of NORC at the University of Chicago, and the Foresight 50+ Panel, the probability-based panel of adults age 50 and older. Five separate surveys were conducted between October 2020 and November 2021 and explored topics related to aging at various points of the COVID-19 pandemic

Women's Reflections on Beauty in the Time of Covid
Marjorie Connelly, *NORC at the University Of Chicago*
Colette Thayer, *AARP*
Patty David, *AARP*

Mirror/Mirror: AARP Survey of Women's Reflections on Beauty, Age, and Media TM is an annual nationwide survey of more than 4,800 women conducted by NORC at the University of Chicago. Results from the survey conducted in November and December 2020 reveals the effect of the coronavirus pandemic on women's attitudes about beauty standards. Many mindsets shifted as women prioritized their overall health and inner peace and cared less about their outer appearance. During a time of heightened stress from the pandemic, women placed a renewed emphasis on their well-

being and self-care. The importance of beauty and personal grooming routines shifted. Women of all ages felt less pressure to keep up their physical appearance to conform to beauty standards, and more embraced natural hair, healthy skin, and comfortable styles. The presentation will include survey data about the impact of the pandemic on beauty and wellness of different generations of women and among white, Black, and Latina women. Newer results on pandemic-related stress and women's priority on self care care from a 2021 survey will also be available in time for the conference presentation offering evidence on whether the 2020 COVID impacts were short term or have been sustained.

OK, Boomer: Exploring Sociopolitical Attitudes within Gen Z
Betsy Broaddus, *NORC at the University Of Chicago*
Dan Malato, *NORC at the University of Chicago*
Marjorie Connelly, *NORC at the University of Chicago*
Jennifer Benz, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*

Americans ages 13 to 24 – Gen Z – grew up in a digital world. Many in this generation also came of age during the coronavirus pandemic, impacting their education, entry into the workforce, and mental health. The Associated Press-NORC Center for Public Affairs Research and MTV, using TrueNorth®, conducted a survey of Americans ages 13 to 56 to understand the sociopolitical attitudes of Gen Z compared to Millennials and Gen X. The survey explores issues of identity, attitudes about social and political issues, sources of stress, outlook toward the future, and more. The data show that while Gen Z distinguishes itself from older generations on a host of factors, the upcoming generation is not monolithic and is often divided along many of the same lines as their predecessors. This presentation will examine these generational differences as well as the diversity in opinion within Gen Z, particularly by race, gender, age, and ideology. By exploring the interactions between generation and other demographics, the data present a nuanced picture of how Gen Z is responding to a tumultuous sociopolitical climate.

In These Unprecedented Times: Age Differences in Mental Health Among U.S. Adults From May-August 2020
Claire Krummenacher, *NORC at the University Of Chicago*
Laura Finch, *NORC at the University of Chicago*
David Sterrett, *NORC at the University of Chicago*
Louis Hawkley, *NORC at the University of Chicago*
Tom Smith, *NORC at the University of Chicago*

In 2020, as Americans grappled with the first summer of the coronavirus pandemic, a longitudinal study conducted by NORC at the University of Chicago with ongoing funding from the National Science Foundation asked Americans about their beliefs, mental health, and outlook before, during, and after the coronavirus outbreak. The study found shifts in attitudes unseen in similar periods of national turmoil, including an all-time low in people saying they are very happy. This study featured three waves of interviews with approximately 2,000 adult respondents between May and August of 2020. The surveys were conducted using NORC's probability-based AmeriSpeak® Panel, which is designed to be representative of the U.S. household population. In 2021, a new study conducted by NORC at the University of Chicago comparing these results with 2018 GSS data reveals that U.S. adults experienced greater loneliness and less happiness in May 2020 than in 2018. In addition, while mental health trajectories between May and August 2020 did not differ significantly by age, older adults reported less loneliness, stress, and negative affect than younger adults. However, despite older age being associated with factors linked to better mental health – namely, higher social satisfaction and lower exposure to COVID-19 – neither of these factors account for the observed age differences in mental health

Can I Get Your Attention Please? Incentives Mode & Message Salience
Household Screening During a Pandemic
Andrew Hupp, *University of Michigan*
Paul Burton, *University of Michigan*
Eva Leissou, *University of Michigan*
Brady T. West, *University of Michigan*

David R. Weir, *University of Michigan*

Every six years the Health and Retirement Study (HRS) adds a new age cohort to its existing longitudinal panel. In 2022 HRS is adding those born 1966-1971. During screening operations, an area probability sample of households is screened for age-eligible adults. Given the narrow age range, many households screen ineligible. If there are age-eligible adults, a household roster is taken to determine to whom we need to speak with. Previous HRS screening efforts have primarily been done in-person, with declining response rates and rising costs. Due to the COVID-19 pandemic and health concerns, in-person screening protocols in 2022 may be more challenging; will a person open the door to someone they don't know, even to talk for only a few minutes? In response to this uncertainty, and to trends in cost and response rates from in-person screening, a self-administered web screener was added as an option. During a preliminary field test we mailed an invitation to 1,000 randomly sampled households to complete their screening questionnaire via the web. An additional 500 households were assigned to the traditional in-person screening method. Half of the web invitations sent via mail contained $2 cash, while the remaining half contained an offer of a $2 post-paid incentive (via an electronic Tango Card). Non-responding cases could receive up to two additional reminders via mail. As part of the final reminder, half of all cases were offered an additional $5 post-paid incentive if a web screener was completed by a particular date. In this presentation, we will compare 1) the effectiveness of the web option compared to the traditional in-person option, 2) cost and response rate differences between the $2 cash and the $2 promised electronic incentive, and 3) the effectiveness of the $5 electronic bonus at pushing households to complete the screener via the web.

Exploring the Effect of Progress Feedback and Incentives on Participant Behaviors in EMA Studies
Chan Zhang, *Zhejiang University*
Florian Keusch, *University of Mannheim*

Ecological momentary assessment (EMA), also referred to as experience sampling method (ESM) or daily diary method, is a type of longitudinal study that administrates brief questionnaires at high frequency to participants, usually multiple times a day. Recently, smartphone-based EMA has regenerated the interests in this method across different disciplines. As with other surveys, getting people to respond and provide quality answers is critical for EMA studies. However, few studies have experimentally investigated the designs of EMA studies and their impact on participant behaviors. To shed some light on this issue, we design an experiment to examine the effects of incentives and progress feedback in an EMA study of the daily lives of college students. The sample is recruited from students of a large Chinese university. The EMA study lasts 21 days and surveys the students at five fixed times on a day. The experiment has a 2 (incentive) x 3 (progress feedback) factorial design. For the incentive, both two conditions offer ¥2 for each completed survey but vary on whether there is a ¥5 bonus for completing all five surveys on a day. For the progress feedback, we have three levels: (1) no feedback at all; (2) textual feedback of the numbers of surveys the participant has completed, missed, or yet to complete for the entire study; and (3) textual feedback (same as the previous design) with the visual presentation of a progress bar. We are interested in how each feature affects participants' behaviors throughout the EMA study, including the completeness and the quality of their answers and any interaction effects between the two design features. The data collection takes place in December 2021, and we plan to present the findings at AAPOR.

Effects of Mode and Incentives on Response Rates, Costs, and Response Quality in a Survey of Opioid Use among Young Adults
John Stevenson, *University of Wisconsin - Madison*
Jennifer Dykema, *University of Wisconsin-Madison*
Nadia Assad, *University of Wisconsin Survey Center*
Sarah Linnan, *University of Wisconsin-Madison*
Jolene D. Smyth, *University of Nebraska-Lincoln*
Kristen Olson, *University of Nebraska-Lincoln*

Younger adults are among the biggest abusers of prescription and illegal opioids and more data is needed to understand their perceptions and awareness of the use of these substances. Unfortunately, they are also a hard-to-survey population, presenting researchers with the challenge of how to collect high-quality data in a cost-effective manner. Two methods that have shown promise in previous research include mail-web mixed-mode designs and the use of incentives. In the current study, neither a mail-only (due to costs) nor a web-only (due to lack of email

addresses) strategy was feasible, and we sought to experiment with different designs that prioritized a web-push methodology. In addition to examining the impact of different combinations of sequential modes, we also aimed to build on recent research demonstrating the effectiveness of including a second (sequential) incentive to convert nonresponders. In Fall 2020, 4,800 adults between the ages of 18 and 25 were randomly sampled from voter registration records in Wisconsin and randomly assigned to eight experimental groups that crossed survey mode and second incentive treatments: three groups used a web-push letter only design; three groups used a mixed-mode design with web-push letters followed by a paper instrument; and two groups used a paper questionnaire-only design. While all sample members received $1 in the first mailing, depending on group assignment, nonresponding sample members either did not receive a second incentive, or were sent an additional prepaid $2 incentive in either a second or third mailing. Overall, groups that received paper questionnaires had a higher response rates and second incentives were effective, but only for paper surveys or mixed-mode designs. Additional analyses examine costs and how data quality and sample representativeness vary across the groups. Recommendations regarding the optimal design that balances quality and cost for this population will be offered.

Effects of Visible Cash Pre-Incentives on Response Rates, Respondent Demographics, and Survey Outcomes
Michelle Cavazos, *RTI International*
Matthew Farrelly, *RTI International*
Robyn Woodlea, *RTI International*

When recruiting households from an address-based sample (ABS), research has shown that providing a small cash pre-incentive with a mailed request to complete a survey is an effective approach to increase response rates. Survey mailings, however, may be considered junk mail by recipients because mailings are often addressed generically to the household (e.g., "Resident"), and therefore could be discarded without knowing that the mailing contained cash. It is important to identify effective methods to inform recipients that survey mailings include a cash pre-incentive to improve response rates and reduce the amount of cash that is discarded. DeBell et al. (2020) conducted an experiment with a windowed envelope that displayed a $5 cash incentive and found that it significantly increased the response rate compared to a standard envelope for a non-response follow-up study. To build on the findings of DeBell et al. (2020), we conducted an envelope experiment with an ABS that was recruited to complete a statewide health survey funded by a state department of health. For the experimental group (N=18,312), the survey invitation mailing included a 9x12 envelope with a large window that displayed the $5 cash pre-incentive, while the non-experimental group (N=18,311) received a 9x12 envelope that did not display the cash incentive. The protocol for the remainder of the recruitment mailings remained the same across both groups. This presentation will provide results of the experiment, including how respondent demographics and survey outcomes compare across groups. Findings of this research will contribute to the literature on the use of visible cash pre-incentives in mailed survey invitations.

Address-Based Sampling Pilot for the Behavioral Risk Factor Surveillance System (BRFSS)
Ting Yan, *Westat*
Carol Pierannunzi, *Centers for Disease Control and Prevention*
Sonya Gamble, *Centers for Disease Control and Prevention*
Reanne Townsend, *Westat*
Gina Shkodriani, *Westat*
Machell Town, *Centers for Disease Control and Prevention*

The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. and territorial residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. The BRFSS sample includes telephone numbers using landline and cell phones. Data are collected via computer-assisted telephone interviewing (CATI). We conducted a pilot study to examine whether it is feasible to conduct BRFSS online by recruiting respondents from an address-based sample (ABS). The pilot study randomly sampled addresses from one state (TX) and employed an ABS push-to-web multimode data collection strategy.   Sampled addresses were sent an advance letter, an invitation to a web survey, a postcard reminder, a nonresponse follow-up letter, and a final letter with an enclosed paper questionnaire. The pilot study includes three experiments. The first experiment assesses the effect on response rate of a web insert providing an example of the survey questions. The second experiment addresses the effect of explicitly stating a deadline on

response rate. The last experiment tests the effectiveness and cost implications of offering differential incentives conditional on the mode of completion. We will report the findings of this pilot study and offer recommendations on optimal strategies for pushing respondents to web for an ABS study.

Interviewer Monitoring & Falsification

Allowing Statistical Checks to Guide Survey Team Attention Toward Problems in Collected Data

Amrik Cooper, *Dobility (SurveyCTO)*
Christopher Robert, *Dobility, Inc.*
Melissa Kuenzi, *Dobility, Inc*

Simple statistical checks conducted during data-collection, like those that flag outliers and potential enumerator effects, can direct scarce quality-control attention toward potentially problematic enumerators and interviews. Such checks can be particularly valuable during piloting and early in the scale-up of a data-collection effort, but researchers facing tight deadlines often struggle to program or configure them before launching data-collection. Evidence is drawn from the SurveyCTO data-collection platform, where a built-in facility to create and run such checks exists, but is used by a minority of researchers. A randomized experiment is used to see whether reminders and shortcuts to create statistical checks can increase their salience or lower the cost of best practice. Finally, for context we provide self-reported details about statistical check implementation, and the prevalence of substitution using equivalent off-platform checks

Defining Quality in Face-to-Face Surveys: A Multi-Indicator Approach to Catching Low-Quality Interviews

Ariana Salazar, *Pew Research Center*
Kelsey Jo Starr, *Pew Research Center*
Neha Sahgal, *Pew Research Center*
Jonathan Evans, *Pew Research Center*
Manolo Corichi Gomez, *Pew Research Center*

Interview quality is a major concern for survey research, especially for face-to-face interviewing as interviewers are often widely dispersed in the field with little or no direct supervision for extended periods of time. During data collection, researchers often try to identify lower quality interviews and, whenever possible, those interviewers who might benefit from retraining for subsequent fieldwork. But setting criteria for identifying problematic interviews and those interviewers whose work can improve is challenging. Indicators such as interview length or non-response, often used as indicators of interview quality, are generally not strong enough on their own to reject an interview or diagnose whether interviewers would benefit from supplemental training.   This paper explains a multi-indicator approach to catching low-quality interviews using data from a recent Pew Research Center survey conducted face-to-face with nearly 30,000 adults in India. Several indicators – including time stamps, geocodes, non-response, recontact protocol and implausible variable combinations – were used to identify cases for investigation. But not all indicators were treated equally – only some were considered sufficient to reject an interview on their own. Most were considered flags for potentially low-quality interviews that would need to be investigated further by the field managers. These weaker, or "soft check," indicators were evaluated as a whole: Flags across multiple indicators were required to prompt investigation into problematic cases or the work of an individual interviewer. This presentation will also touch on how back-checks were used in quality reviews. The data show the advantages of a multi-indicator approach in managing quality control processes for face-to-face surveys.

Using Speech Analytics Software To Enhance an Interviewer Monitoring Program

Jason Rajan, *NORC at the University of Chicago*
Lauren Hartsough, *NORC at the University of Chicago*
Erin Criste, *NORC at the University of Chicago*
Beth Baca, *NORC at the University of Chicago*
Kate Hobson, *NORC at the University of Chicago*

Call centers, particularly of the large inbound customer service type, have increasingly been using speech analytics programs to enhance or replace their agent monitoring systems.   Speech analytics (also called voice analytics) is the branch of data science that deals with computerized processing of recordings and subsequent transcriptions to extract information.   (Mishra and Sharma, 2016). NORC has been investigating the use of one of these packages to enhance their telephone interviewer monitoring program (Rajan et al, 2021). Although there are considerable differences in how this software can be used for unstructured customer service situations vs outbound social science surveys, we have still found several significant benefits in using it. While industry standard for quality programs tends to be monitoring 5 to 10 percent of contacts, Speech Analytics programs can increase the volume of cases reviewed to almost 100 percent of contacts. We've found however that these systems don't fully replace the need for human review of calls in telephone survey quality processes. This talk will overview the promise of the software, and how we are incorporating it into our quality program specifically looking at what it does and does not replace from the existing human call review processes. We'll also address on its reception by interviewers and quality monitors, and investigate the gains in productivity and interviewer success found to date.

Man vs Machine: A Comparison of Multivariate Machine Learning Techniques for Rooting Out Data Falsification
Scott Glendye, *U.S. Census Bureau*
Sheldon G. Waugh, *U.S. Census Bureau, Office of Survey and Census Analytics*
Rafael E. Puello, *U.S. Census Bureau, Office of Survey and Census Analytics*

The U.S. Census Bureau's Field Quality Monitoring (FQM) team in the Office of Survey and Census Analytics created a program to monitor interviewers throughout the 2021 American Housing Survey (AHS) operation. Interviewers were investigated if they were the prime actor in a collection area that had anomalous data in any of a select set of metrics. Anomalies were identified using interquartile ranges (IQR), with collection areas flagged if upper or lower thresholds for identification exceeded 1.5*IQR. This flagging methodology spawned investigations that resulted in hundreds of interviewers counseled, retrained, or observed. However, even more interviewers were flagged and identified as not having any suspicious patterns at all. The two major concerns for the FQM program, and for quality control more broadly, are identifying anomalies sooner and minimizing the costs associated with reviewing false positives. Like the work performed in FQM, a review by Bredl, Winker, and Kötschau (2012) discovered that most survey falsification models consider single indicators, or several indicators considered separately. Following the work of Bredl et al, we similarly believe that a multivariate approach could be more effective at identifying potential falsification sooner, and with greater accuracy. In this paper, we consider several novel multivariate models; Cook's distance, isolation forest, and extreme gradient boosting outlier detection (XGBOD), that may show promising results. This paper uses metrics for item nonresponse rate, time in interview, and final interview outcomes during the 2021 AHS operation, in conjunction with the results of the FQM investigations to generate a training set for multivariate models. We then compare the results of the multivariate models to the IQR approach to see if they outperform IQR in detecting the interviewers sooner and with a lower incidence of false positives .

WFH in Data Collection: Assessing the Future of Remote Interviewing
Gary Langer, *Langer Research Associates*
Patrick Moynihan, *Pew Research Center*
Clark Letterman, *Pew Research Center*
Gary Langer, *Langer Research Associates*

The coronavirus pandemic forced fundamental shifts in data collection in the probability-based survey research enterprise. Face-to-face interviewing shuddered to a halt, replaced where feasible by computer-assisted telephone interviewing. CATI, for its part, moved from call centers to remote interviewing, with interviewers tunneling in from their homes via virtual private networks to connect with call management and supervision systems. How did remote calling perform, and will it remain a viable option in the post-pandemic future? We analyze data from two international Pew Research Center surveys conducted in spring 2021, one in Russia and one in Singapore, comparing productivity and data quality among at-home and call-center interviewers.

Increasing Response Rates & Reducing Nonresponse Bias

Ante Up: Testing the Effectiveness of Bonus Incentives for Panel Participants on Performance

Allysha Kochenour, *Nielsen*
Adam Gluck, *Nielsen*

Nielsen's Portable People Meter panel asks panelists to wear a small device that detects inaudible codes in radio and television broadcasts, which is then used to estimate audience sizes for said broadcasts. Panelists are asked to wear their meter every day, from the time they wake up to the time they go to sleep; however, some do not. To encourage panelists to wear more often, Nielsen periodically offers bonus monetary incentives to panelists if they wear the meter for a minimum, cumulative number of hours in a given week. As long as panelists meet the minimum, they'll earn this bonus incentive. This paper uses the Nielsen Portable People Meter panel to examine the effectiveness of bonus incentives on compliance behaviors. Nielsen periodically tests ways to improve the effectiveness of bonuses by varying incentive structures and incentive types. Recently, Nielsen began testing the effectiveness of bonus incentives using a control group. Bonus eligible panelists were randomly assigned to one of two groups: a test group that received an invitation promising a bonus incentive and a control group that received no invitation or bonus. Participants in the test group who met the minimum wear time, outlined in their invitation, received the bonus incentive. Control group participants who met the minimum wear time received no bonus incentive. We plan to compare the differences in compliance behaviors (meter wearing rates) between the test and control groups to determine the effectiveness of bonus incentives for panel participants in a longitudinal study. These findings can help Nielsen and other institutions running longitudinal panels make decisions on allocating future incentives.

Using Precinct Data to Adjust for Recalled Vote

Jimmy Lederman, *andrew.cmu.edu*
Mickey Jackson, *SSRS*

Differential nonresponse among a certain subset of Republicans is a suspected contributor to bias in 2020 election polling. Specifically, it is suggested that Republicans who answer polls vote differently on aggregate than Republicans who do not answer polls. We propose a method for reducing this error in election polling. A recalled vote adjustment—weighting self-reported candidate choice in the previous election to that election's actual results—is one potential solution. However, the AAPOR task force on 2020 pre-election polling found that such an adjustment did not significantly correct for the 2020 polling error. We hypothesize, however, that the impact of recalled-vote adjustment varies by geography: in particular, the adjustment is most effective where support for Trump is highest. Given this hypothesis, a recalled vote adjustment at a more detailed geographic level may better upweight hard-to-reach Republicans. Specifically, we test a method that groups respondents' precincts based on past Trump performance. Conducting a recalled vote adjustment within every precinct grouping then hopefully upweights the under-surveyed subset of Republicans. Using a September 2020 registration-based survey of the 2020 presidential election in three swing states, respondents were linked to precinct-level results of the 2016 election and reweighted using this precinct-informed recalled voting adjustment (among those who reported voting in 2016). This adjustment was compared to the overall recalled vote adjustment and a similar adjustment using county-level results. We find that that the precinct-informed adjustment reduces bias above and beyond the other approaches in two out of three states. This presentation will evaluate the extent to which these methods are able to reduce bias in election polling and account for limitations in each approach. These results will provide insight into the potential for precinct-informed weighting adjustments to correct for geographically entangled nonresponse errors.

We're Back: What Happens When You Repeat a Non-Panel Survey?

Rizwan Javaid, *Internal Revenue Service*
Scott Leary, *IRS*
Kris Pate, *IRS*
Brenda Schafer, *IRS*
Patrick Langetieg, *IRS*
Jocelyn Newsome, *Westat*
Lavaughn Cadiz Gooden, *Westat*
Martha Stapleton, Westat

Kerry Levin, Westat

The IRS Individual Taxpayer Burden (ITB) survey is an annual multi-mode survey sent to 20,000 individuals in the US. The survey gathers data about the time and money taxpayers spend complying with their tax reporting responsibilities. The annual ITB sample is deconflicted with prior-year survey samples to ensure that no taxpayer will be selected to participate in the survey more than once. However, to better understand the impact on taxpayer compliance burden that resulted from the sweeping tax law changes included in the Tax Cuts and Jobs Act of 2017, the IRS conducted a natural experiment and sent an ITB18 survey to approximately 6,500, ITB17 respondents. The ITB18 invitation letter thanked the taxpayers for responding to the ITB17 survey and invited them to take the survey again for 2018. The Tax Year 2018 responses which reflect changes in the law will be compared to their 2017 pre-law change responses. Research has examined ongoing participation rates for panel surveys. For example, a 2013 study of six national phone panel surveys (Schoeni 2013) found that around 90 percent of prior-year respondents responded to the survey in the ensuing year. The IRS ITB18 follow-on survey provides a unique look at the propensity to respond to an unexpected request to complete the same survey for a second time. In addition, it provides additional insight into who is most likely to respond to the ITB survey. This paper will discuss response rates, frequency and types of customer service calls (e.g., respondents confused by potentially duplicative survey), data reliability, and completion mode (paper vs. web). In addition, response by demographic variables (e.g., strata, age, income, presence of dependents, geospatial location) will also be discussed.

Survey Break-offs and Questionnaire Design in Politically Sensitive Contexts

SangEun Kim, *Vanderbilt University - Nashville, TN*
Carole J. Wilson, *Vanderbilt University*
Elizabeth J. Zechmeister, *Vanderbilt University*

Survey break-offs have implications for efficiency and data quality, but only a small body of research examines their causes and consequences. We present results from a quasi-experiment focused on questionnaire design. While most break-off research focuses on self-administered web surveys in developed contexts, our data are from a national random-digit dial (RDD) computer-assisted phone survey in Nicaragua. The country's politically sensitive and polarized environment created conditions for high break-off rates: over 30% overall. The study's intervention is the placement of personal questions – gender and location (municipality) – at either the beginning (condition 1) or end of the survey (condition 2). Under a responsive design approach, the location of the module shifted mid-fieldwork, an intervention that can be treated as if random under RDD. We find a stark difference in break-off rates: about 60% in condition 1 and 30% in condition 2. We further show that in condition 1, personal information questions are among the strongest predictors of break-offs, yet these questions are not key predictors in condition 2. In addition to cost implications, differential break-off rates have implications for data – if those inclined to break-off when personal questions are asked at the start are fundamentally different from those who remain. In fact, we find only small differences: once demographic variables are controlled, only a small set of questions are statistically different across conditions. In brief, in a politically sensitive context, questionnaires that ask personal information upfront can elicit high break-off rates, but may have only minimal consequences for inferences about public opinion.

Testing the Effect of Three Incentive Groups on Responses in a New Rural State Population-Based Health Survey

Emily Robinson, *West Virginia University - Health Affairs*
Ruchi Bhandari, PhD, MPA, MBA, *West Virginia University School of Public Health*
Angela Dyer, PhD, MSPH, *West Virginia University Office of Health Affairs*
Emily Robinson, MA, *West Virginia University Office of Health Affairs*
Gordon Smith, MD, MB, MBChB, MPH, *West Virginia University School of Public Health*
Christina Mullins, MA, *West Virginia Department of Health and Human Resources*
Cynthia Beane, MSW, LCSW, *West Virginia Department of Health and Human Resources*

A new population-based statewide survey, the Mountain State Assessment of Trends in Community Health (MATCH), was fielded in West Virginia (WV) from September 2021 – February 2022 using an address-based, stratified random sample and two-wave, multi-mode approach (push to web, telephone call-in, and paper). The goal of the MATCH is to obtain representative health data on non-institutionalized adults in every WV county. Therefore, it is important to understand how adults in rural WV respond to different survey incentives since these differences are less known in rural states and vary by survey mode. So, an incentive experiment was designed for the first survey wave, fielding for 2 months, to investigate differences in response and demographics among three different incentive groups: $2 bill pre-incentive, $10 post-incentive, or both the $2 bill pre-incentive and $10 post-incentive gift card offer. The 44,002 households included in the first wave each received four outreach attempts via postal mail. The three incentive groups had 14.6%, 16.4%, and 19.9% response rates, respectively. Responses among the $2 pre+$10 post-incentive group were 69.8% higher compared with the $2 pre-incentive group. Also, individuals in the $10 post-incentive group had significantly lower odds of responding compared to those in the $2 pre+$10 post-incentive group (odds ratio (OR)= 0.79, 95% CI [0.75, 0.84]). Among those who responded, univariate multinomial logistic regression showed that compared to paper survey, web survey had significantly higher response rates in the $2 pre+$10 post-incentive group compared to the $2 pre-incentive group (OR=1.57, 95%CI [1.40, 1.76]). Also, individuals with low educational attainment had significantly higher response rates in the $2 pre+$10 post- incentive group compared to in the $2 pre-incentive group (OR=1.21, 95% CI [1.05, 1.36]). Presenters will share these and other results with other survey practitioners to improve incentive methodology for future rural surveys.

Differential Join Rates as a Proxy for Differential Nonresponse

Tracy Keirns, *University of New Hampshire*
Sean P. McKinley, *University of New Hampshire*

As the cost of traditional telephone surveys continues to increase, and as response rates continue to decline, many survey organizations are turning to online survey panels rather than a traditional RDD methodology for statewide surveys on political and apolitical topics. The University of New Hampshire Survey Center has constructed statewide probability-based panels of New Hampshire, Maine, and Vermont residents. By inviting participants in a randomized way, we are looking to mitigate the major criticism faced by panels constructed from non-probability samples as well as providing viable panels for state-level research which is often unavailable in national panels. Given the changing landscape of survey recruitment methods, nonresponse bias is no longer solely determined by those groups who disproportionately do not respond to a given survey invitation. It is critical that as researchers we consider the upstream nonresponse as an important component of understanding the full extent of bias in our processes. Members of UNH Survey Center panels are recruited using a variety of probability-based sampling methods including traditional RDD telephone surveys, SMS and MMS text message invitations, and ABS sampling. This paper examines differential survey participation rates and panel join rates by recruitment mode and method as a component of the increasing concern over differential nonresponse. We look at the attitudinal and demographic characteristics of those electing to take our recruitment surveys and those who opt into the panels. Early analysis shows support for differential survey participation rates but no differences in join rates, suggesting attention should be focused upstream as well as on the final nonresponse

An Offer They Can't Refuse? Evaluating Engagement Techniques Among Chronic Non-Responders in Probability Panels

Kyle Berta, *SSRS*
Chintan Turakhia, *SSRS*
Jennifer Schmidt, *SSRS*

Respondent engagement and retention are critical in preserving high data quality for probability-based panels and longitudinal studies. In probability-base web panels, panelists can be broadly divided into three categories: some panelists are highly engaged and respond to survey invitations frequently, some are less engaged and respond to survey invitations infrequently and then there are some who go through panel registration only to never respond to any survey invitations (chronic non-responders). Who are these chronic non-responders? Are they a lost cause or is there a way to entice them back into the panel by reaching them via a different mode and an added financial incentive? In this methodological brief we explore the demographic composition of these chronic non-responders. We test whether

reaching these chronic non-responders via a different mode and offering significantly higher incentives is enough entice them to participate in their first panel survey. Furthermore, for panelists that do participate in the first survey, we track their survey participation over time.   For this experiment, we utilize the probability based SSRS Opinion Panel and test engagement outreach to panelists via mail, email, and text messaging.

Quantitative in Election Polling

Shoring up the Levy – Using Combinatorial Approaches to Adjust Election Polls

Clifford Young, *Ipsos*
Chris Jackson, *Ipsos*

   The 2020 election was a rough year for polling in the United States. The polls, on average, missed the mark at the state and national levels by around 5 percentage points and exhibited the largest error in the national vote in 40 years.This came after the industry spent significant time and resources "fixing" the problems that caused polls to pick the wrong winner in the 2016 election. The 2020 AAPOR post-mortem points to several potential issues, but provides no definitive indication as to what went wrong, and as a consequence, the ability of the election polling industry to fix the problems before the next election is in question.   In this paper, we test a novel approach to adjusting election polling estimates. Rather than focus on a single approach which may or may not be accurate, we make the simplifying assumption that all methods have their flaws. As such, a combination of approaches may overcome the inherent error of any single method. In particular, we employ a fundamentals election forecasting model developed in an international context to adjust election poll estimates. In total, we will test the above combinatorial approach in 50 elections in the U.S. and other countries. We will use AAD (average absolute difference) metric as our primary measure of performance. We also discuss how such an approach can be deployed at the aggregate (averaging) and individual levels (weighing).

The Impact of Weighting by Past Vote on Estimates of Pre-Election Voting Intentions: A Case Study Using Australian Polling Data (With Possible Implications for US Pollsters)

Darren Pennay, *Australian National University*
Sebastian Misson, *The Social Research Centre Pty Ltd*
Dina Neiger, *The Social Research Centre Pty Ltd*
Paul J Lavrakas, *The Social Research Centre Pty Ltd*

Applying 'weighting' procedures to reduce bias in the estimates of voting intentions produced by pre-election opinion polls is an important issue for pollsters and of the main issues tackled by the Inquiry into the performance of the 2019 pre-election polls in Australia. This presentation builds upon the work of that Inquiry, and, in particular, looks at the efficacy of adding measures of past voting behaviour to the standard weighting solutions typically used by pollsters. The data for this study are from a pre-election survey conducted by the Australian National University on the Life in AustraliaTM probability-based online panel in April 2019. We find that poststratification to past vote benchmarks reduces bias to a non-ignorable extent when the measure of past vote used for this adjustment is collected from panellists in close proximity to the previous election (i.e. a short-term recall measure of past vote). This reduction in bias is not observed when using a measure of past vote collected from panellists proximate to the upcoming election (i.e. a long-term recall measure of past vote). The AAPOR Taskforce report on 2020 Pre-election Polling found that "polls using partisanship and past vote in their weighting were no more accurate (than polls that didn't)" (p.70). To the extent that the measures of past vote being used in these polls were affected by measurement error, our research suggests that a less biased measure of prospective voting intentions may be obtained if the past vote weighting adjustment is applied to a measure of past vote collected in close proximity to the previous election rather than one collected in close proximity to the upcoming election. Given that 64% of the national pre-election polls conducted in the US in 2020 were conducted on online panels, this is a possible area for exploration by panel providers and pollsters.

Applying Swing Vote Analysis with Multi-level Regression and Poststratification (MRP) to Improve U.S. Pre-Election Polling Estimates

Kaan Cem Ketenci, *University of Michigan*

Various forms of pre-election survey errors cause substantial variation in the U.S. pre-election estimates of vote shares. Due to differential rates of survey non-response among voters of different parties, there appears to be inaccurate large fluctuations in estimated vote shares (Gelman et. al., "The Mythical Swing Voter"). In this paper, a swing vote analysis method is proposed to substantially reduce the variance in pre-election survey estimates and eliminate significant biases due to survey non-representativeness. The paper uses longitudinal panel datasets by the Pew Research Center and ANES (American National Election Studies) covering pre-election and post-election surveys of 2016, 2018 and 2020 U.S. presidential and midterm elections, supplemented by vote validation data. The focus is on the swing vote, i.e., those who respond as intending to vote differently than how they have voted in the most recent previous elections. Estimates of swing between the Democratic and Republican Party, as well as swing between voters of either of the two parties and non-voters / 3 rd party voters are calculated. Multi-level regression and post-stratification (MRP) methods as well as likely voter models are applied to obtain a weighted sample consistent with the distribution of actual voters in significant demographic dimensions. The reduction in variances of vote share estimates is greater especially when swing between voters of different parties is smaller. Hence, this method produces more accurate results when the electorate is very polarized, as repeatedly observed in recent U.S. elections. It is found that variances of estimates of two-party vote shares are reduced by a factor of 3 to 5 by swing vote analysis in each of the analyzed election cycles in this paper. Furthermore, when vote share estimates are decomposed into different demographic dimensions, the findings are consistent with those of the exit polls, hence suggesting improved accuracy in estimates.

Applying Quasi-Experimental Methods to Measure the Impact of Unanticipated Mid-Field Events on Political Attitudes

Michael Jackson, *SSRS*
Jennifer Agiesta, *CNN*

Recent polling errors have prompted exploration of alternatives to random digit dialing for political polling, with the goal of mitigating differential nonresponse. One alternative is address-based sampling (ABS). Advantages of ABS include the ability to offer multiple modes, provide pre- and post-incentives, and use linked auxiliary data to direct follow-up efforts to underrepresented subgroups. A tradeoff is that a mixed-mode ABS with mailing components requires a relatively long field period. This may limit the suitability of ABS for measuring political outcomes such as presidential approval, which are sensitive to fast-breaking external news events. Such events are often unanticipated and therefore difficult to design for in advance. Often, when a high-profile news event occurs during an ABS field period, the only option for assessing its impact is an "after-the-fact" comparison between pre- and post-event respondents. Therefore, in assessing the utility of ABS for political polling, an important question is whether such "pre-post" comparisons can validly estimate the impact of mid-field events. The inferential challenge is that later respondents may differ from earlier respondents on characteristics that are unrelated to the event but could confound observed changes in political attitudes. Using an ABS poll whose field period overlapped with the August 2021 fall of Kabul, Afghanistan, this presentation will assess methods of estimating the impact of high-profile, mid-field events on presidential approval and related outcomes. Simple demographic reweighting will be compared to quasi-experimental methods drawn from the causal inference literature, including propensity score and regression discontinuity approaches. We will evaluate the extent to which these methods can control for underlying differences between early and late respondents and obtain plausibly unconfounded estimates of event-driven changes in attitudes. Results will provide further insight into the tradeoffs of using ABS for political polling and the types of questions that can be answered with such samples.

Demographic Survey Weighting Using the Voter File

Valerie Bradley, *University of Oxford*
Meg Schwenzfeier, *Harvard University*

Widespread bias in 2016 and 2020 pre-election polling has been largely attributed to heterogeneous partisan nonresponse that is unexplained by observable characteristics (AAPOR 2021). In response, many survey researchers have begun to explore new methods for post-survey adjustment that can more flexibly accommodate non-traditional types of auxiliary data. Voterfiles and other population databases contain extensive, highly-predictive auxiliary data, but do often lack good individual-level data on important features, like education and race, which instead are estimated using predictive modeling. However, the best method for incorporating these modeled estimates into weighting and modeling of survey data remains unclear.   Here we compare the performance of two common methods

for incorporating modeled scores into analysis of survey data: 1) direct use of the underlying predicted probability of an individual belonging to a particular demographic category and 2) discretizing the score by assigning each individual to a category based on a single sample from the implied underlying probability distribution. We use simulation studies and 2020 ANES data to evaluate these methods based on the accuracy of resulting survey estimates and each method's ability to preserve underlying joint distributions when multiple scores are used. We also introduce a metric called leverage designed to prioritize auxiliary variables in weighting and analysis. This method differs from other methods of auxiliary variable selection for weighting because it does not require any population data for the features of interest, so can be applied to features only collected in-survey as a method for prioritizing benchmark data acquisition.

Inclusion & Equity in Public Opinion Research & Organizations
Inclusion and Equity in Public Opinion Research and Organizations
Dana Garbarski, *Loyola University*

QUALPOR Panel: Effective Techniques for Increasing Qualitative Credibility & Transparency
Providing Effective Feedback to Maintain Quality and Consistency in a Qualitative Interviewing Team
Casey Tesfaye, *Research Support Services Inc*
Jessie Engel, *Research Support Services*
Daniela Glusberg, *Research Support Services*

Over the course of a number of thematically related qualitive in-depth interview studies, one project team has developed and implemented methods for providing ongoing feedback to interviewers throughout data collection, with a goal of evaluating quality and maintaining consistency. These methods include: evaluating a subset of interviews from each interviewer at prescribed intervals, evaluating interviewer quality along a set of key dimensions, providing interviewers with annotated transcripts when needed, and having regular debriefing sessions that focus on interview methods rather than interview findings. In this presentation, we will discuss the domains that have been most useful for providing effective feedback (such as "Covered all questions" "met question intents" "probed sufficiently" and "avoided leading probes"), the response format that has been most useful for each of the dimensions (not closed-ended), the type of inline feedback most effective in annotated transcripts, and a guide for productive discussion during debriefing meetings.Over the course of a number of thematically related qualitive in-depth interview studies, one project team has developed and implemented various methods for providing ongoing feedback to the interview team throughout data collection, with a goal of evaluating quality and maintaining consistency between interviewers. These methods include: evaluating a subset of interviews from each interviewer at prescribed intervals, evaluating interviewer quality along a set of key dimensions, providing interviewers with annotated transcripts when needed, and having regular debriefing sessions that focus on interview methods rather than interview findings. In this presentation, we will discuss the domains that have been most useful for providing effective feedback (such as "Covered all questions" "met question intents" "probed sufficiently" and "avoided leading probes"), the response format that has been most useful for each of the dimensions (not closed-ended), the type of inline feedback most effective in annotated transcripts, and a guide for productive discussion during debriefing meetings.

Trading Reliability, Generalizability, and Validity for Transferability, Dependability, and Confirmability: Building the Credibility of Qualitative Datasets and Analysis
Danielle Augustine, *The University of Georgia*
Melinda, *Moore*

Qualitative methods are particularly useful in settings when research is addressing complex issues that directly impact the lives of the participants, and that in doing so, you can describe the voice of the individual. Social science research tends to investigate quantified areas and neglect the broader understanding of individual experiences. By shifting from an etic to an emic approach, qualitative methods can identify specific values, beliefs, and behaviors of the individual. However, any advantages from the qualitative method are at the expense of reliability, inclusiveness, and generalizability of the findings. To address those concerns, researchers can instead focus on credibility, transferability, dependability, and confirmability of the qualitative research. Techniques to establish credibility are prolonged engagement in the field, persistent observation , triangulation, peer debriefing, negative case analysis, and member

checking. Techniques to establish transferability are thick description of the findings, in order to apply the research to another setting. Techniques for establishing confirmability include triangulation, audit trails, and reflexivity to address issues of researcher bias. Using the researchers' 15 years of experience in qualitative methods in a governmental survey research and evaluation unit at a public university, this panel will discuss working with focus group and interview datasets and its' subsequent analysis, including data collection, preparation, and analysis within a team of researchers. Included in the discussion are techniques for determining the credibility of qualitative data while working within the constraints of client-contracted research demands. Study examples include focus groups and interviews with law enforcement agency officials working on human trafficking, STEM workforce development, early care and learning, among others.

Ethical Access to Vulnerable and Hard-to-Reach Populations: Lessons From Qualitative Research Studies With Female Combatants in Northern Ireland and Sex Workers in North America
Danielle Augustine, *The University of Georgia*

Building on the researchers' experiences studying female combatants in the Provisional Irish Republican Army in Northern Ireland (considered a terrorist group by some, a revolutionary army by others) and also with sex workers in North America, this panel will discuss how to ethically access hard-to-reach populations for qualitative research, including how to build trust and credibility, locate gatekeepers in the communities, maintain neutrality and apolitical positions, model non-judgmental attitudes, and at all costs safeguarding confidentiality and anonymity of participants. Internationally, limited research exists on the experiences of combatant women in armed conflict; existing research cites the underrepresentation of the experiences and voices of combatant women in political and peace-making discourses of post-conflict societies. Domestically, limited research exists on the experiences of sex workers and existing research describes the difficulty in gaining access to sex workers in an ethical and confidential manner. Therefore, a foremost initial challenge in conducting qualitative fieldwork with hard-to-reach populations involves gaining access to the sample and addressing researcher status as an outsider to those groups. Establishing trust and legitimacy is paramount to the research process in general and to the access issue in particular. Additionally, issues of access to the population of study may require longer periods of trust building and communication between gatekeepers and the researchers than originally perceived. Confidentiality and the safety of the research participants may influence the methods of data collection, issues of informed consent, and compromise the ability to keep detailed lists or information on locations and participants. Finally, the researcher must constantly weigh issues of personal safety with issues of data collection in the field. Data collection methods for these studies included semi-structured interviews, field notes, observations, photographs, document review, and historical political data. Snowball sampling was used to recruit participants for both studies once a gatekeeper was established.

Increasing Data Transparency and Credibility Through Reflexivity in Cognitive Interviewing Studies
Zachary Smith, *National Center for Health Statistics*

R esults of cognitive interviewing studies are used to improve question design and contextualize survey data for the end user. However, the data on which these studies are based often contain only the transcript of respondents' answers to verbal probing , the notes written down by the interviewer , or, sometimes, both . The nature of th ese data obscures what has become a crucial evaluation criterion for qualitative research more broadly: reflexivity, or the critical examination of a researcher's own role and influence in the generation, analysis, and reporting of data. Far from seeking to eliminate "bias" in qualitative research, reflexivity is intended to increase transparency and credibility by clearly denoting the ways in which subjective decisions and unconscious processes guide the co- construction of data. T hough reflexivity is discussed as part of cognitive interviewer training, cognitive testing reports rarely mention the role of the interviewer in this way. T o address this gap and contribute to the transparency of cognitive interviewing studies, t his paper draws on the cognitive evaluation of questions about cannabis use conducted by the Collaborating Center for Questionnaire Design and Evaluation Research at the National Center for Health Statistics. As part of the study design, for each interview, interviewers were asked to reflect on any ways they felt they influenced the direction of the interview or the data gathered. This study combines these written reflections with in-depth, semi-structured interviews with each member of the interview team. Its goals are to document the ways interviewers perceived themselves as implicated in the process of data generation and to contextualize the interpretation of the cognitive interviewing

findings. Through this process, the paper also offers preliminary recommendations of how to communicate reflexivity in the documentation of cognitive interviewing findings and thereby improve data quality .

The Challenges of Qualitative Research with Persons Living with HIV
Kai Fuentes, *Ebony Marketing Systems Inc,.*

When considering how to conduct qualitative research involving sensitive topics that include hard-to-reach and vulnerable population segments,   research with persons who are HIV+ is pre-eminent. The challenges are many: (1) establishing trust not only with prospective research targets but with treatment facilities and CBOs who are gatekeepers to the population; (2) becoming familiar with the issues and phraseology surrounding the virus; (3) understanding ethnic-bound perceptions and attitudes; and (4) acknowledging personal biases. This session will explore how these challenges impact research design, confidentiality protocols, recruitment strategies, screening, selection of interviewers, interviewer training, IDI and focus group strategies. Ebony Marketing Systems is a multicultural, multilingual market research firm with over a decade of experience in conducting research with persons who are HIV+. Throughout this time, EMS has gained invaluable experience developing strategies, skills, techniques, and technology to effectively garner research insights for its clients. EMS' CEO Kai Fuentes will convey practical strategies to conduct research with this population.

Measuring Sex, Gender Identity, & Sexual Orientation: Findings of the National Academies of Sciences, Engineering, & Medicine

Measuring Sex, Gender Identity, and Sexual Orientation: Findings of the National Academies of Sciences, Engineering, and Medicine
Tara Becker, *National Academies of Sciences, Engineering, and Medicine*
Nancy Bates, *U.S. Census Bureau*
Bianca Wilson, *UCLA*
Kellan Baker, *Whitman-Walker Institute*
Katharine Dalke, *Pennsylvania State University*
Karen Parker, *National Institutes of Health*
Christina Dragon, *National Institutes of Health*

The growing visibility of sexual and gender diversity within the population has prompted a re-examination of the ways in which data on sex, gender, and sexual orientation are collected. As part of these efforts, the Sexual and Gender Minority Research Office of the National Institutes of Health asked the National Academies of Sciences, Engineering, and Medicine to convene a committee of experts to review current data collection practices and develop recommendations for measures of sexual orientation, gender identity, and sex for use in survey research, administrative, and medical and clinical contexts. Implementation of these measures would enhance the quality and comparability of data collection efforts and more improve identification of a broad range of sexual and gender minorities within these contexts. This panel will outline the committee's findings and recommendations and discuss plans to implement them within the federal government. Specifically, the five panel presentations will cover the following topics: 1.   Developing Standards for the Measurement of Sex and Gender (6180); 2.   Balancing Science and Community Innovation in Measuring Sexual Orientation Identity (6056); 3.   Measuring Gender Identity for the National Institutes of Health: Recommendations from the National Academy of Sciences, Engineering, and Medicine Consensus Committee Report (5297); 4.   Developing Measures of Intersex Status/Differences in Sex Development at the Population Level (6120); 5.   Implementing Affirming Sexual and Gender Minority Measures: Making the NASEM Report "Measuring Sex, Gender Identity, and Sexual Orientation" Actionable (6025). Together, this panel of talks will provide a comprehensive review of the conclusions and recommendations outlined in the committee's consensus report.

Developing Standards for the Measurement of Sex and Gender
Kellan Baker, *Whitman-Walker Health*

Sex and gender are often conflated in surveys and other data collection instruments. These two variables, however, are distinct and multidimensional constructs that are both contextually and temporally fluid and may not align with each

other. When measures fail to provide clarity on the dimensions of sex and/or gender that they seek to assess, respondents are left to decide how to interpret the question, resulting in mismeasurement and poor data quality. To address this data quality problem and to advance efforts to field measures that can identify transgender populations, the National Academies of Sciences, Engineering, and Medicine have convened a consensus study committee to develop standards for the measurement of sex and gender. The report also assesses the measurement of sexual orientation, which is linked to sex and gender by the degree to which sexual orientation is commonly defined by the gender of a person's romantic partners in relation to the person's own gender. This presentation will discuss the committee's approach to defining the constructs of sex, gender, and sexual orientation, as well as the principles the committee developed to guide its deliberations on strategies to improve measurement of these constructs. It will also detail the committee's conclusions and recommendations regarding the measurement of sex and gender in surveys and other research, administrative contexts, and clinical settings.

Balancing Science and Community Innovation in Measuring Sexual Orientation Identity

Bianca Wilson, *UCLA*
Karina Walters (Choctaw Nation of Oklahoma), *UW, School of Social Work*
José A. Bauermeister, *School of Nursing, University of Pennsylvania*

As one of the three primary dimensions of sexual orientation, sexual identity is the psychological, community and potentially political component. Sexual identity is the dimension of sexual orientation most consistently tied to experiences of material forms of heteronormative-based discrimination (Sears, Mallory, Flores, & Conron, 2021; Wilson, Gomez, Sadat, Choi, & Badgett) and noted explicitly in non-discrimination laws and policies aimed at protecting or harming sexual minorities (e.g., H.R.5 - Equality Act). Documenting population size estimates, characteristics, and outcomes with regard to sexual orientation identity make it possible to understand changes to and systemic disparities within the U.S. population. In order to support the inclusion of measures of sexual orientation and other domains relevant to sexual and gender minorities, the National Institutes of Health recently sponsored the National Academy of Sciences' Committee on Measuring Sex, Gender Identity, and Sexual Orientation to review and recommend measures for sex, gender identity, and sexual orientation. The Committee evaluated existing approaches to measuring sexual orientation identity with regards to construct validity and item performance. We applied these scientific criteria in the context of a set of principles that emphasized respect for people's autonomy and rights to be represented in population-based data. A key area of debate for the committee and within the broader SOGI measurement field of scholarship is focused on which response options to provide. This presentation will discuss data supporting the final recommendation for how to measure sexual orientation identity, as well as areas in which more research is needed to further efforts to comprehensively and meaningfully represent the range of sexual identities in the U.S.

Measuring Gender Identity for the National Institutes of Health: Recommendations from the National Academy of Sciences, Engineering, and Medicine Consensus Committee Report

Nancy Bates, *US Census Bureau, Retired*

Among other studies, both the Institute of Medicine and the National Academy of Sciences have produced reports concluding that gender minorities experience greater health and economic well-being disparities, patterns of stigma, victimization and violence, lifetime suicide attempts, and substance use disorders compared to non-sexual minorities (IOM, 2011; NASEM, 2020). However, the routine collection of gender identity to monitor such trends (in surveys and other data collections) is scarce. Additionally, there is a lack of formally recognized standards for how to measure the construct of gender identity. To improve the quality of data collection efforts around both sexual and gender minorities, in 2020, the National Institute of Health sponsored the National Academy of Sciences' Committee on Measuring Sex, Gender Identity, and Sexual Orientation to review and recommend measures for sex, gender identity, and sexual orientation in three contexts: surveys and research, administrative, and medical or clinical settings. Specifically, we document the current measures used to assess gender identity, the pros and cons associated with each, and the empirical evidence of their validity as measures in order to develop recommendations on which measure(s) to use. We also discuss specific measurement considerations that arise within each of the three data collection contexts. We conclude with an outline of future research questions that will need to be addressed in order to ensure that the questions used to measure gender identity continue to meet content, cognitive and usability standards that are

necessary to ensure minimal measurement error as the terminology and social and political landscape around gender identity continues to evolve.

Developing Measures of Intersex Status/Differences in Sex Development at the Population Level
Tara Becker, *National Academies of Sciences, Engineering, and Medicine*
Katharine Dalke, *Pennsylvania State University*

The history of intersex persons—those with differences in sex development (Intersex/DSD)—in the United States has largely been one of invisibility and erasure. Though this population faces unique challenges, there have been almost no efforts to document and measure at the population level the social, economic, and health-related disparities and inequities they experience. This absence of information is driven by the underdevelopment of survey measures that can be used to identify this population. Though several measures have been proposed to identify Intersex/DSD populations, to date, none have been tested and validated within both Intersex/DSD populations and the general population. The National Academies convened the Committee on Measuring Sex, Gender Identity, and Sexual Orientation to review the research literature in the U.S. and other English-speaking countries to identify and assess the full range of measures that have been proposed and tested for use in identifying Intersex/DSD status. For each measure, we evaluate the complexity of the question, the populations in which it has been tested, and the results of standard methods of question validation, such as cognitive interviewing and item nonresponse. We then identify the most promising measure(s) that can be used to identify Intersex/DSD populations. We further note that Intersex/DSD status can be considered sensitive health information and provide guidance for when such data collections are appropriate and the care that data administrators should use in protecting this population from disclosure of this information.

Implementing Affirming Sexual and Gender Minority Measures: Making the NASEM Report "Measuring Sex, Gender Identity, and Sexual Orientation" Actionable
Christina Dragon, *The National Institutes of Health*
Karen L Parker, *National Institutes of Health*

Visibility is integral to improving health outcomes and well-being. A critical part of visibility for sexual and gender minorities is representation and identification in demographic data which relies on effective measurement and reliable, and community accepted, constructs. The National Institutes of Health-commissioned NASEM Consensus Panel Report provides a roadmap for the NIH Sexual and Gender Minority Research Office to advise and direct NIH Institutes, Centers, and Offices, along with other federal agencies and partners on measures of sex as a non-binary construct, gender identity, and sexual orientation in surveys and research studies, in administrative settings, and in clinical settings. This presentation will highlight the communications and implementation plans developed for gaining buy-in for and encouraging adoption of the NASEM recommendations. The communications and implementation plans will include tools for different types of organizations and settings to start utilizing the appropriate sexual and gender minority measurement constructs

Welcome to the Metaverse: Using the Web for Questionnaire Pretesting

Can a Web-Based Pretest be as Informative as a Telephone-Based Pretest? A Comparison of Modes in a Pretest of Survey Questions for U.S. Employers
David Kashihara, *Agency for Healthcare Research & Quality*

In an effort to reduce costs and canvass more employers, the Medical Expenditure Panel Survey - Insurance Component (MEPS-IC) performed a pretest of survey questions in 2021 using both telephone interviews and a self-reporting web survey. In the past, questions for the MEPS-IC were tested only using telephone interviews. However, because the primary mode of the MEPS-IC is web and also because of the potential cost benefits of using the web, it is of great interest to see if a web pretest of U.S. employers could produce similar quality information as a telephone pretest. The question topics in this pretest cover the effects of the COVID-19 health pandemic on employment, COVID-19 vaccination incentives offered by employers, employment composition characteristics, and telemedicine. Probes were developed to explore issues such as the interpretation of terms, the level of subject understanding, and the difficulty of providing a response. The pretest was designed to obtain 50 telephone interviews and 100 self-reported web responses from

employers that had responded to the 2020 MEPS-IC survey. This presentation will compare response and data quality characteristics between the two modes of pretest participants. Response will be analyzed based on the difficulty to obtain participation for each mode - in particular, how many sample case attempts were neceesary to reach the desired response quotas by size strata and type of employer. Data quality will be assessed based on the item nonresponse rate, including unanswered probes, and examining the quality of responses to the open-ended questions and probes. The findings from this unique study will inform the MEPS-IC and other surveys of employers about the benefits and the pitfalls that may be encountered when conducting a pretest using a self-reported web survey, compared to those of a telephone interview, to test survey questions

Can Crowdsourcing Replace Cognitive Interviews?
Christine Bottini, *U.S. Census Bureau*
Beth Newman Satisky, *U.S. Census Bureau*
Rachel Horwitz, *U.S. Census Bureau*
Rachel Sloan, *U.S. Census Bureau*

When project timelines are often tight and funding limited, a tool like crowdsourcing offers many benefits. Whereas traditional face-to-face cognitive interviews restrict sampling coverage to areas where staff are located, crowdsourced sampling allows access to a much wider population, including rural and other hard to reach areas. Likewise, where the number of participants in traditional cognitive interviews is generally small and determined by time and resources available, crowdsourcing allows for a large number of participants at low cost, to better reach saturation. While other research on crowdsourcing has looked at the quality of the data received on different crowdsourcing platforms, this study determines whether crowdsourced cognitive interviews yield the same conclusions as traditional cognitive interviews. Two independent analysts compare conclusions from in-person cognitive interviews conducted on questions from a national survey to conclusions drawn from crowdsourced cognitive interviews using the same questions from the national survey along with similar probes that were adapted to a self-administered mode. The current study will help determine how and whether crowdsourcing can be used in place of cognitive interviews. For example, if both types of interviews draw similar conclusions, crowdsourced cognitive interviews could be used in place of in-person interviews in instances where time or resources are scarce or as a final check following changes from the last round of in-person interviews. Alternatively, if the crowdsourced interviews lead to different findings, concurrent use may be best to get a complete picture of potential issues.

What's Missing? Incorporating Unmoderated Pretesting Techniques to Inform Survey Design
Temika Holland, *U.S. Census Bureau*
Elizabeth May Nichols, *U.S. Census Bureau*
Jonathan Katz, *U.S. Census Bureau*

While the Coronavirus pandemic has led to an increase in the use of technology to facilitate remote pretesting, other technology-driven techniques have been used to support survey research. For instance, survey researchers at the U.S. Census Bureau have been incorporating technology-enabled unmoderated testing techniques into cognitive and usability evaluations to help inform survey design.   Unmoderated testing typically consists of a participant interacting with a product (i.e., survey) alone, without the presence of a researcher/ interviewer. In recent years, the Census Bureau has incorporated web probes about the functionality and/or content in self-administered online surveys. This type of unmoderated testing technique, allows for data about reporting experiences to be obtained from a larger, more diverse sample. Additionally, unmoderated tests may more closely simulate a respondent's interaction with a survey instrument and eliminate interviewer effects that moderated testing may present. While unmoderated testing may be beneficial and present a cost-effective solution in many cases,some researchers caution against the use of unmoderated testing as a replacement and recommend it be used to supplement moderated testing techniques. This presentation summarizes recent cognitive and usability evaluations that incorporated both moderated and unmoderated testing techniques for demographic and establishment surveys. After briefly comparing findings from each methodology, this presentation will outline recommendations for implementation, highlight special considerations for incorporating such methods, provide guidelines for the testing protocol and the recruitment of participants, and share some examples of how these data were analyzed and summarized to inform survey design decisions.

Which Type of Probe to Use? Comparing the Data Quality of Closed- and Open-Ended Web Probes
Paul Scanlon, *Centers for Disease Control*

The inclusion of structured cognitive probes in survey questionnaires, often referred to as "web probing," is a reliable and flexible question evaluation method that gives researchers the ability to conduct mixed method studies within fielded surveys. Two major variants of web probing exist: open-ended probes use text fields, while close-ended probes provide a structured list of answer categories. The former can be used to collect qualitative information, akin to the data collected in cognitive interviews. The latter, on the other hand, is ideally used to collect quantitative information on the prevalence of patterns of interpretation or response that have emerged from previous qualitative work. However, few studies have explicitly compared the data quality of the two types of probes. This presentation will present the findings from a series of six case studies and experiments comparing open-ended and close-ended web probes from across four rounds of NCHS' Research and Development Survey. These case studies show that, when designed using previously collected qualitative data, close-ended web probes are generally able to collect similar ranges of patterns of response as open-ended probes do. On the other hand, when close-ended probes were designed without the extensive knowledge that previous qualitative studies provide, open-ended probes tend to capture a wider set of interpretations than the close-ended ones. Furthermore, a series of three other experiments explored how providing respondents with a "other, specify" open-text option as the final answer category in close ended probes affected how respondents answered the probes. The analysis of this set of experiments indicates that the "other" category does not substantially add to the list of patterns of interpretation in cases when the close-ended probe was designed using previously collected qualitative data. This presention will discuss the best practices regarding which type of probe to use that stem from these findings

Creating & Maintaining Diverse & Inclusive Survey Panels: Challenges & Strategies
Equity, Diversity, and Inclusion in Survey Panels
Alissa Stollwerk, *YouGov*
E.D. Bello-Pardo, *YouGov*

Ensuring equity, diversity, and inclusion in survey research is a pressing concern for researchers and practitioners alike. As the populations we survey grow more diverse across various measurements, researchers face challenges to ensure that our methods accurately represent this increasing diversity and that our products respect panelists' lived experiences. Practitioners, on the other hand, are growing increasingly attuned to the need to work with survey panels that are taking careful steps to work toward equity, diversity, and inclusion in their research. At the same time, survey panels like YouGov face various recruitment and retention challenges that make working toward these goals a complicated, yet crucial, endeavor. In this session, we will discuss the various ways through which YouGov has tackled these dual challenges. On the recruitment side of things, we will briefly discuss how we can approach the recruitment of diverse populations to survey panels like YouGov's. We will also discuss how YouGov's sampling methodology works toward creating demographically-balanced surveys. Finally, we will spend most of the talk's time focusing on asking sensitive survey questions in panels like YouGov's and how it is important for researchers to consider how their questions and in-survey experience could influence long-term retention of panelists from diverse backgrounds.

How Does AmeriSpeak Panel Recruit, Engage, and Retain Diverse and Hard-to-Recruit Segments of the U.S. Population?
Ipek Bilgen, *NORC at the University of Chicago*
David Dutwin, *NORC at the University of Chicago*

In this presentation, we will discuss how AmeriSpeak ® (NORC's probability-based panel) addresses challenges many probability-based online panels face with recruitment, engagement, and retention of various hard to reach segments of the population to ensure diversity, inclusion, and equity. We will present both long-term and recent strategies to include, retain, and engage hard-to-recruit/retain segments of the U.S. population. As a part of AmeriSpeak's long term strategy to improve representation and diversity, we will talk about our nonresponse follow-up (NRFU) program during panel recruitment, implemented via enhanced respondent incentives and face-to-face interviewing, and its impact on sample representativeness and survey estimates. Specifically, the NRFU program improves the representation in

household surveys among segments that are typically under-represented in commercially available probability-based panels: young adults, renters, Spanish-language dominant Hispanics, persons with less than a high school degree. Also, we will discuss our sampling strategies to improve coverage and combat nonresponse as a part of AmeriSpeak's long term strategy to improve representation. Lastly, we will evaluate 2021 recruitment and retention efforts including: Prioritization of hard-to-recruit segments using big data modeling, naming the person in the household on the letters, providing enhanced incentives to African American and Spanish speaking households in later waves during 2021 panel recruitment, engagement surveys and communication materials tailored for hard to retain segments of the panel such as Hispanic households and teen panelists.

Ensuring Diversity, Equity & Inclusion in Non-Probability Panels
Nicole Mitchell, *Dynata*

Dynata has over 25 years of experience in building and managing global online panels, with a current scope of 63 million online consumer and business respondents across 90 countries, serving 6000 clients across multiple industries (e.g., market research, media agencies, consulting/investment firms, publishers, corporate) and regions. Our goal is to create panels which provide the respondent volumes and diverse types of samples our clients require for their research; they are not built to be representative of a population, but rather that a nationally representative sample or another needed target can be extracted from the panel. As such, the panel is built to deliver diversity within the US samples we provide. As non-probability panels are built differently than probability panels, diversity within this context also needs to be thought about differently. Providing an inclusive sample can be difficult if there is not an understanding of how to recruit and retain diverse groups. Online coverage, types of sources, recruitment contact methods, messaging to respondents, and the offered incentives/rewards need to encourage maximum diversity among respondents. This includes cultural, racial, religious, age, biological sex, gender identity, sexual orientation, and disability diversity, among others.   However, there are challenges when ensuring inclusion of under-represented groups in panel recruitment, retention, and in survey participation. For example, response probabilities for young males are very low versus other demographics, and unacculturated Hispanics may be online but not reachable via English language recruitment and surveys. Measurement of certain types of diversity can also be difficult.   In this presentation, I will discuss why diverse recruitment and retention is important and how Dynata approaches diversity in non-probability panels (such as recruitment inclusivity through targeting specific audiences on social media sites, or measuring inclusivity through profiling panelists – e.g., disabled veteran status), and best practices for researchers when engaging with diverse populations.

Creating and Maintaining Diverse and Inclusive Survey Panels: Challenges and Strategies
Thomas Guterbock, *University of Virginia*
Heidi E. Grunwald, *Temple University*
Kara Shaner Fitzgibbon, *University of Virginia*

While all survey researchers must contend with diversity, equity and inclusion within their work, from sampling design and questionnaire development to analysis and reporting, there are additional unique DEI considerations in the context of running survey panels. This session will explore the challenges and strategies faced in recruiting a diverse and representative panel, maintaining high engagement across panel segments, and exercising inclusive and equitable practices in interactions with and compensation of panelists. This session includes four panelists, collectively representing commercial and academic, probability and non-probability, and global, national, and regional panels. To address the various DEI considerations in the context of survey panels, the presentations will focus on: innovative strategies, including a nonresponse follow-up (NRFU) program, used to recruit, retain, and engage hard-to-reach segments of the U.S. population (AmeriSpeak); recruitment contact strategies, such as audience-specific targeting on social media, as well as guidance on measuring inclusivity within panelist profiles, and general best research practices when engaging with diverse populations (Dynata); multi-mode recruitment strategies and differential incentive protocols to ensure an engaged and diverse respondent pool while limiting non-response bias (BeHeardPhilly); recruitment and sampling methodologies employed to create demographically-balanced surveys and respectful questionnaire design aimed to improve and honor the in-survey experience of panelists from diverse backgrounds (YouGov). Tom Guterbock, Academic Director at the University of Virginia's Center for Survey Research, will serve as

moderator and Heidi Grunwald, Director of Temple University's Institute for Survey Research, will serve as a discussant to guide conversation among panelists following their brief presentation.

Statistical Models & Algorithms to Address Missing Information

The Geography of Nonresponse. Can Spatial Econometric Techniques Improve Survey Weights for Nonresponse?
Christoph Zangger, *University of Bern (Official)*

Different strategies address unit nonresponse in cross-sectional and longitudinal surveys, with calibration and inverse probability weighting as some of the most common approaches (Särndal and Lundström, 2005). Moreover, it has been recognized that nonresponse varies geographically (Hansen et al., 2007). The geographic clustering of survey nonresponse has helped to identify segments of the population that are less likely to participate (Bates and Mulry, 2011; Erdman and Bates, 2017). As a consequence, researchers have included geographically aggregated measures to account for nonresponse and to construct survey weights (Kreuter et al., 2010). This paper extends this literature by building on the argument that people with similar characteristics tend to live in similar places. The resulting segregation induces a spatial correlation among characteristics which are also used to predict survey nonresponse, for example, education, and that are likely correlated with other measures in surveys. Consequently, the residuals of regressing survey response on a set of available characteristics are themselves spatially correlated, biasing estimates and predictions (Pace and LeSage, 2010). While aggregated characteristics can pick up some of the spatial correlation, there is another, more direct approach that accounts for spatial correlation: spatial econometric models (LeSage and Pace, 2009). These models can directly incorporate other units' response status in the prediction of an individual unit's response propensity, accounting therewith for the socio-spatial interdependence induced by unobserved residential selection. Using Monte Carlo simulations, this paper demonstrates how spatial econometric models improve predicted response propensities, yield more accurate survey weights, and are thus a valid alternative to common inverse-probability weighting approaches, even if the data generating process is incorrectly specified. The results are robust across a wide variety of model specifications, including the underlying response pattern and its spatial correlates.

Transporting Recurrent Neural Networks for Coding Interviewer Question-Asking Behaviors Across Surveys
Jerry Timbrook, *RTI International*

Survey researchers commonly use behavior coding to identify whether interviewers read survey questions exactly as worded in the questionnaire or not (a potential source of interviewer variance). However, manual human behavior coding can be expensive and time-consuming. Recurrent Neural Networks (RNNs) have been used to partially automate this process to save time and money, and have demonstrated reliability comparable to humans when coding behaviors on questions in the same survey (Timbrook and Eck 2019). However, these RNNs were trained under a particular set of essential survey conditions (e.g., a particular organization fielded the survey during a specific time frame using a particular group of interviewers). It is unknown if trained RNNs can be used to code question-asking behaviors on the same questions administered under different essential survey conditions. In this paper, I train RNNs using human-coded transcripts of interviewer question-asking behaviors (whether interviewers asked questions exactly as worded) from 13 questions in the Work and Leisure Today II telephone survey (WLT2; AAPOR RR3=7.8%). I then use these trained RNNs to code the same question-asking behaviors on the same 13 questions administered in a different telephone survey (Work and Leisure Today 1; WLT1; AAPOR RR3=6.3%). All WLT1 question-asking behaviors were also coded by undergraduates (a "typical" human-coded dataset) and master-coded by graduate research assistants ("ground truth") to evaluate inter-coder reliability (kappa). I compare the reliability of RNN coding on WLT1 versus the master coders to the reliability of human undergraduate coding on WLT1 versus the master coders. Preliminary results indicate that for a question asking about the number of individuals in a respondent's household, an RNN trained on WLT2 data coded WLT1 question-asking behaviors with reliability equal to the human undergraduate coders. I conclude with the implications of using RNNs for behavior coding across different telephone surveys.

First-Name Information Particularly Improves the Accuracy of Racial-and-Ethnic Imputations for Females
Marc Elliott, *RAND Corporation*
Ann Haas, *RAND Corporation*

John Adams, *Kaiser Permanente Center for Effectiveness & Safety Research and Kaiser Permanente Bernard J. Tyson School of Medicine*
Amelia Haviland, *Carnegie Mellon University and RAND Corporation*
Jacob Dembosky, *RAND Corporation*
Peter A. Morrison, *Morrison & Associates, Inc.*
Sarah Gaillot, *Centers for Medicare & Medicaid Services*
Jennifer Gildner, RAND Corporation
Loida Tamayo, Centers for Medicare & Medicaid Services

Race-and-ethnicity data on sampling frames enable targeted or stratified sampling and evaluation of non-response by race-and-ethnicity. Such data also promote more accurate stratified reporting by race-and-ethnicity. In many cases, administrative records used for sampling may have incomplete, imperfect, or missing racial-and-ethnic data. One solution is to impute race-and-ethnicity using available administrative data. Surname has long been used; although less predictive, first-name information has been found to improve imputations that already include surname information.   We compared the marginal contribution of first-name information to the accuracy of racial-and-ethnic imputations in a sample of Medicare beneficiaries. We analyzed two scenarios: a) a sparse set of predictors (full name, address, and sex) and b) a rich set of predictors (adding an imperfect measure of race-and-ethnicity, demographics, and Medicare coverage characteristics). We assessed whether gains in accuracy from adding first names as predictors differ by sex and race-and-ethnicity.   We found that in the absence of other predictors, surnames are more predictive of race-and-ethnicity for males than females among non-Hispanic White, Hispanic, and Asian/Pacific Islander beneficiaries. For these three groups, the additional information provided by first names under both sparse- and rich-predictor scenarios improves accuracy more for females than males, but in most cases accuracy remains slightly higher for males. In contrast, both surname and first-name are similarly predictive for males and females among Black beneficiaries, and gains in accuracy when first names are added to imputations are similar by sex. Predictions of Black race-and-ethnicity either excluding or including first-name information have similar accuracy for each sex. For all groups, the addition of first-name information improves prediction accuracy more under the sparse-predictor scenario than under the rich-predictor scenario. Thus, first-name information increases the accuracy of racial-and-ethnic imputations, especially when there is only a sparse set of predictors and especially for females.

Applying Modern Machine Algorithms in Income Imputation of Survey Data

Zheyu Jiang, *UCLA Center for Health Policy Research*
YuChing Yang, *UCLA Center for Health Policy Research*
Todd Hughes, *UCLA Center for Health Policy Research*
Royce Park, *UCLA Center for Health Policy Research*
Ninez A. Ponce, *UCLA Center for Health Policy Research*

Missingness in income information is common in data from many surveys, such as household income in the California Health Interview Survey (CHIS). In CHIS, missing household income has been imputed using the hot-deck method within ranges defined by a set of auxiliary variables such as bracketed income range and/or poverty level. In the past, the hot-deck method relied on traditional regression modeling to impute the missing income of CHIS respondents. In this presentation, we will explore a series of modern machine learning algorithms, including neural network, support vector machine, decision tree etc., to impute the missing income information. We will also compare the performance of each algorithm with the regression model currently used in CHIS in terms of accuracy, running time, and ease of model convergence, etc. The results show the great potential of applying machine learning techniques in population survey data imputation, allowing the use of more variables and more accurate predictions than regression models, with less concerns of model convergence and overfitting issues.

Extending Our Analytic Reach: Multiple Regression for Partial Data

Xiaolu Yan, *Ipsos - Seattle, WA*
Randall K. Thomas, *Ipsos Public Affairs*
Frances M. Barlas, *Ipsos Public Affairs*

Multiple regression is an easy-to-understand analytic tool that can sort through the relative importance of factors predicting key variables. Available multiple regression software packages (e.g., SAS, SPSS) typically delete all cases that have any missing data for predictor or dependent variables. As a result, participants in such studies are required to complete all possible items, often leading to a large respondent survey burden. One way to reduce respondent burden would be to have respondents complete a random subset of questions, such as randomly assigning to any given respondent 10 items out of a possible 20 items. Such fractional or planned missingness designs allow every pair of variables to have an equal number of completes. This can significantly reduce respondent burden, improve data quality, and produces data missingness that is completely at random. In a prior paper, we examined using partial designs and computing key coefficients for two predictor variables. In this paper, we examined a study that had over 9,000 completes where respondents completed every item. First, we randomly created missingness in every predictor variable at varying levels (e.g., 10%, 20%, etc.). Using repeated sampling, we generated correlation matrices for each level of missingness and using different sampling sizes (100, 250, 500). This allowed us to calculate multiple regression coefficients and develop their associated confidence intervals, using the empirical distributions for the coefficients. In this study, we focused on the extensibility of our new procedure to produce coefficients for three and four predictor variables. We show that such fractional designs are relatively easy to implement, reduce respondent burden, and we can easily derive coefficients (e.g., betas, b-values, semi-partial coefficients) with estimated variance around the coefficients that enables significance testing. We also review some limitations on sample size and robustness to violations of data missing at random.

Gender & LGBTQI Roles & Attitudes

Key Predictors of Attitudes Toward Lesbians and Gay Men in Singapore

Benjamin Detenber, *Nanyang Technological University*
Ring Joyce Pang Shu Min, *Nanyang Technological University*
Teo Min Yu, *National University of Singapore*
Feng Siqi, *Nanyang Technological University*

While generally more socially conservative than Western countries, Singapore has seen greater public discussion of homosexuality and debates about equal rights for non-heterosexual people. A nationwide household survey was conducted to assess public opinion related to lesbians and gay men in the Southeast Asian country in the third quarter of 2020. At the time, the initial lockdown for Covid-19 had ended and the delta variant had not yet spread to Singapore, so it was possible collect data following the government and university's health and safety measures. Multi-stage cluster sampling was combined with simple random sampling and late-stage quota filling to reach individuals at housing units across the island. The mode was primarily self-administered computer-assisted personal interviews (CAPI), but interviewer-conducted when requested. Questionnaires were administered to citizens and permanent residents aged 21 and above were in English, Mandarin or Malay. Numerous quality control measures were put in place (e.g., speed checks, screen capture recording, quality check callbacks, etc.), and after screening, the final sample was N = 1158 and the response rate was 51.7% (AAPOR RR Formula 3). The primary criterion variables were attitude toward lesbians and gay men (ATLG; Herek, 1984) and acceptance of lesbians and gay men (adapted from Hong Kong Home Affairs Bureau survey, 2006). A wide range of predictor variables were measured, including demographics (gender, age, education, income, and religion), intrinsic religiosity (Gorsuch & McPherson, 1989), and fundamentalism (Altmeyer & Hunsberger, 1982). Other predictors in the survey were personal values (e.g., filial piety and universalism), belief that homosexuality is a choice (Horvath & Ryan, 2003), and the quantity and quality of contact (Islam & Hewstone, 1993) one had with lesbians and gay men. The study was pre-registered on OSF.io. Analyses are underway and will be completed in the first quarter of 2022.

Male Sports Participation and Beliefs of Male-Dominance: A Cross-National Analysis on Gender Ideologies

Simon Lütkewitte, *Universit _ t Bielefeld*

Male Sports Participation and Beliefs of Male-Dominance: A Cross-National Analysis on Gender Ideologies   Gender equality is a cornerstone of liberal democracy and a frequent subject to research in the social sciences. While research shows progress in many Western societies over the past decades, it is also shows notable segregation by gender in several life domains. A prominent example of continuing segregation by gender is sports participation. Already in the 1980's, Messner argued that sports participation in males preserves and reinforces beliefs of a "natural superiority" over females. Furthermore, it also provides a separated cultural sphere that retreats from the growing gender equality in society and as cultural sphere that cultivates gender ideologies. These gender ideologies in turn help to legitimize gender inequalities in society. Despite an extensive scholarly and public discourse, empirical research remains surprisingly scarce. The proposed research aims at identifying the effects of sports participation on gender ideologies in males. The research draws on cross-national survey data from the World Value Survey and the European Values Study of the years 2006 to 2018 including 48911 males from 54 countries. The use of these two data sources offers the advantage that the male respondents studied are socialized in different societies that differ in terms of gender equality, religiosity, modernization or economic development. Furthermore, both data sets provide comprehensive measurements on gender ideologies. Multilevel mixed response models of the cross-national survey data show that sports participation in males indeed slightly fosters gender ideologies in males. Moreover, it is analyzed whether the effect from sports participation varies across contextual and individual level characteristics. For instance, I find that the educational level as well as the age are moderating factors

Understanding Public Support for Policies Aimed at Gender Parity in Politics: A Cross-National Experimental Study

Timothy Gravelle, *Momentive*
Andrea Carson, *La Trobe University*

Victoria Bustamante Rodriguez, *Momentive*
Lía Acosta Rueda, *University of Melbourne*
Leah Ruppanner, *University of Melbourne*

In spite of decades of attention to the issue, women remain underrepresented in national parliaments. There are larger gender gaps in the legislatures of democracies that have not legislated gender quotas than those that have. Several of the Anglo-American democracies fare poorly in rankings for women's electoral representation, in particular Australia (ranked 57), Canada (59) and the United States (72). Drawing on role strain theory and large-scale, original online survey experiments conducted in Australia (n = 8,413), Canada (n = 9,127) and the United States (n = 8,454), this paper seeks to understand public opinion on non-quota mechanisms that could help attract and retain women (and men) politicians. We construct a vignette describing a hypothetical politician's success in politics, but also their personal challenges with meeting the demands of work and family. The experimental design manipulates both the gender of the politician and level of government (federal or local) in the vignettes. All respondents are also asked follow-up questions about what types of assistance, if any, should be extended to the hypothetical politician to help her/him continue in her/his elected role. We find public support for policies aimed at lessening work–family role strain is higher for a hypothetical woman politician; these include a pay raise and childcare and housework allowances. Respondents were less likely to say a woman politician should change careers or have her spouse exit employment. Still, respondents found a woman less electable (compared to a man) if the public learned she was having challenges balancing work and family. Our results point to important cross-national differences in levels of support, which inform our theoretical models. The study's findings contribute to literature on gender equality and non-quota mechanisms designed to mitigate the gender gap in politics.

The Power of (Observing) the Patriarchy

Tresa Undem, *PerryUndem Research/Communication*

In our research, people struggle to identify ways in which sexism or the patriarchy affects their personal lives. Many link sexism with concepts like sexual harassment and equal pay. Yet, Pew suggests just 25% of working women say they've experienced pay inequality. In a recent PerryUndem survey, just 35% of women say they'd ever felt discriminated against or treated unfairly because of their gender. Such findings led us to a question: How do we solve a problem most of us don't see? And that question sparked a research project: What happens when you recruit 42 diverse people to become observers of their lives over the course of 21 days? What happens when you ask them to take note of how gender and race show up at work? At home? On social media? In their family system? In music? On TV? In spirituality and religion? In health care? We used Schlesinger's Over The Shoulder mobile ethnography app to help real people become ethnographers of their own lives. What we found was, frankly, stunning. Some participants started the research expressing gratitude for gender equality before transitioning to confusion, fear, and at times anger at several biases they started noticing in their everyday life. One participant said: "This [project] has been hard. It forces me to think and admit things that I pretend don't exist. Forces me to look at my own family, my own life, where my future will head. Its challenging and scary." Many participants identified several biases in all areas of their lives - biases which aren't captured in survey research. This is one of our all-time favorite research projects. We'll present some of the most interesting and powerful insights - to show you why.

Loss of Income, Career Sacrifice and Overtime for an LGBTQI*-Friendly Work Environment? A Choice Experiment to Investigate Employment Prefereces of LGBTQI* People

Lisa de Vries, *Universitaet Bielefeld*
Zaza Zindel, *Bielefeld University*

Social acceptance of LGBTQI* (lesbian, gay, bisexual, trans*, queer, and inter*) people has improved greatly in many European countries in recent decades, and equal legal treatment in the labor market has been improved. Still, a not inconsiderable amount of LGBTQI* people report experiencing discrimination in their work lives. Several studies suggest that significant differences in labor market outcomes (e.g., earnings) and tendencies toward occupational segregation, exist by sexual orientation and gender identity. Until today, it is unclear to what extent labor market decisions of LGBTQI* people are driven by their search for a non-discriminatory working environment and how far they sacrifice career opportunities to avoid discrimination. To answer these questions, we designed a discrete choice experiment using fictitious job descriptions and integrated it into a German online survey with the target group of LGBTQI*-people in 2021

(N=7,607). Linked to our research questions, we plan to contrast general job attributes with different LGBTQI* friendly work environment measures. Using utility theory, implementing a discrete choice experiment offers us the opportunity to measure the strength of job attribute preferences and calculate trade-offs toward different attributes. Based on the results, we can calculate to what extent LGBTQI* people are willing to sacrifice income, promotion prospects, and working time for an LGBTQI* friendly work environment. Further, the great number of aspired respondents and a high number of control variables allow us to differentiate between separate groups (e.g., sexual minorities, trans* people, etc.) and, for example, industry sector or occupational status.

Beyond Boy or Girl. How Do Teenagers Conceptualize Gender?
Meredith Massey, *NCHS*

Up until recently, gender has been measured as a male/female binary. However, the American public's conceptualization of gender has shifted dramatically in the last decade and the traditional way of asking about gender on surveys is no longer adequate. In order to understand how adolescents conceptualize gender, the Collaborating Center for Questionnaire Design and Evaluation Research at the National Center for Health Statistics conducted 60 cognitive interviews with youth ages 12-18. These cognitive interviews tested the gender question as part of two separate studies testing questions for the Youth Risk Behavior Surveillance System and a set of questions on teens and work for the National Institute for Occupational Safety. This presentation examines the results of these cognitive interviews and details how adolescents have a very different understanding of gender than both the traditional strictly binary understanding of gender and adults' increasingly broad understanding.

New Perspectives on Health-Related Survey Research: COVID-19 & Beyond
Using Survey Research During the Pandemic to Explore Mindsets, Experiences, and Values that Influence Health
Carolyn Miller, *Robert Wood Johnson Foundation*

As our understanding of what drives health expands and we focus more on racial, social, and economic inequities and their impact on health, we face challenges in measuring these complex and interrelated social issues, obtaining adequate sample sizes for subpopulations of interest, and in making findings actionable in advancing a population health and well-being agenda. In support of the Robert Wood Johnson Foundation's (RWJF) vision to build a Culture of Health, in which everyone has a fair and just opportunity for health and well-being, this presentation will highlight select findings and methodological/research design issues from four survey efforts supported by the Foundation prior to, and during, the COVID-19 pandemic: 1)   American Health Values Survey (NORC, 2016 n=10,574; 2020 n=8261) 2)   Impact and Experience of COVID-19 for Populations at Greater Risk Survey (RAND, summer 2020 through fall 2021, 4 waves, WV1 n=5164; WV2 n=4143; WV3 n=4031; WV n=3583) 3)   Impact of COVID on Households in the US (Harvard/NPR/RWJF, summer 2020 n=3454, fall 2021 n=3616) 4)   Views on the US Public Health System (Harvard, 2021; n=1305) These surveys (all probability-based samples of US adults) focus on measuring public opinion around complex social issues such as individual health values and beliefs, views on the role of government in health, the impact of the pandemic on households across the country, and views of the US Public Health System. Using these unique data, we are exploring mindsets that are fundamental to building a Culture of Health (e.g., deservingness, equity, racial justice, and opportunity) but are inherently difficult to measure and slow to change. The findings from these surveys, and all the survey work presented on this panel, are helping to map a new landscape of public opinion on a wide range of topics fundamental to collective action toward population health, well-being and equity.

Black Americans' Responses to and Experiences with COVID-19
Carolyn Funk, *Pew Research Center*

Understanding what Americans think about health and medical issues took on heightened importance in the midst of the coronavirus pandemic. Early reports showed that Black Americans' often faced higher risks from the coronavirus and they were disproportionately impacted by the disease. Drawing on Pew Research Center surveys collected over the course of the outbreak, I will highlight measures developed to better understand Black adults' concerns about and efforts to protect themselves from COVID-19 as well as their personal experiences with the disease. And, I will look at Black Americans' views related to coronavirus vaccines as a window into how health-related polling can lend insight

into people's views of the medical care system. I will draw on key findings from a new Center survey conducted with a large oversample of Black adults to shed light on how people's views are often tied to broader equity concerns rooted in their experiences of health and medical care in the U.S. and to address the diversity of viewpoints among Black Americans by gender, age, education and other factors.

New Perspectives on Health-Related Survey Research: COVID-19 and Beyond
Jonathon Schuldt, *Roper Center for Public Opinion Research*

The COVID-19 pandemic laid bare the complex and interconnected nature of health and social issues, and the need for public opinion researchers to adopt new approaches. This panel brings together leading voices from the worlds of polling, foundations, and academia to discuss new perspectives on health-related survey research during the pandemic and beyond. Cary Funk from the Pew Research Center will discuss the Center's efforts to better understand the coronavirus concerns and health behaviors of Black Americans, a group that has been disproportionately impacted by the pandemic, and how surveys reveal the ways that coronavirus concerns intersect with broader concerns about social equity. Liz Hamel, Mollyann Brodie, and Ashley Kirzinger from the Kaiser Family Foundation will discuss the innovative sampling strategies employed in the new KFF COVID-19 Vaccine Monitor, as well KFF's broader goals of tracking trends, oversampling members of disproportionately affected groups, and remaining methodologically flexible as the pandemic evolved. Carolyn Miller from the Robert Wood Johnson Foundation will discuss results from four large-scale survey projects supported by RWJF prior to and during the pandemic, allowing a unique picture into the public's changing attitudes and the Foundation's approach to addressing social disparities in health. Finally, Peter Enns, Bethany Mackey, and Kathleen Weldon from the Roper Center for Public Opinion Research at Cornell University will discuss Roper's new Health Poll Database, a collection of 70,000 health-related polling questions that allows users to search questions, visualize interconnected health topics, and plot over-time trends in unprecedented ways. Across presentations, the role of survey research in understanding and addressing racial, social, and economic disparities in health outcomes will be considered.

Pivoting in a Pandemic: Adapting an Ongoing Survey Research Program to Fill a Need
Liz Hamel, *Kaiser Family Foundation*
Mollyann Brodie, *KFF*
Ashley Kirzinger, *KFF*

KFF has served as a non-partisan source of information on the U.S. public's views and experiences with health care for more than 25 years. Before the pandemic, our most visible project was the KFF Health Tracking Poll, a (mostly) monthly series of dual-mode RDD telephone surveys that largely covered topics of health care access, affordability, and opinions on key health care debates and legislation including the Affordable Care Act. Early in the pandemic, we added questions about COVID-19 to our ongoing surveys to measure things like worries, behaviors, economic impacts, and views of the government response. In late 2020 as it became clear vaccines were on the horizon and surveys showed vaccine attitudes diverging by partisanship and race/ethnicity, we saw a need for a new ongoing survey effort to track the public's evolving attitudes and experiences with COVID-19 vaccines. We launched this new effort, the KFF COVID-19 Vaccine Monitor, in December 2020 and as of October 2021 have interviewed over 16,000 respondents. This presentation will discuss key substantive and methodological decisions made in adapting our survey methodology to achieve several goals, including 1) having a robust, probability-based methodology to track trends over time; 2) adequately representing the views and experiences of groups like Black and Hispanic adults who were disproportionately impacted by the pandemic; and 3) remaining flexible enough to respond to events on the ground. Our strategies included adding oversamples of Black and Hispanic adults to our RDD surveys, as well as using online probability panels to supplement these surveys for populations like rural residents and parents. The presentation will also review some of the key findings from the Monitor and demonstrate their usefulness for informing the U.S. vaccination effort.

The Roper Center's Health Opinion Database: Making 70,000 Health Survey Questions Available
Kathleen Weldon, *Roper Center, Cornell University*
Bethany Mackey, *Roper Center, Cornell University*

The Roper Center for Public Opinion Research and the Robert Wood Johnson Foundation partnered on a four-year project to create the Health Opinion Database. This database makes responses to more than 70,000 survey questions related to health from more than 750 survey organizations available in a single location (https://www.healthpolldatabase.org). This presentation will highlight the key features of the Health Opinion Database, including the ability to find results via a keyword search and filters or by searching the node network, which shows links between health topics in the database. We will also demonstrate additional data visualization tools that highlight results from survey questions and allow users to plot over-time trends. We will conclude by showing how to find data and questions used in other presentations on this panel via the Health Opinion Database

Does Our Message Speak to You? Tests of Survey Communications Messaging

How to Increase Corporate Employee Mailer Engagement and Survey Participation. A Review of Data Collected and Best Practices

Jason Telner, *IBM*

Tyler Waite, *IBM*

Robert Medl, *IBM*

Jenny Weinberg, *IBM*

Mark Malesta, *IBM*

Corporate teams spend countless hours crafting corporate communications and survey invitations that will stand out in cluttered inboxes, yet research on effective strategies for increasing click rates for corporate electronic mailers (including surveys) in large enterprise organizations remains scarce. The corporate user research team conducted an exploratory analysis of 3,709 corporate electronic mailers and survey invitations in order to identify the most effective techniques for increasing electronic mailer click rates and survey participation. Statistical analyses were conducted on the click rates for electronic mailers, as well as survey link clicks inside of electronic mailers. Our analyses revealed significantly higher click rates for electronic mailers sent at noon relative to those sent in the morning or afternoon. Notification electronic mailers (i.e., successful account setup) had significantly higher click rates than electronic mailers designed to create awareness or request an employee action (i.e., compliance requests). Specifically for survey mailers, a significant effect of the sender was found, with higher survey link clicks occurring when the survey was sent by an executive rather than a generic email. Survey link clicks peaked during the height of COVID (March-April 2020) and decreased dramatically by March 2021 and throughout the remainder of 2021. Finally, an analysis of survey Net Promoter Scores was conducted to evaluate the effectiveness of survey reminder emails. The analysis suggested that Net Promoter Scores remain statistically unchanged even after including the Net Promoter Scores from reminders. Findings will be discussed in detail along with best practices to maximize engagement and responses when designing corporate electronic mailers or survey invitations.

Effect of Logo Visibility on Respondents' Confidence in Government Surveys

Lin Wang, *U.S. Census Bureau*

Elizabeth Nichols, *U.S. Census Bureau*

Alda Rivas, *U.S. Census Bureau*

Shelley Feuer, *U.S. Census Bureau*

It is a common practice to place a logo on a government survey as an indication of sponsoring organization's legitimacy. Nowadays, mobile surveys are a popular mode of data collection. Due to small smartphone screens, mobile survey designers have to make use of every bit of the screen to maximize data quality. A question is thus raised: If logo presence is reduced, how will it affect one's confidence in government surveys? In this paper, we report on an experimental study that investigated the effect of logo visibility on respondents' confidence in government surveys. Eighty-two participants were randomly assigned to one of three logo placement designs: (1) A logo on every survey screen, (2) A logo on the survey title screen only, and (3) No logo on any survey screen. Every participant read an invitation letter on paper and then completed a 6-question survey with the assigned logo on a smartphone. Upon completion of the survey, the participant was asked three questions regarding awareness of the survey organization, perception of the logo, and interpretation of the logo; then selected a preferred logo placement and provided a reason

for the selection. The study found that, (1) about one tenth of participants reported seeing a logo on the survey; (2) among those who saw the logo, about a half of the participants interpreted the logo as an indicator of a government survey or of not being a scam; (3) about two thirds of participants preferred for the logo to be displayed on every screen while less than a third preferred it on the title screen. The findings suggest that a logo on mobile survey is likely to be overlooked, however, it does somewhat increase respondents' confidence in a government survey.

Impact of Differential Appeals in an Advance Letter Mailing on Phone Recruitment for a Longitudinal Survey

Chrystine Tadler, *NORC at the University Of Chicago*
Megan Bjorgo, *NORC at the University of Chicago*

The Medicare Current Beneficiary Survey (MCBS) serves as the leading source of information on the Medicare program and health care costs for the Medicare population. 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. In response to the coronavirus (COVID-19) pandemic, data collection shifted from in-person to phone outreach and survey administration. This shift brought concerns of lower response rates typically seen with phone surveys, as well as incomplete data since phone numbers were not available on the sample frame. This placed increased emphasis on the importance of alternate forms of respondent outreach, such as mailings and respondent-driven toll-free calls The MCBS has long used an advance letter, sent to all sampled beneficiaries in advance of their first contact, to legitimize and notify them that an interviewer would soon be contacting them to request their participation in the survey. In Fall 2021, we implemented a split ballot experiment testing the effect of different appeals within the advance letter, specifically testing new language prompting panel members to call the survey toll-free number for more information and to schedule an interview. Of nearly 15,000 sample members, 7,742 were sent an advance letter with revised appeal, containing bolded text inviting beneficiaries to call the 800-line. The remaining 7,206 were sent the standard MCBS advance letter which tells beneficiaries to expect a call from an interviewer. This presentation will share detail on the revised appeal, impact on completion rates and timelines, and impact of the advance letter appeals on interviewer effort required to complete a case. Findings will help inform how advance materials can be leveraged to increase respondent cooperation and improve efficiency.

Impacts of Presurvey Messaging Content on Response Rates in Low-Income Countries

Steven Glazerman, *Innovations*
Andrew Dillon, *Northwestern University*
Dean Karlan, *Northwestern University*
Michael Rosenbaum, *Innovations for Poverty Action*

Survey methodologists have long experimented with different ways to improve response rates by pre-contacting sample members with messaging designed to predispose them to participation and cooperation. The methodological literature includes many studies of postcards, letters, endorsements, and prepaid incentives or gifts. For researchers working in low-income countries, however, the insights from this research have little relevance. Mailing gifts or letters is rarely feasible and the social and economic context for survey participation behavior is vastly different from that of high-income countries. Building on a randomized experiment that we implemented to test the use of SMS messages to improve response rates to RDD surveys (Dillon et al. 2021), we designed an experiment to test more variations in pre-survey message content for followup surveys in five countries: Burkina Faso, Colombia, Cote d'Ivoire, Rwanda, and Zambia. We designed messaging to appeal alternatively to salience, self-interest, or a combination of the two. To test how salience works, we randomized whether the pre-survey SMS included information learned from the first round of the survey and what form that took: general or specific. We also randomly varied whether the information was about food insecurity or unemployment, two of the survey's major areas of inquiry. To test how self-interest works, we cross-randomized a reminder about the monetary incentive. There were ten treatment arms in total. We examine the impact of these survey messages on several outcomes. The first set is successful contact, cooperation, and completion: Which type of message content most improves response rates? The second set of outcomes is sample composition: How does message content influence the representativeness of the sample? The third set of outcomes is the survey responses themselves. Does message content translate into different inferences from the study itself?

Establishing a Data Collection Strategy: How & Why Establishments Respond

The Effects of Mode and the COVID-19 Health Pandemic on Survey Response and Data Quality in a National Survey of U.S. Businesses

David Kashihara, *Agency for Healthcare Research & Quality*

In the Spring of 2020, many businesses across the United States were impacted by the COVID-19 health pandemic. A few types of businesses flourished while many others suffered greatly, with some businesses closing temporarily or permanently and many of those remaining open suffering losses of customers and revenue, causing them to shed employees. While these things were starting to occur, the 2020 Medical Expenditure Panel Survey - Insurance Component (MEPS-IC) was about to begin data collection operations and was facing pandemic-based issues itself, including the closure of telephone and survey processing facilities, interviewers not equipped to work remotely, and staff scrambling to develop alternative methods to survey businesses that had many more reasons not to respond. The COVID-19 health pandemic forced the large, nationally representative MEPS-IC survey to change its primary administration mode from mail to web. A recent MEPS-IC study found that smaller businesses generally preferred to respond by mail while larger businesses generally preferred to respond by web (Kashihara, 2019). In addition, another MEPS-IC study found that data quality was better for web respondents than for mail respondents (Kashihara, 2019). This analysis will focus on both of these past trends using 2020 MEPS-IC data to gauge the impact of the health pandemic and the primary mode change on survey response and data quality. This presentation will illustrate patterns of response characteristics by mode, firm size, and industry. Data quality will be assessed by examining edit failure and item nonresponse rates for selected survey questions. These findings will be compared to those from prior years to help measure the overall impact of the COVID-19 health pandemic on the survey data. Other surveys of U.S.businesses will find this information very useful as they, and the MEPS-IC, continue to collect data in the pandemic and post-pandemic environments.

Examining the Consistency of Responses Between a Quick Response Module and the Associated Full-Scale Data Collection

Caren Arbeit, *RTI International*
Michael Yamaner, *National Center for Science and Engineering Statistics at the National Science Foundation*

Every year, over 700 institutions provide information on science, engineering and health graduate student enrollment, postdocs and nonfaculty researchers to the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS), sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) and by the National Institutes of Health (NIH).At the end of the 2019 data collection in March of 2020, many coordinators struggled to complete the 2019 data collection in March of 2020 due to the COVID-19 pandemic. In fall of 2020, the GSS fielded a 26 question (including 5 open-ended text responses) COVID-19 Impact Module hosted within the main GSS web application. The module included items on the ability to complete the GSS survey, the impact on graduate student enrollment and funding, as well as the impact on postdocs and doctorate-holding nonfaculty researchers, including asking about hiring freezes. Several questions asked about the specific impact of COVID-19 on temporary visa holding students and postdocs. The module had an over 90% response rate, although many items had lower response rates. With full-scale data collection completed, we can compare the final response data from schools to their responses to the COVID-19 Impact Module. This presentation specifically focuses on the following questions: did coordinators who indicated that the survey would be more difficult in 2020 have higher item non-response rates on the base GSS? How accurate were the directional estimates provided by respondents? Were trends different for coordinators who responded 'don't know' or skipped an item than for coordinators who provided a valid response? The discussion will focus on the lessons learned about the alignment between the responses on the survey module compared to the detailed items in the full survey.

Survey-Based Network Analysis of Interagency Collaboration Mechanisms and Program Performance at the U.S. Government Accountability Office

David Dornisch, *U.S. GAO*

Frequently, in the U.S. GAO's work on interagency collaboration understanding the types of mechanisms created to support program performance is a key issue. GAO has developed a typology of collaboration mechanisms that has

been used in many of its reports. Examples include working groups, memos of understanding, interagency training, and joint information-sharing. GAO has established that their effective use can in turn have positive effects on program performance. One method that GAO has used to measure and evaluate their functioning is survey-based network analysis, which is a method that allows for a comprehensive understanding of the quality and quantity of collaborative relationships among all of the agencies or organizations involved in relevant program areas. This presentation will review recent examples, detailing results related to collaboration mechanisms and associated program performance, in particular: 1. Indian water infrastructure programs : GAO surveyed 42 representatives of federal agencies involved in the administration of programs designed to support the development and sustainment of water systems on Native American lands in 6 states. The survey asked respondents to answer a battery of questions on the particular drinking and wastewater activities that they work on with each other, the particular mechanisms that they use to do this, and their needs for the enhancement of that collaboration in the future. Based on this survey, GAO reported results relating to current and future collaboration needs. 2. Land-Mobile radio systems interoperability . GAO surveyed 60 representatives of federal agencies involved in coordinating emergency management efforts, focusing on questions related to the interoperability of land-mobile radio technologies across the agencies. GAO was able to report an array of results related to the integration, fragmentation, and effectiveness of the overall system of LMR technologies across the federal government

Barriers to Response to Annual Economic Surveys: Talking with Non-respondents
Melissa Cidade, *US Census Bureau*
Kelsey, *Drotning*

As in social surveys, establishment surveys are experiencing decreasing response rates. However, while a robust literature on non-response in social surveys is always being refined, non-response on establishment surveys receives periodic attention. This study seeks to affirm and update prior work on non-response in economic establishment surveys by interviewing representatives of 19 non-responding companies to one or more annual economic survey(s) fielded by the U.S. Census Bureau. These interviews reinforced that the response process – and barriers to that process – for establishment surveys are markedly different than those for social surveys; this paper outlines the factors that survey researchers can mitigate (including clear communications, timely updates, and more), and those that are outside of the control of the survey researcher (including firm structural issues and changes).

Does a Stronger Confidentiality Pledge Influence Respondents' Perceptions? Exploratory Research to Understand the Intersection of Privacy and Reporting Behaviors on Business Surveys
Gerson Morales, *US Dept of Energy/Energy Information Administration*

Higher trust in the government has been associated with higher levels of unit response rates and lower item non-response rates on business surveys (Childs, Fobia, King, & Morales, 2019). Unfortunately, according to the Pew Research Center (2021), public trust in government has been steadily declining, which can negatively influence responses in federal surveys. According to Kampen, Van de Walle, and Bouckaert (2006) trust in the government is important because it helps governments make decisions without using forceful actions. The U.S. Energy Information Administration (EIA) wanted to understand the perceptions respondents have towards EIA's ability to protect the confidentiality of their information and the respondents' perceived trust towards EIA.   Furthermore, EIA studied potential differences in perceptions between respondents protected under the Confidential Information Protection and Statistical Efficiency Act (CIPSEA, a stronger privacy pledge) and respondents protected under Freedom of Information Act (FOIA) exemptions. EIA conducted exploratory research using a web survey that included coverage of respondents from various energy sectors: Petroleum, Coal, Natural Gas, Uranium, Electric Power, and Renewable Energy. The sample size of this study was 978 respondents (172= CIPSEA, 806=FOIA Exempt). Overall, respondents had positive perceptions towards EIA's commodity-specific business surveys. Interestingly, respondents under the CIPSEA had lower perceptions of trust than respondents under FOIA exemptions. Participants had a high level of trust in EIA's promise of keeping data confidential (96%). The majority of respondents (93%) trusted that EIA will use the information they report only for statistical purposes. Lastly, results showed that respondents had c oncerns about EIA sharing data with academic institutions (52%= CIPSEA, 47% FOIA Exempt) and federal agencies (61%= CIPSEA and 52%= FOIA Exempt). Their concerns with EIA sharing their data with colleges/federal agencies stemmed from the security of the data (intruder hacking the system) and unauthorized personnel accessing their data.

Measuring Diversity, Equity & Inclusion: A Psychometric Analysis of Workplace Climate

Donald Levy, *Siena College*
LB Hannahs, *Tangible Development*
April Backus, *Siena College*

The Siena College Research Institute, in partnership with Tangible Development, LLC, developed and administered a Diversity, Equity and Inclusion (DEI) Measurement Tool- the Climate Survey. To date, the DEI Climate Survey has been fielded to over 7,900 respondents across 15 organizations. The instrument, a psychometric tool, asks employees about their individual experiences both in the workplace and in their lives, which when aggregated, provides the employer with an overall DEI Climate score for their organization. This presentation walks through that instrument development process and provides an overview of the application of the instrument through a case study analysis of one organization.   The DEI Climate Survey was developed through an iterative process whereby over 100 indicators were drafted and then tested multiple times using national panels to narrow down to the 45 existing indicators. When submitted to a factor analysis, the final indicators load on four constructs- institutional structure, institutional culture, personal beliefs and personal practices. While the indicators themselves vary independently, they are correlated at a statistically significant level with each of the derived underlying constructs.   The survey is delivered to all employees of an organization, asking respondents to indicate how descriptive each of the statements are of either the organization or themselves. Those indicators are then converted into scores and aggregated into a 'Percent of the Possible (POP)' Score for each construct and for the organization as a whole. The survey also boasts a robust demographic profile as well as a Tolerance Thermometer that gauges the perceived experience of groups of employees across 15 different identities. Together the DEI Climate Survey, Tolerance Thermometer, and robust demographic profile allow organizations to benchmark against themselves, set goals, and drive organizational change.

Panel Conditioning: Responding Once, Twice, Three Times or More

Setting Up An Online Access Panel Of People Of Immigrant Origin In Germany

Almuth Lietz, *German Center for Integration and Migration Research*
Madeleine Siegel, *DeZIM*
Sabrina Jasmin Mayer, *DeZIM*
Jörg Dollmann, *DeZIM/MZES*
Mariel McKone Leonard, *DeZIM*
Jonas Köhler, *DeZIM*

A multitude of different, often commercial, access panels offer researchers new opportunities, e.g., to quickly survey appropriate samples in the context of external events and crises or to track long-term attitudinal trends. However, large segments of the population, such as individuals who migrated to the host country themselves or who have at least one parent born abroad, or ethnic minorities are usually underrepresented in these infrastructures. This hinders the study of reactions of immigrant-origin groups towards specific external events, such as responses and threat perceptions following racist violence. The German Center for Integration and Migration Research (DeZIM) aims to counteract this trend and is building a more integrated data infrastructure in the form of an online access panel. In early 2021, the setup for the DeZIM.panel started, a non-commercial online access panel that is representative for several major groups of people of immigrant-origin as well as the native population in Germany. The sampling design is based on a two-stage cluster sampling procedure. First, municipalities, then personal addresses were randomly drawn from population registration offices. Subsequently, individuals were classified using onomastic methods. The DeZIM.panel is designed as multi-topic survey with constant core modules, and further, provide researchers to hand in their own research questions. Regarding panel management, we place great emphasis on keeping the panel readiness high and minimizing panel attrition. For this purpose, we use e.g. questionnaires in several languages, feedback routines as well as incentives. Data is provided by an in-house research data center. Our paper will provide an overview of the general setup of the DeZIM.panel as well as show data for the first topic-specific wave after recruitment, which will be conducted in November 2021. We intend to share best practices and discuss the challenges and potential biases within the panel.

Experimental Evidence on Panel Conditioning Effects of Increasing the Surveying Frequency in a Probability-Based Online Panel

Carina Cornesse, *German Institute for Economic Research*
Annelies Blom, *University of Mannheim and University of Bergen*
Ulrich Krieger, *University of Mannheim*
Tobias Rettig, *University of Mannheim*
Marina Fikel, *University of Mannheim*
Marisabel González Ocanto, *University of Mannheim*
Katja Möhring, *University of Mannheim*

In March 2020, with the onset of the pandemic, the German Internet Panel (GIP) intensified its longstanding bi-monthly data collection with an additional 16-week daily rotating panel of weekly questionnaires: the Mannheim Corona Study (MCS). The aim was to collect detailed information on the rapidly changing social, political, psychological, and economic situation of the German population. Built into the design was a control sample that continued with the standard implementation cycle of the GIP. In this presentation, we contribute to the panel conditioning literature with this large-scale experiment. The GIP is a probability-based online panel infrastructure covering a variety of questionnaire topics in sociology, political science, and economics. The GIP at the time consisted of panelists from three recruitment rounds in 2012, 2014, and 2018. The experimental study randomly assigned 1/8th of the GIP sample to each day of the week. Each week, on the same originally assigned weekday, the 7 treatment groups were invited to complete a new COVID-related online questionnaire within 48 hours. The remaining random 1/8th of the GIP sample serves as a control group. The design thus offers a unique opportunity to study panel conditioning effects on respondents of different panel maturity. First results show little impact of the MCS on panelists' participation behavior in regular GIP waves during and after the MCS fieldwork. Furthermore, MCS participants provide similar answers as the control group on various typical GIP questions, such as political interest or satisfaction with the government. However, we find that MCS participants are significantly more likely to report having installed the German COVID-19 contact tracing app than the control group, indicating potential topic-specific conditioning effects. Since the GIP still continues operation every other month, our presentation will include further analyses into the long-term effects of the MCS on different types of GIP question content.

Do Panel Effects Exists? Results from a Year 2 Follow-up Experiment to Assess Panel Conditioning Effect 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*
Keshav Vemuri, *NORC at the University of Chicago*
Meimei Zhu, *NORC at the University of Chicago*

Panel conditioning is the effect observed in a person's survey responses that is influenced by their panel tenure and panel experiences. Some panel conditioning effects may improve data quality and some may degrade data quality. The principal aim of the study is to investigate any potential conditioning effects on survey responses in an AmeriSpeak Panel survey fielded one year apart to the same panelists, seeking to understand whether and in which situations survey data quality improves or degrades as a result. We will present about the follow-up data collection effort and analysis to the cross-sectional analysis conducted using the first wave of data and presented at 2020 AAPOR in which we found a relatively small percentage of questions exhibited substantive panel conditioning effects. This was found to mostly occur with less tenured panelists who tended to select more extreme responses for questions with extreme response options. We found little evidence that more tenured panelists were different from less tenured panelists in their willingness to offer an opinion, whether they were more knowledgeable/interested in topics frequently asked in surveys, or more likely to provide moderate and/or no opinion responses to attitudinal questions. For Year 2, we will extend the cross-sectional analysis and examine less and more tenured respondent behavior longitudinally. We will analyze and present findings as to whether differences between responses to the same survey to the same respondents fielded 12 months apart may be due to panel conditioning, further enhancing our understanding of potential panel effects. Our research design uses a randomized experiment and statistical matching to make high/low tenure panelists as equivalent as possible, seeking to account for any panel attrition effects.

Does Participating in a Long-Term Cohort Study Impact Research Subjects' Longevity? Experimental Evidence from the Wisconsin Longitudinal Study

Rob Warren, *University of Minnesota*
Andrew Halpern-Manners, *Indiana University*
Jonas Helgertz, *University of Minnesota*

Cohort studies—research projects in which the same people participate on multiple occasions—are essential sources of information for scientific advances in public health, medicine, the social sciences, economics, and beyond. The resulting longitudinal data play a sustaining role in building knowledge about public opinion, health, human behavior, social relationships, inequality, and human development—but they also pose several complicated methodological problems. Among the least well-understood of these problems is "panel conditioning," or the bias that results when the very act of participating in the cohort study itself alters participants' outcomes or attributes. We will describe an experimental study of panel conditioning effects on an outcome that has clear and well-known social and behavioral antecedents: adult mortality. Given the potential for panel conditioning effects on a variety of health-related outcomes, we ask whether long-term participation in an intensive longitudinal survey of health behaviors and outcomes impacts the timing of participants' death. We address two specific questions: First, to what extent does repeatedly answering survey questions—especially about health behaviors/outcomes—influence the timing of respondents' death? Second, are mortality effects from panel conditioning distributed evenly across the population or do they vary systematically depending on respondents' sociodemographic characteristics? Our data permit a direct comparison between (1) people who—strictly by chance—completed a survey just once in 1957 and (2) their peers who—also by chance— completed the same 1957 survey and then have been in the Wisconsin Longitudinal Study (WLS) ever since. The WLS is an intensive health-focused study that has interviewed participants for 60+ years. If we observe significant differences between the two groups with respect to timing of death, we can infer that these differences are due to behavioral changes induced by panel conditioning

Phone Calling – Fuhgeddaboudit? The Implications of Mode Effects In Pre-Election Polling On The 2021 New Jersey Governor's Race

Ashley Koning, *Rutgers University*
Debbie Borie Holtz, *Rutgers*
Jessica Roman, *Rutgers*
Kyle Morgan, *Rutgers*

Since 2016, pre-election polling has been under heightened scrutiny: polls have frequently underestimated Republican political candidates as the tool becomes increasingly politicized and distrusted, particularly among Republicans at the mass level. While the recent AAPOR report stated that underestimation in 2020 was widespread regardless of mode or sample type, our polling during the 2021 New Jersey gubernatorial election potentially shows otherwise. New Jersey history indicated an uphill battle for Democratic Governor Phil Murphy heading into his reelection, but polling in the state showed a consistent double-digit or high single-digit lead up until the election. Yet much like Biden in 2020, Murphy's healthy margin was reduced to a few points when all votes were finally counted. Pundits decried that pollsters once again missed the mark, leaving many to wonder how a Republican challenger with little name recognition in a blue state almost beat a governor with healthy approval ratings. A look at mode in two Rutgers-Eagleton Polls leading up to the election may shed light on the closeness of the race. In two New Jersey statewide dual-frame telephone surveys, respondents from the cell phone sample frame contacted by live one-to-one SMS messaging and pushed to a self-administered online version of the survey were much more likely to indicate a preference for the Republican, Jack Ciattarelli, compared to those called on the phone from the same cell frame and from the landline frame. A slight majority of online respondents, in fact, chose the challenger, while a slight majority of those called chose Murphy. Yet on all other survey measures, online and phone respondents were similar. Observed modal differences allude to the need for further investigation of social desirability bias, as well as mode and interviewer effects, in pre-election polling to better understand potential nonresponse bias among certain segments of Republicans.

Benchmarking Probability-Based New England Web Panel Results in 2020-21

Zachary Azem, *University of New Hampshire*

Sean McKinley, *University of New Hampshire Survey Center*

Conducting RDD telephone surveys has become increasingly untenable financially and has been made more difficult by the COVID-19 pandemic. Low response rates have also called into question the ability of RDD surveys to accurately predict election outcomes. Some organizations have turned to building probability-based web panels in order to reduce costs yet to also maintain or even improve data accuracy. But while the practical appeal of probability-based web panels is quite clear, many have been hesitant to embrace them out of concerns that their accuracy in predicting known benchmarks, such as elections or vaccination rates, will not be as good as more traditional methods such as RDD telephone polling. The University of New Hampshire Survey Center has constructed probability-based panels in New Hampshire, Vermont, and Maine. This paper seeks to build on earlier research by comparing results from probability-based state panels to actual results from several elections in 2020 and 2021, including the 2020 New Hampshire State Primary, the 2020 General Election, and the 2021 Maine Referendum Election, as well as COVID-19 vaccination rates in all three states. Furthermore, it will discuss lessons learned and best practices that have emerged from utilizing these panels, in addition to potential non-response bias and how this issue can be mitigated through the use of additional weighting variables. In doing so, this paper hopes to demonstrate the viability of probability-based panels as an effective and durable alternative to more traditional survey methods.

What are You Doing with Whom, & Where? Measurement Error in the Home, Wallet, & Out & About
Sampling & Mode Effects in Election Polling

Revisiting the Gold Standard: What Election Data and Health Surveys Say about Probability and Nonprobability Samples

Peter Enns, *Cornell University*
Jake Rothschild, *Verasight*

Probability sampling, such as random digit dialing and random address-based recruitment, has long been the gold standard for survey research. However, declining response rates and high costs associated with these traditional recruitment methods have made non-probability samples, such as inviting respondents to opt in through online advertisements, more attractive to many. Yet, non-probability samples may offer reduced accuracy due to incomplete population coverage, systematic differences between those who opt in or out, and varying methods. Of particular concern is that while non-probability samples sometimes yield accurate results, non-scientific sampling means on average non-probability samples may lead to less accurate results. In light of increasingly sophisticated weighting procedures, developments in likely voter screens, and declining traditional survey participation rates, we seek to add to the literature comparing probability and non-probability methods in three ways. First, we analyze 355 national polls from the 2020 U.S. presidential election and more than 100 surveys measuring Covid-19 vaccination rates in 2021. We compare these polls to the respective benchmarks of final 2020 election results and vaccination rates tracked by the Centers for Disease Control. Our results show that surveys using non-probability samples are at least equally accurate as those fielded with probability samples, and in fact are more accurate in certain circumstances. These results hold when we compare average error, median error, and the least accurate polls from both methods. Second, for surveys with available individual-level data, we analyze unweighted results to evaluate how much of the accuracy of non-probability surveys comes from modeling the data. Preliminary results suggest that probability-based samples are more representative than non-probability samples, but that modeling leads non-probability samples to perform as well as probability-based data in the contexts under study. We conclude by discussing the implications for collecting, reporting, and consuming modern survey data.

Mixed Modes of Data Collection and Administration and New Sources of Samples in the US 2020 Election. How Did They Fare?

Claire Durand, *University De Montreal*
Tim P. Johnson, *NORC/UIC*
Luis P. Pena Ibarra, *Université de Montréal*

This paper presents the results of an analysis of the performance of the polls in the 2020 U.S. Presidential election that fare better, that is, with estimates within one percentage point of the final result during the last 10 days of the campaign. We find that, out of 47 polls, 12 polls fit the criteria, two of the four polls that used IVR as the main mode, three of the eight

polls using interviewer-assisted telephone polls and seven of 34 online polls.   We investigate the characteristics of these polls and find that seven of them – three telephone polls, two online polls and two IVR polls – use a combination of modes of data collection and administration. In addition, we find that, among the seven online polls that were closer to the final result, five used Random Device Engagement (RDE) to recruit respondents. In this paper we investigate further these results in order to better understand why we may conclude to a superior performance of some combination of modes or some specific sources of sampling. We rely on the information provided by the pollsters, either through their web site, their poll report, or answers to our requests for information. We conclude that, in the future, we will need to examine more seriously the methods used by pollsters and differentiate the polls using criteria that are more precise than only mode of administration or data collection.

Survey Participation to Advance Community Impact and Inclusion: New Platforms and Possibilities

Survey Participation to Advance Community Impact and Inclusion: New Platforms and Possibilities
Elissa Schuler Adair, *Lines for Life*

What are You Doing with Whom, and Where? Measurement Error in the Home, Wallet, and Out and About

Surveys and Financial Transaction Data: Comparisons of Different Data Collection Methodologies
Andrew Warren, *Financial Health Network*
Kevin Foster, *Federal Reserve Bank of Atlanta*
Marco Angrisani, *University of Southern California*

Research on personal financial behavior and activities can be limited by data collection challenges. Individuals' personal finances are often spread across financial institutions and accounts, making comprehensive administrative data collection difficult. Likewise, self-reported data on financial activities via surveys are subject to respondents' abilities to recall detailed and complex information. Using several data sources from the Understanding America Study survey panel, we shed light on differences between financial activity data collected via survey questions and an opt-in financial aggregator.   We create a novel dataset containing detailed survey responses on individuals' financial activities over a period of three days (collected through Diary of Consumer Payment Choice and Survey of Consumer Payment Choice surveys) and merge it with financial transaction data from the same individuals collected during the same period through a financial aggregator (collected via the Financial Health Pulse study). This merged dataset, with a sample size of approximately 300 individuals, allows us to begin to understand the potential systematic differences between these data collection methods.   We find evidence that financial transaction data collected via the opt-in financial aggregator captures more checking account transactions than are reported in survey data, as well as more total non-cash transactions. The number of credit card transactions, by dollar value and volume, are also higher in the data captured via financial aggregator, though the difference is marginally significant. There is no significant difference in the total value of checking account transactions. These results raise important considerations for researchers considering different data collection methods. Finally, it is important to note that, short of having a third, objective, and unbiased record of an individual's financial activity against which self-reports and transactional data can be compared, our results should be viewed as informative but not necessarily conclusive about differences in financial data arising from different data collection methods.

Translation of Social Identity to Economic Identity: Group Economic Performance Thru and Identity Lens via Direct Survey Measurement
Jonathan Nagler, *New York University*
Suzanna Linn, *Penn State University*
Jan Zilinsky, *TUM School of Social Sciences and Technology*

With the rise in identity politics has come a renewed interest in the groups to which people feel they belong, and how those group identities shape political behavior. Evidence suggests identity politics shape evaluations of politicians and political issues. We ask whether identity politics influences views on a fundamental material bread and butter issue of

politics: economic performance. We introduce a measure of economic identity, and thru the use of novel, but direct, survey questions we examine the extent to which social identities translate into economic identities: who people think of when evaluating the economic performance of "people like you." We find that social identity only weakly translates to economic identity, there is little overlap between the groups people say are important to their overall identity and the groups people say are important to their economic identity. We also examine the claim that Donald Trump catalyzed group identity among less educated, white Americans with his claims that people like them were getting left behind. If such claims are true, we would expect less educated, white Americans to be more likely to reference their ethnic group (or, education group) when evaluating how the economy was performing for people like themselves. But the data suggest that while Trump catalyzed social identity among lower educated whites, he did not do the same for their economic identity. This suggests that while social identity may be a lens through which people evaluate politics in many circumstances, it is not a lens which people necessarily use in evaluating economic performance, a key correlate of voting behavior.

How Does Sleep Affect Recall Error?: A Complex Sample Design Adjusted Analysis from 2010 American Time Use Survey
Melike Sarac, *Hacettepe University Institute of Population Studies*

Although 24-hour time-use diaries are designed to minimize recall error, respondents are nevertheless likely to encounter episodic memory failures when recalling their previous day's activities. From total survey error (TSE) perspective, recall error and factors affecting it should be well-investigated to achieve high quality time-use statistics. The current study therefore examines the incidence of recall error in the ATUS in the U.S. population and investigates the impact of respondent ability, including sleep deficiency and sleep onset, and motivation on the likelihood of recall error. Proportion tests and logistic analyses were carried out taking complex sample design of the 2010 ATUS into account so that inference applies to U.S. residents aged 15 and over living in occupied households. This study found that 16.6% of the general population are likely to commit recall errors in the 2010 ATUS and that the strongest predictors of recall error are busyness and labor force status. These results would warrant programming more recall cues or interview warnings to slow interview speed and double check CATI system entries to help interviewers obtain better quality responses from individuals who are busier than average or are either unemployed or out of the labor force.

It's a Bird… It's a Plane… It's a Survey!: Surveying Air Passengers at Three Baltimore-Washington Area Airports
Zoe Padgett, *ICF*
Kenneth Joh, *Metropolitan Washington Council of Governments*

Air passengers are a unique and challenging population to survey. Busy travelers may not have time for an in-person survey while trying to catch a flight, and the reluctance of airlines to provide detailed flight information makes post-travel surveys difficult. In areas with large airports understanding passengers is critical to transportation planning. The Metropolitan Washington Council of Governments has conducted the Washington-Baltimore Regional Air Passenger Survey (APS) at the region's three commercial airports biennially since 2005. The APS provides critical data on air passenger characteristics, ground access information, and factors that influence airport choice. The APS is the key source of observed data for regional air systems planning and informs long-range regional transportation planning. The APS is conducted as a paper-based intercept survey of passengers at boarding gates. The survey used a stratified sampling approach based on characteristics like airline and destination, and data collectors attempt to survey all passengers on each sampled flight. In recent surveys, response rates and response quality have declined with increasing break-offs and item nonresponse. This presentation shares the results of a multi-pronged evaluation to assess the drivers of declines in APS response and data quality while identifying strategies for airport-based field research. We present findings from literature reviews, technical advisory committees, key informant interviews, and airport site visits. We discuss findings on data collection frequency, questionnaire flow, incentives, and interviewer training, and current efforts to improve the survey. Additionally, we address issues of inclusion by considering whether airport employees should be included in the survey given their own transportation needs. While some findings are specific to airport-based surveys, we believe that many are applicable to in-person surveys at-large. Considerations include multiple facets of total survey error, including coverage error, nonresponse bias, and measurement error, and fitness of use criteria including timeliness and coherence

You Can Tell Us: Collecting Data on Sensitive Topics

Web Versus In-Person Interviews on Reporting Sensitive Topics: Findings From 7th Wave of World Values Survey for European Countries

Melike Sarac, *Hacettepe University Institute of Population Studies*
Ismet Koc, *Hacettepe University Institute of Population Studies Department of Demography*

Web surveys have certain advantages although low response and coverage problems may arise due to lack of interviewer assistance, low literacy and technology use. Response accuracy for sensitive topics, that is influenced from social desirability, may increase in web-based surveys, as a result of increased reporting and admitting (Loosveldt and Sonck, 2008; Duffy et al., 2005). This study aims to compare reporting behaviors of respondents from in-person interviews (CAPI or PAPI) and web-based interviews (CAWI). The data come from 7th wave of World Values Survey (WVS) conducted between 2017 and 2020, in selected countries (Denmark, Finland, Germany, Iceland, Netherlands, Switzerland), which used both modes simultaneously. In total, 43 percent of interviews were in-person while 44 percent of those were web interviews. Mail/post respondents consist of 13 percent of total interviews. To reach study objective, sensitive variables referring social values, attitudes, stereotypes, norms were selected, and their reporting levels were assessed. Our preliminary findings suggest that stating importance of religion in life as 'very important' is much more frequent for CAWI respondents than in-person respondents. Further, CAWI respondents seem to be willing to report that they do not like to have drug addicts, immigrants/foreign workers and heavy drinkers as their neighbors. When ethical values/norms are examined, justifications related to claiming government benefits to which respondents are entitled, avoiding a fare on public transport, cheating on taxes, homosexuality, abortion and divorce is observed much more often in web-based interviews than in-person ones. Initial results introduce adopting web mode which allows flexibility through self-administration, even if main data collection technique is face-to-face. Finally, the study has a potential to contribute to literature focusing on efficiency of mixed-mode surveys under examination of WVS. In further phases, reporting will be examined under control of covariates such as sex, age, income level, marital and employment status.

Mode Effects on Sensitive Questions: Evidence from an Experiment in Jordan

Julie De Jong, *University of Michigan*
Salma Al-Shami, *Princeton University*
Julie de Jong, *University of Michigan*
Zeina Mneimneh, *University of Michigan*

There is ample evidence that survey mode can impact reporting of sensitive information. The impact seems to be relatively consistent, with self-administered modes leading to higher reporting of sensitive information relative to interviewer administered. However, when comparing telephone interviewer-administered modes with face-to-face interviewer-administered modes, the effects are less clear. While the interviewer's physical presence in face-to-face surveys might lessen perceptions of privacy or heighten social desirability concerns, the visual aspect might facilitate rapport between the respondent and interviewer, improving reporting of certain types of sensitive information. With the pandemic forcing many surveys to shift from face-to-face to telephone interviewing, investigating the effect of mode on reporting sensitive questions is essential. This is especially important for long-standing surveys that estimate trends in attitudes and have long relied on a face-to-face mode of administration. In this paper, we use data from an Arab Barometer survey in Jordan leveraging an experimental design to isolate measurement differences in reporting sensitive attitudes and behaviors. We selected an area probability sample and assigned one-third of the households within a Primary Selection Unit to be contacted and interviewed face-to-face and the other two-thirds to be initially contacted face-to-face to obtain contact information, with subsequent interviewing completed by telephone. This design allowed us to maintain the sample and coverage properties for both phone and face-to-face administration, thereby isolating measurement error. The instrument included questions from earlier waves of the Arab Barometer varying on level of sensitivity, including those related to the government, gender attitudes, and religious beliefs. We also included additional measures to identify possible mechanisms behind some of the potential measurement differences including measures of trust and social desirability concerns. Findings from this experiment will contribute to a better understanding of measurement mode effects on reporting sensitive information and the potential mechanisms, particularly in a non-western setting

Testing for a Mode Effect When Collecting Data on Rape and Sexual Assault: A Comparison of Self-Administered Versus Interviewer Administered Surveys.

David Cantor, *Westat*
Darby Steiger, *Westat*
Reanne Townsend, *Westat*

This presentation will provide results of a study that tests for the effects of survey mode for the collection of rape and sexual assault victimization (RSA) among a general population of women age 18-49. The study was funded by the Bureau of Justice Statistics and was designed to improve the measurement of RSA on the National Crime Victimization Survey. Collecting reports of RSA on a survey is extremely sensitive. Based on this, one would expect there to be a significant difference between self- and interviewer administered modes. However, unlike research on other types of sensitive behaviors (e.g., drug use, sexual activity), there has not been a great deal of experimental research comparing these two modes when collecting reports of RSA. In fact, a number of surveys that have been conducted using telephone interviews, using special procedures to increase confidentiality and privacy, that have found very high self-report rates of RSA. However, it is hard to assess whether the procedures used to maximize privacy and confidentiality reduce or eliminate the mode effect, or if they produce high rates because of the use of different question wording. This presentation reports on the results of a study that compares estimates from a survey administered with audio computer assisted self interview (ACASI) to a parallel random digit dial telephone survey of females age 18-49. The questionnaires administered were designed using very similar survey questions. Contrary to the published literature on mode effects, this study did not find a significant difference in RSA prevalence rates between modes. This paper will explore potential explanations for the non-significant effect related to question wording, differential non-response, coverage, and procedures used to preserve privacy and confidentiality.

Respondents' Perceptions and Preferences in Sensitive Topics

Mina Muller, *Ipsos Public Affairs*
Zachary Lewis, *Ipsos Public Affairs*
Mary Noorzai, *Ipsos Public Affairs*

When asking about highly sensitive topics such as illegal activities, violence and victimization in a survey, researchers use different strategies to gain cooperation before the survey begins. Some of these strategies are providing an introduction text about the topic, reassuring of confidentiality, asking respondents to complete the survey in a private setting. However, language meant to reassure participants may actually cause discomfort and could cause higher item nonresponse and drop off rates. This research addresses respondents' perceptions of sensitive questions and provides insight into factors that may encourage or discourage participation on a web survey of sensitive topics. The goal of this presentation is to provide empirical evidence and recommendations on respondent perceptions and preferences to help researchers mitigate participation hesitancy and improve data quality in social and behavioral research on sensitive topics. The data for this paper is gathered via a genpop sample conducted via the largest probability-based online panel in the U.S. (KnowledgePanel). Respondents are asked about different types of sensitive topics such as sexual violence, aggravated assualt, and personal larceny. We also evaluate the level of burden and emotional distress respondents anticipate experiencing when answering questions about these topics, and reasons for reluctance in responding or providing accurate responses to these types of questionnaires. Furthermore, the research investigates factors that promote participation such as confidentiality and safety measures established, preferred survey language and collection mode. Additionally, we analyze findings by key demographics and socioeconomic indicators.

Effect of Question Reader's Sex on Substance Use and Mental Health Reports in an Audio Computer-Assisted Self-Interview (A-CASI) Survey

Sunghee Lee, *University of Michigan*
Aaron Nesbitt, *University of Michigan*
Karen Gates, *University of Michigan*
Anqi Liu, *University of Michigan*

Survey interviews are a social interaction between researchers and respondents. For interviewer administered surveys, interviewers are involved in this interaction on behalf of the researchers. Because of the social nature of the interviews,

characteristics of interviewers serve as additional visual or audio cues to the interaction. In particular, respondents may perceive questions about attitudes or behaviors that they have associated with inferred physical, attitudinal or behavioral characteristics of the interviewer (e.g., attitudes towards a racial/ethnic group; religiosity) as socially sensitive depending on the interviewer characteristics. This may prompt respondents to offer answers that conform to social norms rather than their "true" state or belief, as shown in face-to-face and telephone surveys. This line of research, however, for the Audio Computer-Assisted Self-Interview (A-CASI) setting is limited but has shown that the sex of voice affects responses. The results are somewhat mixed: some favoring data quality with the female voice and others favoring matching the sex between the ACASI voice recording and respondents. This study attempts to extend this literature by experimentally examining the sex of ACASI voice in a survey of persons who inject drugs (PWID) where respondents will be randomly assigned to either male or female voice. Our analysis will focus on the main effect of the sex of voice on responses given to substance use and mental health questions. We then will examine the interaction between the sex of voice and the sex of respondents. This study is expected to shed light on the instrument design for hard-to-survey vulnerable population subgroups such as PWID.

You Don't Have to Tell Me Twice: Dependent Interviewing & Previously Reported Data

Respondent Opinions of Previously Reported Data in Web Surveys

Joseph Rodhouse, *U.S. Energy Information Administration (EIA)*
Kathy Ott, *National Agricultural Statistics Service*

A common complaint among respondents that are repeatedly sampled in establishment surveys is that they have to answer many of the same questions over and over again despite no changes in their situations. When this happens, respondent burden and frustration with the survey organization can increase, potentially hurting the likelihood of future cooperation. In a qualitative study conducted at the National Agricultural Statistics Service (NASS) to determine how the Agency could better serve its population, responses included "make questions easier to answer," and "use data I have previously reported to either NASS or its affiliated agencies." As a result, NASS is carried out an experiment in the 2020 Census of Agriculture Content Test where sample units' previously reported data (PRD) is pre-filled in the answer cells of the web mode of the survey. The motivation for incorporating PRD is to address the concerns of NASS's survey population, and help make response faster and easier, hopefully resulting in a reduction in burden overall. To test these goals, the web mode with PRD concluded with a series of eight questions designed to gauge respondent opinions of PRD in their surveys (including questions designed to assess burden and satisfaction). As it is likely respondents' opinions of PRD will vary, an analysis of the factors related to satisfaction and dissatisfaction of its use will be examined. Most studies of PRD are qualitative and have limited respondents. The sample size of this study is 9,000 and will likely produce one the largest quantitative measures of respondent opinions of PRD in the survey literature to date. Conclusions, recommendations, and future research based on respondent opinions of PRD in web surveys will be discussed.

Testing the Measurement Error of Dependent Interviewing in a Self-Administered mode

Curtiss Engstrom, *University of Michigan*
Dr. Jennifer Sinibaldi, *National Center for Science and Engineering Statistics*

Memory of past responses to a longitudinal data collection can be lost over time. As a result, measurement error in longitudinal studies can increase between waves. A remedy that promotes consistency between waves is dependent interviewing, where responses to questions from previous waves are provided to respondents. This method not only helps respondents remember their previous responses but also reduce respondent burden when the information is the same. However, there is concern that some respondents may further reduce their burden by not updating the information from last wave. Also, dependent interviewing is traditionally conducted in interviewer-administered modes, such as face-to-face interviewing, and not in self-administered modes such as web. The Survey of Doctoral Recipients (SDR) conducted a web pilot study in 2020 implementing two different dependent interviewing conditions and a standard web survey (control group). In our analysis, we compared the two dependent interviewing conditions against each other and against the control group looking at different indicators of measurement error. We looked at possible relationships between experimental group and the number changes in both closed and open-ended questions. We also examined perceived response burden for respondents across groups. We evaluate whether the changes and consistencies are indicative of increased or reduced measurement error.

Operationalizing the Use of Previously Reported Data on the 2022 Census of Agriculture

Gavin Corral, *USDA*
Linda J Young, *USDA*

The USDA National Agricultural Statistics Service (NASS) is incorporating the use of Previously Reported Data (PRD) into its Census and survey programs. One major goal is to have PRD integrated into its 2022 Census of Agriculture. In preparation for the Census, several tests on the potential impact of PRD on data quality, respondent experience, and respondent burden are being conducted. Concurrently with testing, NASS is vetting and training users in the new PRD delivery system. In this paper, the results from studies embedded in (1) the 2020 Census of Agriculture Content Test and (2) the 2020 September Crops Agricultural Survey are discussed. Furthermore, the lessons learned from the first operational use of the PRD delivery system during the 2021 June Agricultural Survey are presented. The results indicate that PRD positively influences the respondents experience, reduces perceived respondent burden, and may lead to increased data quality by decreasing the number of spurious changes, thereby increasing the accuracy of survey estimates. Areas of future research are considered.

Use of Dependent Interviewing in Establishment Surveys: What We Know, Don't Know and Where We Need to Go

Heather Ridolfo, *Energy Information Agency*
Gerson Morales, *U.S. Energy Information Administration*

Dependent interviewing is a method of presenting survey respondents with previously reported data to reduce response burden, increase efficiency and reduce measurement error, specifically spurious change in survey data over time. This method is commonly used in panel surveys and cross-sectional surveys where respondents are repeatedly sampled. In dependent interviewing, previously reported data can be presented as a reminder to respondents of their previous answers or used to probe on inconsistencies in reporting. Research has shown that the presentation of previously reported data can have implications for data quality. However, most published research on dependent interviewing focuses on household surveys and less is known about its utility in establishment surveys. Establishment surveys pose unique challenges to data collection, including identifying the appropriate person to answer survey questions. Individuals within an organization may have varying degrees of knowledge and access to data, which may lead to differing interpretations and responses to survey questions. Responding to survey requests may also require delegation of survey tasks to different respondents within an organization. In order for dependent interviewing to be effective, the previously reported data must be presented back to the previous respondent or to another respondent within the organization who has equal knowledge and capacity to answer it. Given the scarcity of published studies on dependent interviewing in establishment surveys, it is difficult to determine how to best implement dependent interviewing in this setting. To address this, the Energy Information Administration is currently conducting an evaluation of the impact of dependent interviewing on user experiences and data quality in establishment surveys. This presentation will present results from a literature review, highlighting areas where additional research is needed in establishment surveys, and results from a survey of agencies within the Federal Statistical System on their experiences with dependent interviewing in establishment surveys.

Collecting Income Data in a Longitudinal Setting: An Experimental Study

Nicolás Libuy, *Centre for Longitudinal Studies, UCL Institute of Education*
Emla Fitzsimons, *Centre for Longitudinal Studies, UCL*
Nicolas Libuy, *Centre for Longitudinal Studies, UCL*

Participants of the Millennium Cohort Study (MCS), a UK cohort of about 19,000 individuals born around 2000, will have reached early adulthood by the time of the next wave of data collection. It will be the first time to collect detailed income measures from the study members, and in the interests of longitudinal continuity, the measures we select at this baseline adult wave will be carried forward in future waves – so the choice of measure is particularly important at this stage of the study. Measuring income is known to be challenging due to its many components, which tend to vary greatly by socio-economic status. Surveys of parents in earlier waves of MCS include a long and detailed module on income, and also included single, stand-alone questions on total take-home income. In this study, we aim to test the properties of different measures of income, with the primary aim of testing a shorter measure that reduces the

extensive set of questions on benefits against the 'gold standard' longer measure to minimise respondent burden. We run an experimental survey on Prolifics using Qualtrics (N=1,000), with respondents allocated to one of four groups, in order to study: 1) correlation between a shorter income module, which mainly reduce the extensive set of questions on benefits, with longer measures, 2) whether the detailed income modules provide close estimates to single questions on income, 3) whether single questions on income perform better in a banded fashion vs. an unbanded (open-ended) one. All of these will be analysed for the whole sample, as well as for subgroups as defined by a variety of socio-demographic characteristics

What Can We Learn? Designing, Conducting & Disseminating Experimental Research

Survey Use in Substantive Research: Trends Across Disciplines

Chase Harrison, *Harvard University*
Ádám Stefkovics, *Harvard University*
Dominic Skinnion, *Harvard College*
Angela Eichorst, *Harvard College*

Survey research is widely used across a variety of academic fields. During the past decade, technological and sociological changes have led to a diversity of approaches used in survey research. On the one hand, many survey research tools (such as access to internet sample frames and computer assisted questionnaire development tools) have become much more accessible to researchers. On the other hand, professionally conducted probability based survey research interviews have increased astronomically in cost. This paper documents and evaluates the use of surveys in published in academic research across several fields from 2000 through 2020. Data will be drawn from all articles published between 2010 and 2020 in the top three ranked (impact factor) journals in medicine, sociology, and political science. We will examine the proportion of studies in these journals that use surveys, and report trends or changes. We will evaluate the source of the survey (secondary analysis or custom survey), the purpose of the use of the survey (generalizable estimates; conditional analysis; experimental manipulation); the type of research design (cross-sectional survey, repeated cross section, or longitudinal); the type of population (general population or special population); the type of sample frame used, and whether probability or non-probability sampling methods are used. We will also evaluate the extent that that the AAPOR Transparency Initiative disclosure elements are available in the journal article or through supplemental exhibits or archives..

Three Classes of Wicked Problems in Statistical Information Systems: Stakeholder Priorities, Design Decisions and Shared Interpretation of Results

John Eltinge, *United States Census Bureau*

Practical work with statistical information systems (including traditional sample surveys, as well as the integration of survey and non-survey data) has encountered remarkable opportunities and challenges over the past several years. These arise in part from increasing heterogeneity in priorities expressed by key stakeholders; complex design decisions required to balance multiple dimensions of quality, risk and cost; and the increasing need for practical two-way communications with data users regarding published results, policy implications, and related uncertainties.   These issues involve complex trade-offs among competing interests of multiple stakeholder groups.   Moreover, realistic negotiations to address these issues often encounter complications due to the highly contingent nature of the priorities of some internal and external stakeholders. Somewhat similar complexities arise in the design of many other public-oriented programs, e.g., environmental policy, transportation policy, and drug-control policy; and have led to a controversial literature under the rubric of "wicked problems."   In keeping with the conference theme of "Advancing Inclusion and Equity through Data Collection, Measurement and Community," this paper suggests that some practical ideas and analytic approaches from the "wicked problems" literature can help improve our understanding of, and responsiveness to, diverse and competing groups of stakeholders served by statistical information systems. Throughout the paper, four cross-cutting examples motivate and illustrate the principal issues and conclusions: (i) diagnosis and mitigation of survey nonresponse bias, especially for relatively small and hard-to-measure subpopulations; (ii) movement from a traditional survey to a system that integrates multiple sources of data and produces additional statistical series; (iii) design of new statistical information systems to measure rapidly evolving phenomena; and (iv)

implementation of some features of the Foundations of Evidence-Based Policymaking Act of 2018.   Each of examples (i)-(iv) involve challenges related to stakeholder priorities, design decisions and interpretation of empirical results.

Making Lotteries Legible: Designing Natural Experiments

Daniel Goldstein, *New York City Department of Housing Preservation & Development*
Elyzabeth Gaumer, *New York City Department of Housing Preservation & Development*

Often described as the gold standard for program evaluation, randomized experiments are used to eliminate confounding and reduce selection bias in measuring the impact of an intervention. But investigator-controlled randomization is often infeasible or cost-prohibitive or both. Natural experiments, on the other hand, often leverage existing lotteries that public agencies use to distribute constrained resources. When randomization is conducted with programmatic intent rather than strictly for research, however, complications can arise in conducting proper analyses, and care must be taken in constructing weights and specifying models. Researchers facing complex lotteries need to design weighting schemes and analysis plans such that controls are appropriately matched to corresponding treatments to recover the estimand(s) of interest. Affordable housing lotteries in New York City provide ample opportunities for these sorts of research designs. In this paper, we present case studies to show how housing researchers in an applied government setting in New York City put these concepts into practice. We describe the existing lottery process, potential study designs, weighting approaches, and estimation strategies to show the challenges and opportunities for leveraging existing randomized allocation programs. We include recommendations for other researchers seeking to analyze complicated program lotteries.

The Benefits of Reporting "Likely Effects" in Probabilities or Percentage Points

George Terhanian, *Electric Insights*

In a 1976 interview, pollster Lou Harris derided peers, whom he dubbed "political eunuchs," for believing their job was done once they had tallied scores, such as presidential approval ratings. He directed his criticism at George Gallup, who had argued that pollsters should be "fact-finders and scorekeepers, nothing else." In contrast to Gallup, Harris believed that all social scientists and all pollsters had a fundamental duty to advise on how to raise scores on key measures.   In other words, he felt compelled to describe the "likely effects" of possible actions in everyday language. For John F. Kennedy, for instance, Harris had to estimate the precise vote-share increase JFK would earn if he emphasized specific issues in the run up to the 1960 presidential election. But Harris's approach did not scale—he began his career when the computing equivalent of a modern laptop weighed more than 25 tons. That made it difficult to carry out logistic regression analysis, though it was, and is, "the standard way" (Gelman & Hill, 2009) to predict binary outcomes like vote choice. Interpreting standard output from logistic regression proved to be a problem, too. Most people—not only JFK in 1960—do not understand how to act on findings or recommendations reported in logits, odds, or odds ratios. That is probably why DeMaris (1993) observed "...there is still considerable confusion about the interpretation of logistic regression results." And why Gelman and Hill (2009) asserted "...the concept of odds can be somewhat difficult to understand, and odds ratios are even more obscure." The aim here is to assess and communicate the possible benefits of reporting "likely effects" in probabilities or percentage points (in addition to logits, odds, or odds ratios), using Pew data on attitudes toward driverless vehicles to illustrate key points

Why Survey Experiments Fail: Evidence from Re-analyses of 100 Experiments.

Jeremy Freese, *Stanford University*
Tamkinat Rauf, *Stanford University*
Jan Gerrit Voelkel, *Stanford University*
James N Druckman, *Northwestern University*

Population-based survey experiments have become a popular methodology for causal inference across the social sciences. Like all research, published experimental studies suffer from publication bias, making it difficult to glean the factors that constitute "successful" research designs and to learn from less successful ones. We study the efficacy of survey experimental designs using a sample of experiments fielded via the Time-sharing Experiments for the Social Sciences (TESS) platform. Investigators propose projects to TESS; these are peer-reviewed; and selected proposals are fielded using an online probability sample of US adults (currently NORC AmeriSpeak; formerly the Ipsos

KnowledgePanel). We sampled 100 such experiments fielded between December 2012 and February 2020. Using the research proposals submitted by investigators, we drew up a comprehensive list of hypotheses in each study, implying nearly a thousand specific tests. We look at both the success of specific tests within a study and whether a study had more successful tests than expected by chance. To understand sources of variation in study and hypothesis success rates, we examine a range of attributes, including features of the individual studies (such as the type of experiment, clarity of the research proposal, and sample size), differences across hypotheses (such as outcome measurement and whether the hypothesis entailed mediation or moderation), as well as any seasonal effects due to the timing of data collection. We also inspect whether the investigators' gender, discipline, or seniority has any bearing on their experiments' success. The findings of this study help better apprehend how various aspects of survey experiments impact the chances of success, help test researchers' preexisting intuitions about features of good designs, and identify aspects of survey designs that have greater need for creativity and improvement.

Innovations in Address-Based Sampling

Have Your Cake and Eat It Too: The Utility of a Prepaid Sample Component in bolstering the Representativeness of an Address-Based Sample Design

Arina Goyle, *SSRS*
Susan Sherr, *SSRS*
Royce J. Park, *UCLA Center for Health Policy Research*
Christine Loveridge, *The Center for Health Information and Analysis*

Address-based sampling (ABS) with push-to web designs have helped combat the precipitous drop in response rates and the resulting sharp rise in costs associated with random digit dialing (RDD). However, they do not provide a panacea for improving response from certain key harder-to reach-subgroups. Adding a phone component targeted to sample prepaid or "pay as you go" cellphone users, and including screening criteria for desired groups, could help bolster response from key subgroups while reaping the benefits of an ABS design. Previous research (McGeeney, 2015) has shown that prepaid phone users are more likely to be non-white, lower income, and less educated and live in urban areas compared with other cellphone users. These are key demographics that tend to be underrepresented in typical ABS studies. While including a prepaid phone sample can be a costly endeavor, employing this type of sample to supplement the overall study can bolster the representation of certain groups without overwhelming the budget. The current research explores the utility of adding prepaid phone sample components to major state-level health care studies that primarily use an ABS push-to-web and telephone nonresponse follow up design for data collection: the 2021 California Health Interview Survey (CHIS) and Massachusetts Health Insurance Survey (MHIS). The analysis will compare key subgroup representation between prepaid and ABS samples and will also explore the effectiveness of adding additional screener questions to distill the prepaid sample by targeted subgroups. Preliminary results suggest that prepaid sample respondents are more likely to be Hispanic, African American, less educated, lower income and uninsured. They also suggest the utility of screening questions in sampling some more elusive groups such as lower income Hispanics and lower income African Americans. This paper will also explore regional differences in, and discuss potential barriers to, the inclusion of prepaid cell phone sample.

Examination of Auxiliary Information Appended to Address Based Samples in Mixed Mode Studies

Ipek Bilgen, *NORC at the University Of Chicago*
David Dutwin, *NORC at the University of Chicago*
Grace Xie, *NORC at the University of Chicago*
Chang Zhao, *NORC at the University of Chicago*

The use of mixed mode methodology has been increasingly common during recruitment and data collection, especially in general population surveys. To be able to contact and interview individuals via various modes, researchers need auxiliary contact information to be appended to their sampling frames. One of the more commonly used strategies in multimode studies is to use address-based sample (ABS) frames given the efficiency and coverage properties of the ABS frames that are created using USPS Delivery Sequence File (DSF). In these studies, appending auxiliary contact information such as name, phone, and email to addresses is needed to enable the use of various modes during recruitment and increase the efficiency of the outreach. The incidence and accuracy rates of the appended auxiliary

contact information match impact the success of the outreach efforts in multimode studies. Accordingly, we have conducted a study using NORC's AmeriSpeak probability-based panel and asked a representative sample of panelists whether initials, phone numbers, and emails provided by various data sources are currently used by anyone in their household. We have examined which data sources or combination of data sources produces the most optimal incidence and accuracy rates. Using additional data on the efficacy of the appended data by the providers, we built models designed to predict whether a given data point would in fact be accurate to a given sampled household. In this presentation, we will explore incidence and accuracy rates of appended names, emails, and phones (cell versus landline) both at the household and individual level. We will present our examination of how much modeling strategies improve the accuracy of the auxiliary data sources for various populations. Overall, we will illustrate the incidence and accuracy rates of various strategies of deploying appended auxiliary contact information to utilize such data more efficiently and effectively in multimode surveys.

Supplementing Address-Based Sample With Prepaid Cell Sample to Help Improve Sample Representativeness

Jonathan Best, *Social Science Research Solutions (SSRS)*
James McKinstry, *SSRS*
Arifan Hasanbasri, *SSRS*
Christine Loveridge, *Center for Health Information and Analysis*
Huong Trieu, *Center for Health Information and Analysis*

With RDD telephone sample response rates continuing to decline and the associated rise in data collection costs, address-based samples have become a viable alternative to the dual-frame RDD samples that have been the mainstay of public opinion research for the past 20 years. The Computerized Delivery Sequence File offers an electronic frame from which to draw probability samples with very high coverage of the US population. However, one of the challenges with address-based samples is that they tend to result in samples that are less socioeconomically diverse and skew older and more educated. To help improve ABS representativeness, it can be combined with sample of prepaid cell numbers. These prepaid or "pay as you go" cell numbers, which are not attached to a contract with a cell phone carrier, are more likely to be used by groups that are typically underrepresented in ABS, including non-white, lower income and less educated. This paper investigates several methodological issues that arise when combining ABS and prepaid cell samples. First, we discuss a straight-forward weighting adjustment that can be used to combine the two samples. Second, we explore a method for determining appropriate sample allocations across the two sample frames. Finally, we do an analysis to explore the relative costs of collecting data by ABS and prepaid cell sample, both overall and by key subgroups.

Assessing General Population Representativeness in the British Open Register

Joseph Phillips, *Pennsylvania State University*
Nikhil Sengupta, *University of Kent*

In Britain, the Open Register is a potentially useful tool for researchers to use for postal surveys of the general population. Unlike the Postcode Address File, which contains all households, the Open Register also contains names, making it possible to personalize invitations (a strategy known to increase response rates). It is derived from the Electoral Register, and since it is a legal requirement to register to vote in Britain, the Electoral Register should be fairly comprehensive. However, the Open Register excludes three subgroups: people ineligible to vote (who cannot be on the Electoral Register), people who refrained from registering to vote despite the legal requirement, and people on the Electoral Register who explicitly opted out of inclusion in the Open Register. In this paper, we attempt to quantify coverage bias in the Open Register by comparing households on the Open Register with those on the Postcode Address File on a variety of local-level characteristics.

Using an Interactive Mapping Dashboard to Conduct In-Office Housing Unit Listing using County Assessor Tax Parcel Data

Lee Fiorio, *NORC at the University Of Chicago*
Peter Herman, *NORC at the University of Chicago*
Chang Zhao, *NORC at the University of Chicago*
Ned English, *NORC at the University of Chicago*

Frames for in-person household surveys are typically sourced from lists of addresses such as the USPS Delivery Sequence File (CDS or CDSF). However, due to CSD under-coverage in rural areas, some additional housing unit listing is still necessary to enhance the CDS and minimize overall coverage bias. In-person address listing is labor intensive and costly, but the growing availability of new kinds of data and advancements in interactive web mapping applications have opened the potential for CDS address list enhancement to be conducted in-office using GIS. In particular, the rise of digitized tax parcel data files created and maintained by county assessor's offices throughout the US offer a source of high-quality address information that can potentially be used to supplement CDS address frames for household surveys. This presentation will discuss the suitability of county assessor data for address listing and describe NORC's efforts to integrate the CDS with county assessor data in a Remote Listing Dashboard used to improve the address frame for NORC's 2020 Decennial National Master Sample. The Remote Listing Dashboard is an R Shiny and JavaScript web app that allows in-office listers to add and remove addresses from the CDS by first joining CDS addresses with vendor-provided county assessor tax parcel data through deterministic record linkage techniques and mapping the linked results against publicly available satellite and building footprint data as well as project-specific sampling geographies. Differences between census-provided housing unit counts and address counts from the CDS are mapped to show which areas require listing, and in-office listers select missing addresses to add to the frame from county assessor tax parcel data. This presentation will cover the development of the Remote Listing Dashboard and provide an assessment of where in-office CDS enhancement is adequate and where in-person is still needed, such as tribal lands.

New Techniques for Combining Data from Different Samples

Universal Adaptability: A New Method to Draw Inference from Non-Probability Surveys and Other Data Sources

Christoph Kern, *University of Mannheim*
Michael P. Kim, *University of California Berkeley*
Shafi Goldwasser, *University of California Berkeley*
Frauke Kreuter, *Ludwig-Maximilians-Universität München*
Omer Reingold, *Stanford University*

The gold-standard approaches for gleaning statistically-valid conclusions from data involve random sampling from the population. Collecting proper random samples, however, can be challenging, so modern statistical methods, including propensity score reweighting, aim to enable unbiased inference when random sampling is not feasible. We put forth a new approach for making inferences based on data from a source sample that may differ in composition in unknown ways from an eventual target population. Whereas propensity scoring requires a separate, target-specific estimation procedure for each different target population, we show how to build a single estimator that allows for efficient and accurate estimates on any downstream target data. We demonstrate, theoretically and empirically, that our target-independent approach to inference, which we dub universal adaptability , is competitive with target-specific approaches that rely on propensity scoring. Our approach builds on a surprising connection between the problem of inferences in unspecified target populations and the multi-calibration problem, studied in the burgeoning field of algorithmic fairness . We show how a post-processing algorithm from the multi-calibration framework can be employed to yield valid inferences from a single source sample across a diverse set of target populations.

But What's My Margin of Error? Estimating Sampling Variance When Blending Nonprobability and Probability Samples

Michael Jackson, *SSRS*

The low cost of nonprobability sample sources has prompted continued research into methods of obtaining reasonable population inferences from such samples. A key challenge is that selection into nonprobability samples is related not only to demographics, but also to behavioral and attitudinal characteristics for which external benchmarks do not exist. "Hybrid" samples, blending data from probability and nonprobability sources, offer one potential solution that balances the low cost of nonprobability sampling with the rigor of probability sampling. Such designs typically use the probability-based portion of the sample to produce "internal benchmarks" for non-demographic characteristics that are correlated with selection into the nonprobability sample. The combined probability-nonprobability sample is then weighted to these internal benchmarks in addition to traditional demographic benchmarks. This approach has been

shown to reduce selection bias in some circumstances. However, an unanswered question is how practitioners can obtain reasonable estimates of sampling variance from hybrid samples. The critical feature for reducing selection bias in such designs—weighting to benchmarks obtained from a portion of the sample itself—adds a source of variability that may be difficult to account for using methods that were originally developed for probability-only samples. Using simulation methods and real-world hybrid datasets, this presentation will assess the accuracy of standard errors and confidence intervals produced for hybrid samples using existing off-the-shelf variance estimation methods. Simpler methods (adjustment by an unequal weighting effect and Taylor linearization) will be compared to more rigorous but harder-to-implement methods (replication). Sensitivity to key study design features (the choice of weighting variables and the size of the probability portion of the sample) will be assessed. Results will provide insight into whether and how traditional measures of sampling variability can be applied to hybrid samples, and thus assist practitioners in weighing the benefits and challenges of such designs.

Transforming Communities through The American Rescue Plan Act: Involving Stakeholder Opinion in COVID-19 Pandemic Recovery and Revitalization
Parker Quinn, *Polco*
Erin Caldwell, *Polco*

In March 2021, the federal government funded the American Rescue Plan Act (ARPA), a 1.9 trillion aid package that provides financial aid to families, governments, businesses, schools, and nonprofits impacted by the COVID-19 public health crisis. As part of this package, every local government is provided funding to help with COVID-19 recovery and the Treasury has urged state and local governments to engage their constituents in developing plans to use these payments to catalyze broader economic recovery and rebuilding. This presentation will meta-analyze resident surveys of more than 100 communities conducted between 2019 and 2022 to demonstrate the impacts of COVID-19 on more than 40 aspects of community livability including housing, employment, health care, food security and more. Areas considered most important for government intervention will be identified by comparing both stated and revealed importance analyses.   Researchers also will demonstrate the importance of the use of the surveys for ARPA program evaluation and Treasury compliance through an analysis of more than 20 local resident and business surveys hosted in 2021 and 2022.

Equity in COVID-19 Vaccination and Seroprevalence Rates
Ronaldo Iachan, *ICF*
Kristie Clark, *CDC*
Yun Kim, *ICF*
Adam Lee, *ICF*
Lee Harding, *ICF*

Coronavirus (COVID-19) remains a serious problem in the United States with more than 45 million cases and 743,000 deaths reported through October 29, 2021. The epidemic has had a disproportionate effect on communities of color and persons with lower levels of socioeconomic status, and has spread from larger metro areas to smaller cities and then to rural communities. This presentation examines potential disparities in both the disproportionate prevalence of COVID-19 as well as COVID-19 vaccination rates among racial and ethnic minority groups and groups with lower levels of socioeconomic status (SES). We analyze cumulative data from our repeated cross-sectional seroprevalence study of residual sera from national commercial clinical laboratories using data aggregated over 15 rounds of antibody testing done every two weeks and across all states. Further, we merge data on demographic and socio-economic status (SES) characteristics at the county level from the American Community Survey (ACS) separately with a) daily and weekly vaccination data reported by states to the CDC, and b) our aggregate seroprevalence data. In addition to presenting bivariate analyses of vaccination rates and seroprevalence by demographic and SES measures, we will discuss our development of mixed multilevel models for seroprevalence that account for county-level variations in these measures. We conclude with a discussion of how these models can suggest which groups are subject to the greatest disparities in seroprevalence and vaccination.

A Bayesian Model for Inference on Multiple Panel Public Opinion Surveys
Brittany Alexander, *Texas A&M University*

Arnold Vedlitz, *Texas A&M University*

Multi-wave surveys, which track individual's opinions over time using multiple administrations of the same survey, are common. Since they involve repeated measurements, the results from one survey date, or wave, are correlated with the next wave. However, combining information from two multi-wave surveys with different numbers of waves or dates administered is non-trivial. We present a case study using Bayesian inference to combine two panels about terrorism policy from 2016. The first panel was a large 1730 individual two-wave probability-based panel with dropouts taken six months apart in May 2016 and November 2016. The second panel was a non-probability panel that had six waves taken every month from June 2016 to November 2016 and had 779 respondents also with dropouts and includes a 108 person replenishment sample. We present an extension of multilevel regression with poststratification to model this data set, with an additional level of partial pooling across time and a multivariate likelihood for the repeated measures. We find this model produces more precise population estimates at individual time points without sacrificing the accuracy of predictions of individuals

May 13, 2022 | 9:30 am – 10:15 am | Poster Session #3

Impact of Processes of Social Change on Central American Migration to the USA

Clemente Quinones, *Georgia Gwinnett College*

Over the decades, several million Central American migrants have sought opportunity, refuge, and stability in the United States. Over the past 30 years, for example, this migration increased by 137 percent between 1990 and 2020, from 6.82 million to almost 16.2 million (UNDESA, 2020). Just between 2018 and 2021, we have observed a series of Central American migrant caravans bursting into Mexico with the purpose of crossing the Mexico-USA border to get to the USA. The conventional scientific views tell us that this migration is either a response to development disequilibria (economic problem) or a function of static 'push' and 'pull' factors (de-Hass 2021). However, migration is a phenomenon that cannot be seen in isolation from the broader processes of social adjustment, natural and anthropogenic climate change, and governance failure. The aim of this paper is to analyze the extent to which these broader processes of change and governance failure have affected regional migration. To test this hypothesis the author uses data from the Latin American Public Opinion Project (LAPOP) and one of the Emif data base: Migrantes Procedentes del Sur con Destino a Estados Unidos, 2020 or Migrants Coming from the South with the USA as Destination, 2020 .

Leveraging Surveys and Platform Data to Understand the Behaviors and Attitudes of U.S. Adults on Twitter

Colleen McClain, *Pew Research Center*
Regina Widjaya, *Pew Research Center*
Gonzalo Rivero, *Pew Research Center*
Aaron Smith, *Pew Research Center*

Social and computational scientists alike have long focused on Twitter and its users in studying online discourse, social dynamics, and political participation, among other societal phenomena. While roughly a quarter of Americans use Twitter, previous research shows a small group of highly active users produce most content from U.S. adults on the platform. Studying the ways in which these highly engaged tweeters differ from other users is an important research task. This research provides new insights into these differences using nationally representative self-report data directly linked to survey respondents' digital traces. To explore these issues, Pew Research Center conducted a survey of U.S. adults who use Twitter and asked respondents for consent to share their Twitter handles for research purposes. Results from the combined analysis reveal a familiar theme – a minority of Twitter users produce the vast majority of tweets – and probe deeper to explore the differences between those who tweet at "high volumes" and other users. Key differences include their motivations for using the site, their views of major problems on Twitter and how Twitter has impacted them personally, and their experiences on the platform.   More broadly, the research discusses Twitter users' views on how the platform has impacted them on a personal and societal level, as well as the extent to which users

themselves have encountered harassment or misleading information on the platform. In line with other work suggesting Americans struggle with understanding their privacy online, it also reveals the extent to which Twitter users are (un)aware of their privacy settings by pairing self-reports with ground truth data from their accounts. Results are discussed in light of the broader social media research environment and new directions for blended data research.

Measuring and Modeling Neighborhoods

Cory McCartan, *Harvard University*
Jacob R. Brown, *Harvard University*
Kosuke Imai, *Harvard University*

With the availability of granular geographical data, social scientists and practitioners are increasingly interested in examining how residential neighborhoods are formed and how they influence attitudes and behavior. To facilitate such studies, we develop an easy-to-use online survey instrument that allows respondents to draw their neighborhoods on a map. We then propose a statistical model to analyze how the characteristics of respondents, relevant local areas, and their interactions shape subjective neighborhoods. The model also generates out-of-sample predictions of one's neighborhood given these observed characteristics. We illustrate the proposed methodology by conducting a survey among registered voters in Miami, New York City, and Phoenix. We find that across these cities voters are more likely to include same-race and co-partisan census blocks in their neighborhoods. Net of other factors, White respondents are 6.1 to 16.9 percentage points more likely to include in their neighborhoods a census block composed entirely of White residents compared to one with no White residents. Similarly, Democratic and Republican respondents are 8.6 to 19.2 percentage points more likely to include an entirely co-partisan census block compared to one consisting entirely of out-partisans. We also show that our model provides more accurate out-of-sample predictions than the standard distance-based measures of neighborhoods. Open-source software is available for implementing the proposed methodology.

Satisfaction With Social Support, Public Spaces, and Virtual Socialization During the COVID-19 Pandemic Among LGBTQ Young Adults

Elisa Borowski, *Northwestern University*
Amanda Stathopoulos, *Northwestern University*

MOTIVATION Socialization is important for maintaining social support and coping capacity, but physical distancing measures during the coronavirus (COVID-19) pandemic have impacted face-to-face interactions. Beyond the pandemic, lesbian, gay, bisexual, transgender, and queer (LGBTQ) individuals have been at greater risk for social isolation and lower quality of mental health. Traditionally, the LGBTQ community has relied on physical leisure spaces for safe social gatherings and identity development, but throughout the pandemic, many of these spaces have moved to the virtual realm. RESEARCH QUESTIONS This research examines changes in social support satisfaction, feelings of welcomeness in outdoor public spaces, and barriers to virtual socialization experienced during the pandemic, focusing on marginalized groups. This investigation begins to shine light on pandemic-related mobility shifts from in-person to virtual and their impacts on social network expansion, contraction, and maintenance. METHODOLOGY Data was collected in February 2021 using a web-based survey deployed to 313 individuals of ages 18 to 34 years in California, Texas, and Illinois. This mixed methods survey was designed to quantify changes in strong and weak social ties prior to and during the pandemic using a retrospective egocentric social network name generator and place generator, respectively. Attitudes were measured using 5-point Likert-like scales and open-ended free response questions. Data analysis is performed using t-tests and latent class analysis to examine differences in pandemic-related experiences for marginalized sociodemographic groups. FINDINGS Results from a t-test analysis comparing groups indicate that LGBTQ respondents experienced significantly worse impacts to sleep, mental health, and familial relationships during the first year of the pandemic compared to non-LGBTQ respondents, as well as significantly worse satisfaction with received social support and available outdoor public spaces. Initial results from a latent class analysis support this finding and suggest greater social support satisfaction for heterosexual, white males during the pandemic.

Practice of GIS Applications to Estimate Land Usage- Qatar 2021 Agriculture Census.

Elmogiera Elawad, *Qatar University*
Mohamed Agied, *Social and Economic survey research institute at Qatar University (SESRI)*

Lina. Bader, *Social and Economic survey research institute at Qatar University (SESRI)*
Semsia Al-Ali Mustafa <semsia@qu.edu.qa, *Social and Economic survey research institute at Qatar University (SESRI)*
Rima Charbaji Elkassem, *Social and Economic survey research institute at Qatar University (SESRI)*

Practice of GIS Applications to Estimate Land Usage- Qatar 2021 Agriculture Census. The Social and Economic survey research institute at Qatar University (SESRI) conducted a census of agriculture 2021.The census was planned to improve the techniques of measurement, counting and data collection, more than 100 enumerators and field supervisors was assigned to performed field operations from Jan 2021 to July 2021   This paper aimed to compared the measure method used in the 2000 Census of Agriculture and 2021 census in terms of cost, speed and accuracy   The results showed that the new GIS technique used in the 2021 Census had the fastest completion times and immediate result manipulation. The imputation of GIS developed as tools to measure land usage in the census achieved the goals set forth during Census planning,

Improving Physician Survey Response Rates: Compelling Evidence for Non-Contingent Incentives
Emily Hachey, *Social Science Research Solutions (SSRS)*
Eran Ben-Porath, *SSRS*
Gillian SteelFisher, *Harvard TH Chan School of Public Health*

The field of survey research is continuously looking for new ways to increase response rates in order to improve data quality along with productivity. Building on previous research, this poster focuses on one feature of mailed survey invitations that has the potential for this type of improvement: the use of non-contingent incentives. Non-contingent incentives refer to incentives offered in the invitation to participate in a survey, irrespective of whether the invited person completes the survey or not. By contrast, contingent incentives are furnished to respondents only after they complete the survey, and are typically larger than non-contingent incentives.   The three-month fielding for survey of primary care physicians (PCPs) in the U.S. (n=397) provided an opportunity to test the effect of these two incentive types on response rate. For this study, we mailed survey invitation packages to named physicians. The findings we report are based on an experiment (n=1,125) embedded in that process, which tested the effectiveness of two incentive types on reaching higher response rates : a $25 cash, non-contingent incentive; and a $75 contingent incentive (provided by check or electronic gift-card upon completion). The findings were statistically significant and substantively clear: smaller, non-contingent incentives yielded much higher response rates, and did so at a lower cost-per-complete. Respondents who received non-contingent incentives were also more likely to complete the survey by hardcopy. This poster will present data on response rates based on contingent and non-contingent incentives, demonstrate demographic differences between those who completed the survey in each group, and consider the circumstances most appropriate for including non-contingent incentives in invitation packages.

Nothing but the Truth: a Systematic Review on Socially Desirable Responding in Surveys
Emma Zaal, *University of Groningen*
Yfke P. Ongena, *University of Groningen*
John. C.J. Hoeks, *University of Groningen*

 Socially Desirable Responding (SDR) - presenting a better version of yourself by providing biased self-reports - is inevitable in interactions about behavior. SDR creates serious problems for communication research that uses interviews and questionnaires. The phenomenon can significantly affect study outcomes by making respondents' answers to self-report items diverge towards perceived social norms, and by increasing occurrences of non-response (leaving questions unanswered). For instance, people have a tendency to overreport healthy behavior such as exercising and to underreport unhealthy behavior like substance use. It is generally recognized that SDR is a serious concern that needs to be dealt with urgently. However, to date we are still in the early stages of unraveling the mechanisms underlying SDR and the complex interplay between characteristics that play a role in it. In addition, we currently do not know what the most optimal ways are to measure, predict and reduce SDR in survey research.   Before attempting to control for and reduce SDR, it is necessary to find out what is known about the factors leading to misreporting and non-response due to SDR. Therefore, a systematic review was carried out that collected and investigated recent research literature published on SDR. The review provides a shortlist of the factors that are currently known to predict SDR best. This set of factors includes individual characteristics (e.g., personality, age), aspects of the

specific topic under discussion (e.g., exercising and alcohol consumption), question wording manipulations (steering respondents away from SDR), and context features (e.g., mode of administration). In sum, the review provides 1) a general overview of factors related to measuring, predicting and reducing SDR; 2) theoretical and practical knowledge gaps that still exist in relation to studying SDR and 3) gives recommendations on what types of studies are needed to help capturing SDR and reducing its impact.

Impression Management in a Health Survey: Questionnaire, Personality, and Cultural Trait?

Evgenia Kapousouz, *University of Illinois - Chicago*
Timothy Johnson, *University of Illinois at Chicago*
Allyson Holbrook, *University of Illinois at Chicago*

Researchers often need to rely on survey data to explain causal inferences or predict future behaviors. There is the underlying assumption that respondents answer truthfully, and the data are accurate (Green, 1996). Although this is true most of the time (Fowler, 2009), they should be aware when the data are more likely to have measurement error. One source of measurement error is social desirability bias. Previous research has shown that social desirability bias is a questionnaire characteristic, meaning that the topic or the wording of the question may influence respondent answers. Some research has also shown that social desirability bias may be a personality trait, meaning that some respondents are more likely to provide socially desirable answers than others when confronted with the same stimuli. It is also considered a cultural trait because for two main reasons: (1) socially desirable behavior differs by culture and (2) some races/ethnicities may be more concerned about the impression they make. The current presentation will examine whether social desirability is a questionnaire, personality, or cultural trait or all three of them. We will use a face-to-face survey conducted by Survey Research Laboratory in Chicago. The survey used a stratified sampling by race/ethnicity using four groups: White non-Hispanic, African American non-Hispanic, Mexican American, and Korean American. The survey's main goal was to investigate how cultural orientation affects survey response behavior in a culturally diverse sample using health questions. Using an SEM model, we plan to investigate whether questionnaire, personality, and/or culture affect impression management and if impression management affects latencies and the number of questions respondents refused to answer based on question sensitivity. A better understanding of social desirability bias will help researchers in the future to appropriately design questionnaires, being mindful of the wording of the questions and the target population.

Facing COVID-19 Challenges: Assessing Impact of Change in Data Collection for a Mixed-Mode Survey of Behavioral Health Enrollees

Felix Rodriguez, *Washington State University*
Rose L. Krebill-Prather, *Washington State University*
Kent J. Miller, *Washington State University*

Surveys have been useful in gauging client evaluation of health services. The COVID-19 pandemic in 2020 affected not only the providers and users of health services but also the conduct of survey operations. As an academic survey organization, we report on our experience launching an annual survey of Medicaid behavioral health enrollees in Washington state. The survey has relied primarily on two modes for completion: telephone with web option. The pandemic-related restrictions resulted in the loss of our on-campus pool of student interviewers and staff reduction in our telephone lab. Consequently, we had to shift our data collection strategy to encourage greater use of a mail-in paper survey. To monitor completion, we randomly divided a sample of 17,064 adult enrollees into three data collection strategies. In the first strategy, we offered survey completion by web, mail, or phone; in the second, web and mail; and in the third, web only. The analysis evaluates broadly the shift in our data collection strategy and particularly the effectiveness of offering a mail-in paper survey. We then compare enrollee evaluation of behavioral health services before (2019) and during the pandemic (2020) on measures of access to services, participation in treatment goals, appropriateness and quality of services, general satisfaction, perceived outcome of services, and feelings of being stigmatized, while controlling for survey mode and client characteristics such as gender, age, race/ethnicity, disorder type, and urban or rural residence. The study hopes to generate lessons about the effects of a transformative event, the COVID-19 pandemic, on survey implementation, on one level, and on enrollee assessment of behavioral health services on another level.

Online Learning During the Covid-19 Pandemic: Parental Experiences and Attitudes in the State of Qatar

Haneen Al-Qassass, *The Social and Economic Survey Research Institute (SESRI)*
Engi Assaad Ahmed Elmaghraby, *Social and Economic Survey Research Institute, Qatar University*
Abdellatif Sellami, *Qatar University*
Semsia Al-Ali Mustafa, *Social and Economic Survey Research Institute, Qatar University*

This study examines parental experiences and attitudes toward online learning in the State of Qatar during the Covid-19 pandemic. The data are from the "Qatar Semi Annual Survey" (QSAS) – a telephone survey implemented by SESRI in May 2021 and involving Qatari nationals and resident expatriates with school-aged children. Using a Logistic regression analysis, our findings reveal statically significant associations between parental experiences and attitudes and salient demographic variables, including gender and education level. Parents' nationality and employment status did not emerge as important predictors of their experiences and attitudes during their online learning during the Covid-19. The study concludes with a discussion of the findings and offers some key recommendations for educational policy-making, and provides suggestions for potential research in the future.

Using a Predictive Search Database for Address Collection in Surveys

Shauna Yates, *RTI International*
Jerry Timbrook, *RTI International*
Ashley Wilson, *RTI International*
Shauna Yates, *RTI International*

Collecting accurate respondent addresses is often a critical study activity (e.g., for mailing incentives, conducting follow-ups). However, addresses entered by respondents are not always accurate and can require costly post-survey data cleaning or result in undeliverable mailings. Predictive search databases (PSDs) can autocomplete respondent addresses as they are typed to match entries in the USPS's Delivery Sequence File (DSF), potentially increasing the quality of address data and reducing the need for post-survey data cleaning. However, the data quality and cost implications of using a PSD for collecting addresses is currently uninvestigated. In this paper, we compare the quality and cost implications of collecting address data with and without a PSD. We use address data collected without a PSD from n=1,620 respondents of a nationally representative postsecondary student financial aid study (base-year) and compare it to address data collected with the Experian PSD from the same n=1,620 respondents in a follow-up longitudinal field test study (follow-up). Both studies were mixed-mode web/telephone surveys. We examine three indicators across these two studies: 1) time to complete the address question (in seconds; a proxy for burden), 2) item nonresponse (i.e., any incomplete address fields), and 3) validity of complete addresses (i.e., the number of addresses resulting in undeliverable contacts). Preliminary results indicate that approximately 1.2 percent of addresses collected in base-year without a PSD were returned as undeliverable and required manual post-survey data cleaning by study staff to correct. Conversely, all addresses collected in the follow-up with the PSD were error free and valid (i.e., no undeliverables), and did not require post-survey data cleaning. We conclude the paper with methodological approaches that can improve the functionality of predictive search forms to reduce missingness and decrease cost with improved data quality.

The National Survey of Children's Health: Annual Innovations Following a Major Redesign

Jessica Jones, *HRSA/MCHB/OER*

The National Survey of Children's Health (NSCH) is an annual, addressed based household survey designed to produce national and state levels estimates on the health and wellbeing of children and their families. Following a significant redesign to the survey's sampling frame and mode of data collection from 2012-2016, the NSCH has continued to innovate in each of the five completed cycles of data collection. These innovations, while varied in nature, were designed to increase survey response, lower operational costs, improve data quality, and/or reduce respondent burden. This poster will describe the compendium of NSCH innovations and experiments undertaken from 2016-2020 across the following four categories: sample design, contact strategies, mailed materials and incentives, and questionnaire design. A description of sample design innovations will overview how the creation and refinement of a child flag to identify the household addresses most likely to include children has improved sampling efficiency and reduced costs. Similarly, we will describe experiments to determine the utility of universal versus targeted "web push" recruitment strategies and

reminder postcards. The mixed results of mailed materials innovations, which found that only some color-based envelope enhancements improve response, will be presented. These results will include an overview of the operational constraints which limit the continued use all desired improvements. Additionally, we will describe how non-conditional cash incentives, provided in varied amounts and to the respondent at different points in the data collection timeline, both improve overall response and reduce differential non-response. Finally, we will describe the annual review of the NSCH questionnaires, and improvements made to reduce respondent cognitive burden and completion times.

The Role of Questionnaire Design in Measuring Workplace Discrimination: The Deliberative, Transparent Process Needed to Improve the Rigor of Data Collection

Jessica Starace, *John J. Heldrich Center for Workforce Development, Rutgers University*
Carl Van Horn, *John J. Heldrich Center for Workforce Development, Rutgers University*
Maria Heidkamp, *John J. Heldrich Center for Workforce Development, Rutgers University*
Sean Simone, *John J. Heldrich Center for Workforce Development, Rutgers University*

Producing a high-quality, impactful survey research study typically combines rigorous sampling and administration methods with a purposefully designed instrument that can reliably measure the information sought by researchers. Survey researchers agree on the importance of engaging in a thorough, "informed" questionnaire design process when designing an instrument, but few organizations provide the survey's audience with detailed background about that process when releasing survey data. If the data is only as rigorous or useful as the instrument itself, then engaging in deliberative questionnaire design activities is an essential facet of a high-quality survey research project; that approach likely varies based on the population of interest and research topic. In 2022, the Heldrich Center will collect data describing the perceptions of American workers about workplace discrimination and opportunities for advancement using a nationally representative survey. Heldrich aims to distinguish this project from others by leading an intensive questionnaire design process of this population and on this topic through stakeholder engagement. This project will fill research gaps in knowledge of how inequality manifests within organizations and ways to increase employer strategies to reduce inequalities in the workplace by collecting reliable evidence documenting the perceptions of workers, and do so based on a collaborative, transparent questionnaire design process. Stakeholders from employer associations, employment policy experts, and workers themselves, among others, will be engaged by Heldrich researchers using in-depth interviews, focus groups where feasible, and pre-tests to inform the development of the survey questionnaire. Researchers will document the iterative process by which the project questionnaire is developed through these activities for the project's target population of workers. This effort will contribute to the ongoing conversation of improving rigor in survey research practice, describing the ways in which organizations might increase the legitimacy of their own projects by communicating a deliberative questionnaire design strategy to audiences.

Where Does Japanese Cultural Nationalism Come From: A Mixed-Method Study

Jiangxue Wang, *University of Michigan*

Early research on cultural nationalism mainly focused on state-driven nationalism. State-led nationalism undoubtedly remains essential in analyzing the patriotic sentiment of a nation. However, in an era characterized by national populism and cultural globalization, it is necessary to pay close attention to cultural nationalism derived from cultural products. By analyzing data from the Second to Third Wave of Asian Student Survey, the first part of this paper finds that Japanese students have the strongest tendency only to consume indigenous popular culture. In the meantime, there is a significant positive correlation between the strength of cultural nationalist sentiment and the consumption frequency of Japanese popular culture among Japanese students. The second part of this paper explores three possible explanations for this phenomenon through semi-structured interviews with Japanese students. First, some of the core values of Japanese traditional culture have been internalized into Japanese popular culture. These core values are considered to be the reason why Japanese popular culture is recognized worldwide. This recognition reinforces Japanese cultural confidence. Second, in terms of American popular culture, the cultural nationalist sentiment among Japanese students seems to be triggered by the discomfort with American cultural dominance. Third, Japanese students tend to feel a strong "sense of distance" when consuming foreign popular culture. One may argue that this "sense of distance" comes from a strong sense of "cultural uniqueness" among Japanese youngsters. In conclusion, the study presents in this paper shows that although there is a significant positive correlation between the consumption

frequency of indigenous popular culture and the degree of cultural nationalist sentiment among Japanese youth, the causality between the two events needs to be addressed more specifically in future studies.

The Shy Respondent and Propensity to Participate in Surveys and Voting
James Dayton, *ICF International*
Randy ZuWallack, *ICF*
James Dayton, *ICF*

The term "shy voter" emerged from British electoral politics as a potential explanation for underestimates of the Trump vote in 2016 and 2020 pre-election polls. The hypothesis that voters most supportive of Trump were underestimated in polls has not been demonstrated. This study explores whether a segment of the population wants to stay out of sight (shy respondents) and how this affects their likelihood of survey participation and voting.   We used a commercial non-probability panel which relies on financial incentives not social utility to obtain respondents, and neither survey topic nor sponsorship was revealed, to maximize the likelihood of representing "shy" respondents within a demographically representative national sample of nearly 2,000 adults. Our focus is on the relationship of "shyness" to survey participation and voting, not population estimates.   We used questions about preferring "to stay out of sight and not be counted" and alternately showing "I am proud of who I am" to construct a survey "shyness" scale. Nearly a quarter of this sample (24.4%) both agreed with the former and disagreed with the latter statements (score 6-8). We found this shy respondent measure was related to stated willingness to participate in future surveys, as well past participation in surveys. Although shy respondents were less likely to vote than others, a majority "always" vote in Presidential elections. Indeed, in our sample, shy voters represent 20% of those who always vote in Presidential elections, 17% in Congressional elections, and 16% in party primary elections. The relationship between shyness and demographics is generally slight and usually non-significant. Hence, demographic weighting is unlikely to correct for underrepresentation of the shy respondent.   These findings suggest further research on "shy respondents" may improve understanding of pre-election polling errors and motivations for survey participation.

Pandemic Pivot: The Shift From CAPI to CATI, and Single-Device vs. Multi-Device Approaches
Melissa Kuenzi, *Dobility*
Dr. Christopher Robert, *SurveyCTO*
Amrik Cooper, *SurveyCTO*

During the ongoing Covid-19 pandemic, researchers pivoted in significant numbers from in-person survey modes, like CAPI, to computer-assisted telephone interviews (CATI), as it enables safe interactive interviews between enumerators and respondents. As longtime providers of a CAPI platform with basic multi-device CATI capability, we observed this disruption in our industry from a unique vantage point. Drawing from SurveyCTO user data based on platform usage, interactions with users, and survey results, we explore the rise of CATI in 2020-21 and the challenges of pivoting from CAPI during the pandemic. We also dive into a comparison between different methods for CATI. We analyze the "multi-device" method of enumerators calling respondents on one device while recording answers on another vs. the decentralized, single-device method we designed with our smartphone- and tablet-based users in mind. We share evidence for both methods based on SurveyCTO user data, and discuss the advantages of each for researchers. We present these findings on the impact of Covid-19 on survey research modes in the context of a world still being shaped by this ongoing global crisis.

Conspiratorial Thinking, Political Rhetoric, and Views Toward Immigrants in the U.S.
Melissa Shannon, *NORC at the University Of Chicago*
Claire Krummenacher, *NORC*
Emily Swanson, *Associated Press*
Hannah Fingerhut, *Associated Press*

Immigration continues to be a priority issue for many Americans and one that has been linked to rising racial, cultural, and political polarization within the American electorate. Drawing upon the political rhetoric on immigration, The Associated Press-NORC Center for Public Affairs Research conducted a nationally representative survey of American adults with 4,000 AmeriSpeak and opt-in panelists to understand the relationships between geography, partisanship,

beliefs about immigrants, and support for different immigration policies. We measure the pervasiveness of replacement theory and investigate the social and political factors that contribute to belief in conspiracy theories on immigration. We also examine how racial resentment, engagement in conspiratorial thinking, and contextual factors like increasing immigrant populations moderate the effects of partisanship on views toward immigrants.

Impact of COVID-19 on Attitudes Towards Flu Vaccination

Melissa Dunn Silesky, *Lines for Life*
Chelsea Barth, *The Public Good Projects*
Erika Bonnevie, *The Public Good Projects*

Introduction: Seasonal influenza affects millions of people across the United States each year. The purpose of this paper is to understand the impact of COVID-19, including the introduction and authorization of the COVID-19 vaccinations, on attitudes towards flu vaccination. Methods: By May 2022, we will have results of three cross-sectional surveys conducted before April 2020 that exclusively measured attitudes towards flu vaccination, and three surveys since October 2020 that measured attitudes towards both flu and COVID-19. All surveys were fielded in the 9 Kaiser Permanente (KP) states and 6 non-KP states. Results from the surveys were analyzed using IBM SPSS Statistics and RStudio Statistics software. Data were compared using a 2-sided Pearson Chi-squared test and multivariate logistic regression analysis where appropriate. Significance was determined at p≤0.05. Results: Results will capture two aspects of our study: (1) changes in measures that have been asked in all surveys from pre- to during- the COVID-19 pandemic, and (2) changes in measures about both flu and COVID that have been asked since October 2020. Since the start of the COVID-19 pandemic, the percent of respondents in both KP and non-KP states that received or intend to receive the flu vaccine has increased. However, we have seen changes in motivating factors to receive the flu vaccine; fewer respondents said they would get the flu vaccine to protect "friends and family," and more respondents said they would get the flu vaccine if "everyone else was getting it." We will also present patterns found on measures that examine both COVID-19 and flu vaccination, such as "Getting a flu shot helps to protect healthcare resources, which might be stretched thin this flu season, due to coronavirus."

State-Level Shifts: How Socioeconomic and Demographic Change Influence Public Opinion on Immigration in a Diversifying America

Michael Lenmark, *Stony Brook University*

Past contextual research on public opinion and immigration has emphasized the influence of local socioeconomic and demographic changes on immigration attitudes. However, this body of research has neglected the study of such changes at the state-level. This gap in the literature is surprising, given that states are increasingly bearing the costs of immigration enforcement and are passing their own immigration policies to bypass federal congressional deadlock. Further, much of this research focuses only on immigration attitudes of White Americans, in part due to a lack of representative data. This characteristic omission leaves us with an incomplete picture of a deeply divisive political cleavage in a rapidly diversifying country. To address these outstanding issues in the literature, I employ a longitudinal multilevel design to leverage the Cooperative Congressional Election Study's unique representativeness in each of the fifty U.S. states. I incorporate both individual-level and contextual state-level data to examine how socioeconomic and demographic changes over the past ten years have influenced public opinion on pathways to citizenship, border security, immigration enforcement, and immigrant employment across a diverse and representative sample of the United States

All Scales Are Not Equal: A Cross-Cultural Comparison of Bipolar and Unipolar Scales

Randal Ries, *IBM*
Phyllis Mancino, PhD, *IBM*
Peter Guzewich, PhD, *IBM*

U nipolar and bipolar scales are widely used to capture attitudes and opinions in surveys; however, s cale polarity (unipolar and bipolar) and its effects are often overlooked, particularly in the context of cross-cultural research.   R esponses to a series of agreement/opinion statements on an employee survey collected from eight countries (Brazil, China, France, India, Italy, Japan, UK and U.S.) were analyzed. The countries selected are known for their extreme

response styles in survey taking based on Gert Hofstede's seminal research conducted on IBM employees (Hofstede, 1991).   Our overall finding is consistent with a recent analysis of survey responses collected on b ipolar and u nipolar scales published in the Social Science Computer R eview . This paper concludes that respondents due to inconsistent responses the scales should not be viewed as interchangeable (Hohne, Krebs & Kuhnel, 2020).   The data analyzed in the Hohne research was collected on 5-point Likert type scales while our employee data was collected on 11-point scales. Our analysis indicates that uni polar scales result in data th a t a re less negatively skewed, have a higher overall mean, and less variance than a bipolar scale . In this research, bipolar scales yield lower proportions of agree responses while the Hohne study finds bipolar scales have higher agree proportions. This difference between the two studies can, at least in part, be attributed to the difference in total response options (5 vs 11). Our r espondents were more likely to select 1, 2 , or 3 (disagreement) on bipolar scales than on unipolar scales .   Cross-cultural comparisons indicate r espondents in India and the U.S. are more lik ely to select positive values (7, 8 , or 9) on a unipolar scale than on a bipolar scale. Participants in France in Japan use the scales interchangeably

Is That Your Respondent or Mine? Geographic Accuracy of Listed Cell Phone Numbers in Cross-state Coverage Error
Randal ZuWallack, *ICF International*
Traci Creller, *ICF*
Samantha Collins, *ICF*
Thomas Brassell, *ICF*

The mobility of cell phone numbers is a well-known challenge for random digit dialing (RDD) cell phone samples of specific geographic areas. RDD frames will have overcoverage, reaching people residing outside the geographic area who have an in-area phone number, as well as undercoverage due to being unable to reach people residing inside the geographic area who have an out-of-area phone number. The Behavioral Risk Factor Surveillance Survey (BRFSS), a state-based survey conducted in all 50 states plus the District of Columbia, presents a unique opportunity to study these errors. On one hand, the overcoverage and undercoverage are mitigated because states will "share" any mis-assigned respondents so they can be counted in the state where they live even if they are surveyed by a different state. However, states differ in the optional modules and state-added questions they field, so respondents who are surveyed by another state will only be partially complete by the receiving state's standard. One of the benefits of transitioning the 2021 cell phone sample to Marketing Systems Group's Advanced Cellular Frame (ACF) is better accuracy based on auxiliary information about geographic location for some cell phone numbers, or "listed" numbers. The percentage of listed sample within each state generally ranged between 40-55%. In this multi-state analysis, we examine the accuracy of the geographic information for three groups: 1) listed numbers with in-area telephone exchanges, 2) listed numbers with out-of-area telephone exchanges, and 3) unlisted numbers. We then compare the frame coverage improvements in 2021 as compared to 2020 at a state and county level. Results should help inform best practices for coverage error in RDD surveys.

Mixed Method Research on Barriers to Workplace Inclusion and Employee Retention for First Generation Professionals
Rodney Terry, *U.S. Census Bureau*
Luke Larsen, *U.S. Census Bureau*
Matthew Virgile, *U.S. Census Bureau*
Aleia Clark Fobia, *U.S. Census Bureau*

We present results of qualitative research and a survey of federal employees regarding barriers to workplace inclusion and employee retention, respectively. Barriers to workplace inclusion based on protected characteristics – such as gender and race/ethnicity – persist, despite efforts to reduce them. In a study of whether low socioeconomic status (SES) is also a diversity characteristic for which there may be barriers to workplace inclusion and advancement, qualitative research was conducted with First Generation Professionals (FGPs) from the U.S. Department of Commerce (Terry & Fobia, 2019). Defined as professional employees who are the first in their immediate families to hold a white-collar professional position, results found that FGPs reported experiences and circumstances that reflect a lack of resources that are typically available to people with middle and upper-SES backgrounds. This disparity in resources resulted in experiences of barriers to workplace inclusion and career advancement. However, study participants also identified supporting efforts that would help reduce these barriers, including orientation sessions about workplace

culture, networking skills training, and career advancement counseling. Finally, despite the disparities and barriers associated with being an FGP, many participants also described positive aspects of FGP status. To expand upon this qualitative research and increase understanding of FGPs' workplace experiences and outcomes, we present findings from a survey of employee retention among U.S. Department of Commerce employees. Our analysis compares results between FGPs and continuing generation professionals (CGPs) regarding interpersonal experiences at work, job satisfaction, intent to leave their position, reasons for intending to leave their position, motivations to stay, and related topics. Our analysis also investigates whether results from the previous qualitative research align with the survey data on workplace retention and aims to provide additional insight regarding barriers to workplace inclusion for FGPs. Finally, we discuss research implications and future research directions.

Participation, Completion, and Breakoff in the U.S. Army's Annual Leadership Survey

Ryan Riley, *ICF*
Matthew D. McDonough, *ICF*
Maryann Stassen, *ICF*
Meghan I. Huntoon, *U.S. Army Center for the Army Profession and Leadership (CAPL)*
Jon. J. Fallesen, *U.S. Army Center for the Army Profession and Leadership (CAPL)*
Matt Jans, *ICF*

Members of the US Army, both soldiers and civilian employees, are surveyed frequently. As a result, gaining participation is challenging. Long questionnaire duration and other design features that can lead to breakoff are of particular concern. This presentation reports on response rates, breakoffs, and questionnaire duration in the Army's annual leadership survey conducted in 2020. The survey is part of a recurring (since 2005) study that captures assessments from the field about leadership and leader development to inform senior leaders about upward or downward trends in the quality of leadership, the effectiveness of leader development experiences, and situational factors affecting leadership. The survey sample consisted of nearly 300,000 Army personnel, including officers, warrant officers, and noncommissioned officers (NCOs) in the active and reserve components, and Army civilian employees. The survey protocol included an email invitation and up to four email reminders to complete a self-administered web questionnaire. The overall response rate was 7.7%. This presentation will discuss survey completion per contact, and across the different kinds of survey participants. For example, the invitation message generated 25% of total respondents, with highest participation among Army civilians (33%) and lowest among warrant officers (20%), while, the fourth reminder message was still useful, yielding about 15% of the total respondents. Additionally, breakoff (i.e., leaving the questionnaire and not returning) occurred most often within the first one-third of the questionnaire, and was highest among company grade officers (30%) and junior NCOs (29%), and lowest among Army civilians (10%). Results will be discussed in the context of instrument design to reduce duration and breakoffs, and surveying heavily-surveyed organizations.

Which Populations are Being Measured and Who is Left Out of Online Surveys: Sample and Expenditure Differences in a Multi-modal Household Survey

Samantha Cole, *U.S. Census Bureau*

The Consumer Expenditure Survey provides monthly data on expenditures, income, and demographic characteristics of consumers in the United States. Data are collected from randomly selected addresses by the U.S. Census Bureau for the Bureau of Labor Statistics. There are two key components of the survey: a quarterly interview and a two-week diary recording of household expenses. This research uses data from the diary portion of Consumer Expenditure Survey. Due to the COVID-19 global pandemic, for the first time in the survey's operational history, respondents were offered a choice between an online diary and a printed diary, while previously the paper diary was their only choice. Households were not eligible for the online diary if they lacked cellular or in-home high-speed internet (i.e., they were part of the digital divide), or if they infrequently went online. This paper asks several research questions, including how socio-demographics differ by mode choice and by household eligibility for the online diary. Household socio-demographics include cohort, education, and receipt of government resources such as welfare, food stamps, or reduced rent. The number of expenses and types of expenses will also be also analyzed by socio-demographic characteristics and by modal choice/eligibility. Data for this descriptive research study includes all households eligible for the Consumer Expenditure Diary in a 12-month timeframe (September 2020 through August 2021). Methods include frequencies,

percentages, and cross-tabulations. The discussion section will explore whether household's ineligible for the online diary have distinctive socio-demographic characteristics and/or spending patterns, different from households that had a choice in modes. It will discuss how these differences may translate to some groups being underrepresented in online only surveys

Effects of COVID-19 on a Nationwide Survey of Public Schools

Sarah Heimel, *U.S. Census Bureau - Washington, DC*
Beth Newman Satisky, *U.S. Census Bureau*
Rachel Horwitz, *U.S. Census Bureau*
Josh Neufelder, *U.S. Citizenship and Immigration Services*
Deanne Swan, *National Center for Education Statistics*
Jana Kemp, *American Institutes for Research*

The School Survey on Crime and Safety (SSOCS) was one month into data collection in 2020 when schools across the country had to shift to online learning, due to COVID-19. SSOCS asks schools about crime and safety topics such as the presence of security staff at school, the availability and student use of drugs and alcohol, and the frequency of fights and bullying at school. With schools suddenly transitioning from in-person classes to virtual classes, some survey questions (such as how many crimes occurred at school and how common cyberbullying was among students) had a different meaning for late-responding schools in 2020 than had been anticipated. In an attempt to distinguish what changes in survey estimates could be attributed to measurement error or response bias and what changes were true changes in the survey estimate, we conducted a difference-in-difference analysis of critical survey estimates. This analysis used 2018 SSOCS data in addition to the 2020 data and examined how the nationwide school closures impacted the resulting survey estimates

Cheering for the Home Team: Changes in Attitudes Towards Hosting the FIFA 2022 World Cup in Qatar

Semsia Jatic, *Qatar University*
Abdoulaye Diop, *Qatar University*
Lina Bader, *Qatar University*
Engi ElMaghrabi, *Qatar University*
Haneen AlQassass, *Qatar University*
Buthaina AlKhulaifi, *Qatar University*

In less than a year, Qatar will become the first Arab nation to host the FIFA World Cup, a global event that draws together millions of fans from around the world to support their national teams. While the international community has been vocal about its criticism of the decision to have the event hosted in Qatar due to its poor legacy of migrant worker abuse, little is known about the attitudes of those currently living in Qatar (both Qatari citizens and expatriate residents). The Social and Economic Survey Research Institute (SESRI) recognized this lack of nationally representative data and in 2015 conducted a first-of-its-kind survey of the views and attitudes of Qatari nationals and expatriate residents towards hosting the FIFA World Cup and its perceived socio-economic impacts. A similar and more detailed study followed, conducted in three waves (in 2018, 2019, and 2021), tracking the changes in attitudes over time as the event approaches. Using the nationally representative survey data, this research will examine changes in respondents' attitudes towards a series of issues, including benefits of hosting the World Cup for the nation, economic, social and global image effects, as well as their interest and excitement for the upcoming event.

Are Pre-Notification Letters "Old News"? Evaluating the Impact of Prenotification Letters on RDD Cell Phone Samples

Silpa Sevakula, *ICF*
Samantha Collins, *ICF*
Randal ZuWallack, *ICF*
Traci Creller, *ICF*
Thomas Brassell, *ICF*
Rebecca Wood, *TX DSHS*
Wendi Gilreath, *WA DOH*
Mark Serafin, *WA DOH*

Prenotification letters have the potential to increase response in random digit dialing (RDD) telephone surveys by alerting households to the impending call and making them more likely to answer the phone. Addresses are obtained by matching sampled telephone numbers addresses from telephone directories and other commercial databases. A criticism of prenotification letters is that they only reach a nonrandom subset of sampled numbers and may introduce differential response between those who receive a letter and those who don't, potentially biasing results. Further, the value of prenotification letters for cell phone samples has been limited by low numbers of matching addresses. This study reports on the tracking of pre-notification letters mailed for both the Texas and Washington Behavioral Risk Factor Surveillance Survey (BRFSS). The match rates in Texas and Washington were 15.9% and 14.7%, respectively for statewide cell phone samples selected for conducting the 2020 BRFSS. These address matches represented about 30% of the completed surveys in both states. Changes to the sampling frame used for the 2021 BRFSS resulted in a large increase in the address match rates in the two states, over 50% in each state. The matched addresses now represent nearly 80% of the total completed surveys in the two states. With this drastic increase in address matching, we explore three research questions related to sending pre-notifications. First, we explore whether prenotification letters increase overall response rates to the survey. Second, we examine the cost benefit trade-off of sending prenotification letters. Finally, we compare the demographic differences between completed surveys from the matched addresses and those from the nonmatches.

Measuring Social Interaction with Smartphone Data

Sonja Malich, *Institute for Employment Research*
Sebastian Bähr, *Institute for Employment Research (IAB), Germany*
Georg-Christoph Haas, *Institute for Employment Research (IAB), Germany; University of Mannheim, Germany*
Florian Keusch, *University of Mannheim, Germany*
Frauke Kreuter, *Ludwig Maximilian University of Munich (LMU), Germany; University of Maryland, USA*
Mark Trappmann, *Institute for Employment Research (IAB), Germany; University of Bamberg, Germany*

As a complex and difficult to grasp construct, social interaction is usually measured through observational studies, retrospective self-reports, or diaries. While each of these traditional survey methods has its own advantages, they can be burdensome for participants and are vulnerable to recall bias. Moreover, individuals increasingly interact with each other via smartphones but research often has overlooked this dimension of social interaction so far. Interactions via smartphone are documented automatically in the phone's communication history. Smartphone data provide rich information about social behaviors over the course of a day at a fine-grained level over a long period of time. At the same time, these data overcome limitations of traditional survey methods by eliminating response burden and recall bias. We therefore aim to measure social interaction in new ways using call and text message log data as well as app data. To this end, we use data from the IAB-SMART study which used a smartphone app to collect survey and sensor data from 623 members of the German Panel Study Labour Market and Social Security (PASS) over a period of six months. The potential of these data is further enhanced by the special study design through which the smartphone data can be linked to the PASS and to administrative data on participant's employment biographies. The aim of this study is to demonstrate new ways of measuring social interaction, uncover patterns of social interaction, and compare the identified behaviors across different social groups.

Online Echo Chambers and Opinion Polarization About Algerian Elections: Did the Facebook News and Information Divide the Algerians?

TARIQ SAIDI, *University of khenchela*

Facebook, according to many studies, has been the most important resource of information for numerous Algerians about Algeria's protest movement and through it, the users also got informed about one of the most controversial movement's topics which is the presidential elections, and shaped their opinion towards the participation and voting or the Abstention. This study explores the correlation between Facebook contents and opinion polarization Based on one of the selective exposure and reinforcing model perspectives which is "Echo chambers" that suggests that the tendency for select information that adheres and reinforces the opinions may play a crucial role in shaping users' opinions about the elections during the protest movement and after that. this process of getting information from polarized groups and pages might affect the public debates and discussions between Algerians. This study aims to test the hypothesis that

Facebook's contents (news and information) have contributed to dividing Algerian citizens into segregated clusters which are known as "echo chambers". we are going to analyze massive users' posts on Facebook pages and groups with the comments of users interacting with the elections from the beginning of the Algerian protest movement (Hirak) on 22 march 2019 until the day of the local election was held on 27 November 2021.

Assessing the Impact of Political Ideology on Public Health Surveillance Survey Participation

Thomas Brassell, *ICF Macro*
Dave Roe, *ICF*
Randy ZuWallack, *ICF*
Samantha Collins, *ICF*

Several years ago, research exploring the impact of party affiliation and political ideology on telephone response rates found little to no relationship between the two. The current political climate in the U.S., including strong divisions centered around the COVID-19 pandemic (such as the clear politicization of COVID-19 vaccinations, mask wearing, and other public health measures implemented to reduce the spread of the virus) has called the impact of political ideology on public health surveillance survey participation into question again. Our current efforts explore the impact of environmental political ideology on response to public health surveys over the past five years. Understanding the impact of politicization on public health data collection is immensely important to public health surveillance. Underrepresentation of subpopulations may limit the ability of public health researchers to develop policies and programs geared towards addressing the health needs of these communities. To assess the impact, we combined data from multiple representative state-level telephone surveys from 2016 thru 2021 with county-level voting results from the 2016 and 2020 presidential elections. Using this data, we investigated the relationship between the immediate political environment and telephone data collection performance metrics and whether the existence or strength of any relationships have changed over time. Specifically, we used county-level voting results to assign a political ideology to an area. We then assessed the relationship between assigned political ideologies and telephone survey performance metrics (e.g., response rate, contact rate, cooperation rate, etc.). This presentation will present the results of these analyses while discussing potential methods for mitigating the potential negative impact political ideology may have on public health surveillance

Communicating Mixed COVID-19 Vaccine Messages: How the Development, Production, and Distribution was Characterized in Leading National and Regional News Media

Thomas Christie, *University of Texas - Arlington*
Dr. Charla Markham Shaw, *University of Texas Arlington*

This study examines the nature and impact of the coverage of COVID-19 vaccination issues in leading U.S. and regional news media. The 2020-2021 worldwide outbreak of the COVID-19 virus and subsequent focus on the development, production, and distribution of a vaccine frequently evoked misunderstanding and fear as intense media coverage influenced opinion formation and varying degrees of public acceptance of the vaccine, particularly among underserved populations. As the nature and extent of media coverage precedes the formation of public opinion and public policy during such health crises, this study uses the agenda-setting and agenda-building traditions of communication research as a foundation for understanding the linkage of this COVID-19 coverage to public policy and public opinion. Key research questions guiding this study were: 1. What attributes of the COVID-19 vaccine issue received prominent coverage in leading national/regional media outlets? 2. How did the leading national/regional media coverage of these attributes compare with the issue attributes and levels of concern noted in public opinion polling? This study incorporates past research of mass media coverage of major health issues and their influences on public opinion formation. Results from a quantitative content analysis and a qualitative thematic analysis reveal the unsettled nature of this news coverage and some notable differences in coverage among The Wall Street Journal , The New York Times , and The Dallas Morning News . The analysis reflects on characteristics of communicating fear and uncertainty and notes observed public policy consequences during and following the media coverage of vaccine development, distribution, and acceptance during the COVID-19 outbreak. The analysis provides grounding needed to address health disparities among high-risk and underserved populations, including ethnic minority populations

An Ode to WAPOR: Public Opinion in International Contexts

Change or Continuity? Attitudes and Perceptions About the Reformed Employment System in Qatar

Semsia Jatic, *Qatar University*
Semsia Al-Ali Mustafa, *Qatar University*
Muzna Madeeha, *Qatar University*
John Lee Holmes, *Qatar University*
Kien Trung Le, *Qatar University*

In the last decade, policies managing Arab Gulf countries' migrant workers, commonly known as the kafāla (sponsorship) system, have witnessed both change and continuity. Governmental attempts to reform this highly restrictive employment system, while delicately and prudently managing the many contesting interests, have faced multiple obstacles and varying degrees of success. The changes have been taking place in a direction cautiously applauded by international observers, yet skeptics maintain that across the Arab Gulf states much of the system remains intact. In their view, migrant workers are still legally bonded to powerful employers and laboring under what they characterize as abusive working conditions. Qatar's labor laws in particular have been in a state of accelerated change amidst immense global scrutiny ever since the country was awarded the rights to host the upcoming 2022 FIFA World Cup. Given the unsettled state of Qatar's labor laws and the debate surrounding its treatment of migrant workers, this paper examines how changes in the employment system shape the lives, perceptions and attitudes of those most affected, i.e. the migrant workers and their Qatari employers. Our assessment of these changing perceptions and attitudes is based primarily upon a nationally representative telephone survey of 2,760 respondents (857 Qatari nationals, 1,012 high income and 891 low income expatriates) conducted from September 2020 through January 2021. Understanding public attitudes is key to determining whether progress has been made and to identify remaining challenges that impede progress towards realistic goals. Scholars assert that any successful societal transformation would be difficult to bring about without sufficient public and corporate support. Therefore, understanding continuity and change in the perspectives of the local population (Qatari citizens and resident expatriate workers) with respect to reform of the labor laws is fundamental.

Attitudes and Opinions on Aid and Poverty Reduction; Mukuru Slum - Nairobi, Kenya

Angela Pashayan, *Howard University*

The poorest of the poor live in challenging settings with logistical problems for conducting research. Therefore many development programs, including those backed by the US State Department, are targeted to reach the top-tier of the lower-class and the lower tier of the middle-class, leaving the poorest of the poor neglected. This research gathers public opinons directly from the poorest of the poor, residents of Nairobi's second-largest slulm (Mukuru), population estimated at 825,000. Opinions are collected through in-person interviews and focus groups using CAPI methodology and cultural sensitivity. The data serves as a rare look directly into the lives of slum dwellers. This work is intended to extend the demographic reach of poverty research to the income level that is often without voice.

Populism in Eastern Europe. New Types of Populism - the Romanian Case

Dan Sultanescu, *Center for Civic Participation and Democracy, SNSPA, Romania*
Dana Sultanescu, *CPD SNSPA*

We assume that there is a different impact on growing populism in a European country like Romania, in an electoral context (like 2019, with two rounds of elections – European and Presidential) and post-covid context (2021-2022). Different academic studies of populism identify two important ideas to define populism: (1) "the people" is the main source of legitimacy, and the politicians must be accepted as serving the people, and (2) the regular people and the politicians elite are two very different groups, with the public being much more fit to rule and govern than the corrupted elite. There are a lot of new measurements of populism, using these approaches, and our idea is to replicate this in a country like Romania. We intend to measure populism with two questions: 1) ordinary people would do a better job/do no better solving the country's problems than elected officials and 2) most elected officials care/don't care what people

like me think. These measures are used by scholars studying populism to capture attitudes about an antagonistic relationship between elites and the people (Stanley, 2011; Spruyt et al. 2016; Schulz et al. 2017). Romania is a relatively new democracy, with less civic participation and a lot of traditional values, and society strongly affected by internal and external struggles (the effects of anti-Western propaganda, the crisis in the European Union etc). In Romania, the populist threat should be much more visible and more active than in other European countries. Because of that, we are interested in adding a third variable when measuring populism (3) the level of rejecting all politicians as corrupt and as people that should be sentenced to prison. Using data from 7 surveys from 2019-2022, we will segment the public into different categories, and see if there are different types of populism.

Afghanistan: Afghan Public Opinion Before and After the Return of the Taliban
Matthew Warshaw, *D3 Systems, Inc*

Afghanistan fell to Taliban control on August 15, 2021. This has had a profound impact on the people of Afghanistan. ACSOR Surveys and D3: Designs, Data, Decisions have been polling in Afghanistan since 2003 and the return of the Taliban regime has impacted our polling operations. At the writing of this abstract, we have shifted our polling to CATI only to protect our interviewers, and continued our "Afghan Futures" polling series. We have data from prior to and after the fall of the U.S. backed Afghan government and we intend to collect more data before the AAPOR conference in May 2022. We propose to present a current view of conditions in Afghanistan for average Afghans related to future of their country. We will look at the results from our national surveys and discuss what comes next for 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 over ten national face-to-face surveys. We have presented findings from our surveys at AAPOR in the past and believe now is an important time to present again as Afghans chart their way under the new Taliban government.

Evaluations of Democracy through Citizen's Conceptualization Lens
Carmen Le Foulon, *Centro De Estudios Públicos*

Although most citizens value democracy as a form of government, they are highly critical of its functioning (e.g. Norris, 2011). This research seeks to shed light on this tension by exploring how individuals conceptualize democracy - procedures, political rights, social benefits (e.g. Dalton, Shin and Jou, 2007) and the interrelation between conceptions of politics and democracy as how the shape citizen's understanding of the role of representatives. It proposes that evaluations of democracy are associated with distinct patterns of association between the different conceptualizations. This study relies on data from a non-probability two-waves survey (recruited through Netquest) that includes Chilean residents from all socioeconomic aged 18 and above.

The Longest War: Examining American Public Opinion Towards The War In Afghanistan (2001 – 2021)
Ryan Tully, *Ipsos*
Randall K. Thomas, *Ipsos*

On October 7, 2021, the United States and its allies began their invasion of Afghanistan as a direct response to the September 11th terrorist attacks. Over the next two months, the U.S. and allied forces routed Taliban government forces, as well as al-Qaeda militants, throughout the country. On November 13th, 2001, the U.S. and allied forces took the capital city of Kabul in a siege lasting less than 24 hours. What appeared to a decisive victory in the war was actually a prelude to a nearly 20-year stalemate. On August 15th, 2021, the war ended by coming full circle with Taliban forces taking Kabul and regaining ruling power in Afghanistan. This study examines public opinion towards the War in Afghanistan over its 20-year duration. Using historical polling data, the study shows widespread public support at the start of the war. Public support for the war was so high that polling firms largely ceased asking about it by 2003. The study shows that when substantive polling on the war resumed in the late 2000s, support for the war had polarized largely along political party lines. The study concludes by looking at polls conducted on the Ipsos KnowledgePanel® immediately before and after the withdrawal of U.S. forces in Afghanistan. The study will analyze this cross-sectional data to determine if previous widespread support for the withdrawal was impacted by the Taliban's dramatic return to power. The study will further analyze these results based on political ideology and party affiliation.

The Role of Information & Faith in Democracy: Celebrating the 10th Anniversary of the AP-NORC Center for Public Affairs Research

Faith or Fraud: Americans' Confidence in Their Electoral System

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

The issue of electoral integrity has taken center stage in American politics. Following Donald Trump's election in 2016, the president declared that millions of illegitimate ballots were cast. Throughout his presidency, prominent Democrats raised concerns that Russian interference in the 2016 election helped him to victory with the possibility of further tampering in 2020. Then, after his defeat in 2020, Trump again cast doubt on the veracity of the results, culminating in the January 6 th insurrection at the U.S. Capitol. In the middle of this environment, COVID-19 introduced expanded access to mail-in voting, and with it came additional claims by prominent politicians about the accuracy of electoral outcomes. But how do most Americans feel about their country's elections? Using AP VoteCast and other surveys, The Associated Press-NORC Center for Public Affairs Research has conducted a series of studies over the past several years exploring Americans' faith in their electoral systems. The findings paint a grim picture: many Americans lack faith in the integrity of U.S. elections, especially at the national level. A sizeable minority has little or no confidence in the accuracy of vote counts. They express concern about the interference of foreign governments and the vulnerability of voting systems to tampering. It is a partisan issue, as well: on the one hand, Democrats worry that those who are eligible to vote won't be allowed to, and on the other hand, Republicans think people who shouldn't be voting will have their ballots counted. Using these AP-NORC surveys, this presentation will analyze trends and demographic differences in faith in the electoral system to explore the depths of America's confidence crisis.

Important Information on Misinformation: Exploring the American Public's Attitudes toward the Spread of Misinformation

David Sterrett, *NORC at the University Of Chicago*
Claire Krummenacher, *NORC at the University of Chicago*
Jennifer Benz, *NORC at the University of Chicago*
Trevor Tompson, *NORC at the University of Chicago*

The spread of misinformation has had significant effects in recent years on American politics and society. Researchers, policymakers, and the media continue to try to better understand both the causes and consequences of the growing spread misinformation. A key part of addressing this issue is exploring public behaviors and attitudes toward the spread of misinformation. A September 2021 Pearson Institute/AP-NORC Poll provides strong evidence of the wide public concern with the spread of misinformation. The nationally representative survey with 1,071 adults was conducted by the Pearson Institute for the Study and Resolution of Global Conflicts at the University of Chicago and The Associated Press-NORC Center for Public Affairs Research, using AmeriSpeak®, the probability-based panel of NORC at the University of Chicago. The poll shows that 95% of Americans say the spread of misinformation is a problem when it comes to getting information about current events and important issues, including 81% who say it is a major problem. The research provides insights into what demographic groups are more or less likely to believe they have been exposed to misinformation or have themselves spread misinformation. In addition, the survey highlights which domestic and foreign actors people believe are responsible for spreading misinformation and what actors Americans say are most responsible for addressing the problem. The findings from this study offer valuable insights for researchers and policymakers about how the public views the spread of misinformation and who they believe should address this major problem.

Eroding Faith in Government and Institutions

Hannah Fingerhut, *Associated Press*

American public opinion research over decades has documented a steady decline in trust in institutions, both public and private. In the last 10 years, the AP-NORC Center for Public Affairs Research has chronicled and analyzed the crisis of

confidence as political polarization has intensified and facets of U.S. democracy have been questioned. This paper will draw from AP-NORC surveys in recent years to show the American public has low trust in the very institutions that continue to be relied on for information and services, including the news media, religion and government. The paper will explore how that plays out in specific instances, such as the COVID-19 pandemic. It will then seek to understand what undergirds the decline in trust in government in particular by focusing on the erosion over the last five years in American confidence in government to protect many of their basic rights.

A Unique Model to Bring Journalists Together with Social Scientists: 10 Years of the AP-NORC Center
Trevor Tompson, *NORC at the University Of Chicago*

Now more than ever there is a need for high-quality, trustworthy, fact-based, data-driven insights to inform key policy debates. The hyper-partisan environment that breeds mistrust in information requires trusted, non-partisan institutions to step forward and try their hardest to break through with insights that all parties can trust. It is in that spirit that The Associated Press and NORC at the University of Chicago formed a partnership to bring data-driven insights to the public by informing fact-based journalism. At the time the partnership was formed, there was a lot of anxiety among AAPOR members that the tradition of high-quality media polling might well come to an end in the United States. The financial crisis had hit news organizations especially hard and in-house polling units were being eliminated or faced severe budget cuts. Recognizing this challenge, AP and NORC formed a joint venture that aimed to reinvent how public opinion research was conducted for news, looking to a different funding and implementation model that would help sustain high-quality social science research that was intended to inform the best of journalism. Since then, the AP-NORC Center has been a success story. The partnership now routinely brings data-driven insights that inform not just AP's coverage but that routinely reach more than two thirds of the world's newsrooms and more than half the world's population. As an introduction to a panel that highlights some of the best of the 10 years of AP-NORC research, we will discuss the challenges, success and long-term impact that the partnership has had on AP's journalism, on research at NORC, and the influence it has had on key policy debates.

The Role of Information and Faith in Democracy: Celebrating the 10th Anniversary of the AP-NORC Center for Public Affairs Research
Trevor Tompson, *NORC at the University Of Chicago*

The nation is facing a crisis of faith in democracy, and declining trust in institutions is just one symptom. In the 10 years since the founding The Associated Press-NORC Center for Public Affairs Research, AP-NORC researchers have tracked the views of the public on their confidence in democratic institutions, and have attempted to delve deep into the data to understand what leads people to trust or distrust news and information. This panel celebrates the 10th anniversary of the unique joint venture between The Associated Press and NORC at the University of Chicago by synthesizing hundreds of surveys and research reports that the Center has released to the public. The paper presentations will focus on research about how and why people trust and consume news, chronicle the decline in confidence in the electoral system, explore the views of the public on misinformation, and provide a synthesis of research on declining trust in insitutions and how this impacted the public's reaction to the COVID-19 pandemic. The session will begin with a discussion of the unique journalism-research partnership of the AP-NORC Center including a review of the Center's impact on AP's journalism, its impact on NORC's research, and its contribution to important policy debates.

Understanding News Consumers
Marjorie Connelly, *NORC at the University Of Chicago*

The Media Insight Project is one of the core research initiatives of The AP-NORC Center for Public Affairs Research. Conducted in partnership with the American Press Institute, the Media Insight Project has published several groundbreaking studies aimed at understanding how people consume, understand, and trust news and information in today's rapidly changing media environment. The collaboration has produced insights on how and why Americans choose to pay for news content, and how personal politics impacts where people go, trust and pay for news. The project defined the "personal news cycle" and followed how consumers are exposed to and interact with news throughout the day, whether they seek it out or bump into it while doing other activities. A particular focus of the research has been around how young people consume news content. A recent study around trust in media found that

not all Americans embrace many of the core values that guide journalistic inquiry, and suggested that this disconnect may be one of the root causes of the lack of public trust in media. This presentation will look back at this body of research and illuminate some of the key insights that have emerged from this remarkable research collaboration.

Constant Vigilance: Continuing to Evolve Methods for Ensuring Quality in Interviewer-Administered Data Collection

Constant Vigilance: Continuing to Evolve Methods for Ensuring Quality in Interviewer-Administered Data Collection
Mary Davis, *US Census Bureau*

Survey organizations conduct interviewer-administered data collection efforts with both households and establishments almost every day. While data collection practices are continuing to evolve, so must methods for ensuring the data collected are of the highest quality possible. As noted in AAPOR's Falsification in Surveys report (2020), i t is incumbent on survey research organizations to establish policies and practices aimed at preventing falsification and at detecting and removing falsification should it occur.   The assurance of data integrity in survey research includes the " appropriate use and analysis of process data and data from actual interviews to detect anomalies." (AAPOR, 2003) This session will detail the latest efforts by the U.S. Census Bureau, NORC at the University of Chicago, and Westat to monitor data collected by field interviewers in order to identify and address potential issues that impact survey data quality. The presentations cover recent methodological advances to identify potential falsification, including the development of a Field Quality Monitoring program for ongoing Census Bureau surveys using 'boots on the ground' data analysts in regional offices for investigations; the use of data visualization tools to bring large, complex quality datasets to life for monitoring interviewer behavior; the evaluation of a combination of response and paradata used to detect data quality issues at small geographic levels in the American Community Survey; the use of data science techniques to develop an enhanced falsification detection program for NORC field survey data collections; and the use of machine learning to reduce labor intensive human coding from Computer Assisted Recorded Interviewing (CARI).

Beyond Reinterview: Creating a Near Real-Time Field Quality Monitoring Program Using Advanced Analytics for Detecting and Mitigating Interviewer Falsification
Mary Davis, *US Census Bureau*
Richard A. Denby, *U.S. Census Bureau*

The Census Bureau collects data for multiple surveys on a monthly basis by relying on local interviewers either visiting homes or calling respondents to collect data from sampled households.   Survey interviewers are often cited as a significant source of potential measurement error in survey data collection (Groves, 1989).   Measurement error can be introduced when interviewers fail to follow procedures, are improperly trained, or deliberately falsify the data.   Detecting and mitigating these types of errors in any survey collection operation is challenging, especially since falsification is a relatively rare event, but important since it affects overall survey data quality. In response to this challenge, the Census Bureau has created a Field Quality Monitoring (FQM) Program focused on designing and implementing an enterprise approach to near real-time quality monitoring of field data collection activities.   This includes the design and implementation of a quality monitoring program that utilizes data analytics of paradata to monitor, mitigate, and address potential issues with field data collection. This paper will discuss the success of our pilot study designed for the American Housing Survey, utilizing our 'boots on the ground' analysts in each of the six Census Bureau regional offices. Outlier detection was done within three quality components of FR Workload, Interviewer Behavior, and Response Behavior, and included metrics such as refusal rates, partial interview rates, interview duration, and item nonresponse rates. We will also discuss the next steps for the program which includes taking a holistic view of an interviewers work across all surveys they work on.

Picture This: Using Data Visualization to Identify and Investigate Potential Falsification by Survey Interviewers
Laura Hergert, *U.S. Census Bureau*
Melanie A Hollowood, *U.S. Census Bureau*
Scott W Glendye, *U.S. Census Bureau*
Sheldon Waugh, *U.S. Census Bureau*

On any given day, U.S. Census Bureau survey interviewers are visiting homes or calling respondents to conduct interviews with sampled households. Every personal visit or phone attempt is carefully tracked through enumeration devices. Paradata from each attempt, such as time spent interviewing, are automatically recorded and transferred downstream to large databases. While these data can be used by analysts to identify problematic or even potentially falsifying interviewers, it is inefficient and taxing to sift through millions of rows of data stored across multiple tables in various databases.   In response to this challenge, the U.S. Census Bureau's Office of Survey and Census Analytics (OSCA) used data visualization tools to bring the large, complex quality datasets to life for their Field Quality Monitoring (FQM) Program. Paradata from FQM's pilot study, the 2021 American Housing Survey (AHS), first were merged, cleaned, and aggregated to calculate outlier metrics such as refusal rates and interview duration. This paper will discuss how the AHS paradata were transformed through multiple visualization tools, including Tableau, ArcGIS, and Excel, and how those tools were then used by analysts to identify and investigate survey interviewers. The main benefits of using data visualization for a quality program will be examined. In addition, key visualization features, such as tooltips, click-to-filter, map layers, and drill-down tables, will also be discussed.

An Overview of the ACS Field Data Quality Operation
Erik Schmidt, *U.S. Census Bureau*

The American Community Survey (ACS) is a multimodal survey that samples approximately 3.5 million addresses per year and relies on computer-assisted personal interview (CAPI) follow-up with nonresponders. The considerable size of the ACS requires several thousand field representatives (FRs) working ACS nonresponse cases in a given month, in nearly all parts of the United States. In the past few years, the ACS has expanded efforts to ensure high data quality in the CAPI operation. This has included the buildup of a data monitoring and FR follow up operation, including research to identify data quality issues and follow up on those issues in cooperation with staff from multiple divisions and directorates across the US Census Bureau, at headquarters and in the Census field regions. This operation has streamlined ACS's awareness of and ability to address data quality issues and FRs of concern. The operation has the dual benefits of raising awareness of data quality issues prior to final tabulation and improving the quality of future data collection by targeted FR follow up and retraining. In this presentation, I will discuss the general organization of the FR-level data quality operation, the steps taken when a potential data issue is flagged, the FR-level checks themselves, and the operation has evolved in response to various issues we have discovered. Finally, I will discuss the applicability of such data quality control measures across various data collection environments.

Using Data Science to Develop an Enhanced Falsification Detection System
Kate Bachtell, *NORC at the University Of Chicago*
Shannon Nelson, *NORC at the University of Chicago*
Catherine C. Haggerty, *NORC at the University of Chicago*
Lisa Blumerman, *NORC at the University of Chicago*

Interviewer falsification is extremely costly for survey operations and a major threat to data quality. Rising concern about falsification was a driving force behind the Falsification in Surveys report commissioned by the AAPOR Executive Council and published by the AAPOR Task Force on Data Falsification in 2020. In this paper we describe the use of data science techniques to develop an enhanced falsification detection program for NORC field survey data collections. The objective was to streamline quality monitoring processes and identify potential falsification in near real time. A team of data scientists, survey directors, analysts and systems developers leveraged confidential data from the 2019 Survey of Consumer Finances (SCF) as a test case. The team first explored the data using data mining, machine learning, and natural language processing. We then identified survey and paradata variables that were statistically associated with interview falsification. Chief among these were measures of short cutting the interview by reporting the least allowable amount of complexity for a given item (e.g. unmarried respondent, very few financial assets, etc.). The falsification scores generated by these models, along with a wide range of paradata, are integrated in an interactive dashboard built with SAS Visual Analytics. Our paper includes a description of the exploratory analyses conducted, findings from

statistical modeling, and a discussion of operational implications. We propose a strategy for identifying predictors of interview falsification that can be generalized to a variety of field surveys.

Using Machine Learning to Assess Interviewer and Question Performance
Hanyu Sun, *Westat*
Ting Yan, *Westat*

Computer Assisted Recorded Interviewing (CARI) has long been used by field management to monitor interviewer performance and to assess questionnaire items (e.g. Hicks, Edwards, Tourangeau, McBride, Harris-Kojetin and Moss 2010). Prior research found that providing feedback to interviewers based on CARI was effective at improving interviewer performance (Edwards, Sun and Hubbard 2020). Conventionally, a human coder needs to first listen to the audio recording of the interactions between the interviewer and the respondent, and then evaluate and code features of the question-and-answer sequence using a pre-specified coding scheme. Such coding tends to be labor intensive and time consuming. Due to resource constraints, often a small proportion of completed interviews or a selected group of questionnaire items can be evaluated in a timely manner. In this study, we will present a pipeline we developed at Westat that heavily draws on the use of machine learning. The pipeline includes two components: diarization and speech-to-text. The diarization process detects who spoke at which turn in a question-level audio recording. The speech-to-text feature provides conversational turn-level transcript. In the presentation, we will present how to use the pipeline to detect potential interviewer falsification, identify interviewer with undesirable behaviors, and detect problematic questionnaire items. We will evaluate the performance of the pipeline using both mock interviews produced in a laboratory setting, and field interviews from a nationally representative survey. We will also discuss the time and cost implications of using the pipeline as compared to the conventional human coding.

Recruiting Respondents Online: Methods & Sample Sources

Who Do Respondents Recruit in Respondent Driven Sampling Studies?
Ai Rene Ong, *University of Michigan*
Michael Elliott, *University of Michigan*
Sunghee Lee, *University of Michigan*

Respondent Driven Sampling (RDS) is an important method to sample hard-to-reach populations. It is a modified snowball sampling method with restricted peer recruitment (i.e., respondents are limited to a predetermined number of recruits). The efficacy of this method in reaching the population of interest depends on the existing respondents' peer recruitment cooperation and success. Much of the analysis methods using RDS data depends on a several assumptions about the way respondents recruit in their networks, among them that respondents have a known number of contacts, that these contacts are bidirectional (the contacts consider the respondent among their set of contacts), and that the respondents recruit randomly among their contacts. Getting a better understanding of these recruitment dynamics will help researchers plan for data collection challenges upfront (e.g., if specific demographics are harder to recruit, we can increase the number of seeds from that demographic or modify the recruitment incentive). This study aims to explore the characteristics of the respondents who successfully recruit their peers into an RDS study, the characteristics of the people respondents consider recruiting and the characteristics of the people who are recruited. Data from a web-RDS survey targeting the adult LGBT population residing in the U.S. will be analyzed to answer these questions. In this web survey, respondents were presented with a hypothetical RDS study and asked to think about the people they plan to recruit. They were also asked to recruit two people chosen by random from the people they described as their possible recruits. These analyses will also be supplemented with in-depth interviews with different respondents to understand the thought processes behind the decisions further to cooperate with recruitment and the selection of the people in their network to recruit.

Reaching Across the Political Aisle: Overcoming Challenges in Using Social Media for Recruiting Politically Diverse Respondents
Maggie Macdonald, *New York University*
Megan A. Brown, *New York University*
Nejla Asimovic, *New York University*

Rajeshwari Majumdar, *New York University*
Lena Song, *New York University*
Laura Huber, *University of Mississippi*
Sarah Graham, *New York University*
Joshua A. Tucker, *New York University*
Jonathan Nagler, *New York University*

A challenge for public opinion surveys across modes today is achieving representativeness of respondents across relevant political characteristics such as attitudes and demographics. We focus here on overcoming known challenges in online recruiting of Republican respondents for surveys on American politics by testing recruitment strategies using Facebook ads. While the use of Facebook ads for recruitment has increased and offers potential benefits, such a recruitment strategy suffers from a lack of Republican engagement relative to Democrats and Independents. In this project we focus on one feature which may influence a potential participant to click on an ad, consent to begin, and complete a survey: the institution listed as fielding the survey. We argue that a potential participant's trust, or lack of trust, of an institution will affect whether they choose to participate. We focus on the relevant case of academic institutions. Respondents are recruited through three otherwise identical Facebook pages and ads listing : New York University (NYU, a more liberal university), the University of Mississippi (Ole Miss, a less liberal university), and CSMaP Research Center (as the control, a generic research institution). We predict that Republicans will be most likely to both begin and complete surveys from Ole Miss, followed by CSMaP Research Center, and least likely to do so for surveys from NYU. Additionally, we will assess whether the Republicans who agree to participate under each treatment differ from each other in terms of easily observed characteristics such as education and region, and also, more importantly, in their political attitudes. Our findings will be a useful step forwards in overcoming existing challenges in recruiting Republicans for research.

Test-Retest Reliability of Four U.S. Non-Probability Sample Sources

Mario Callegaro, *Google Inc*
Inna Tsirlin, *Google Inc*
Yongwei Yang, *Google Inc*

It is a common practice in market research to set up cross sectional survey trackers. Although many studies have investigated the accuracy of non-probability-based online samples, less is known about their test-retest reliability which is of key importance for such trackers. In this study, we wanted to assess how stable measurement is over short periods of time so that any changes observed over long periods in survey trackers could be attributed to true changes in sentiment rather than sample artifacts. To achieve this, we repeated the same 10-question survey of 1,500 respondents two weeks apart in four different U.S. non-probability-based samples. The samples included: Qualtrics panels representing a typical non-probability-based online panel, Google Surveys representing a river sampling approach, Google Opinion Rewards representing a mobile panel, and Amazon MTurk, not a survey panel in itself but de facto used as such in academic research. To quantify test-retest reliability, we compared the response distributions from the two survey administrations. Given the attitudes measured were not expected to change in a short timespan and no relevant external events were reported during fielding to potentially affect the attitudes, the assumption was that the two measurements should be very close to each other, aside from transient measurement error. We found two of the samples produced remarkably consistent results between the two survey administrations, one sample was less consistent, and the fourth sample had significantly different response distributions for three of the four attitudinal questions. This study sheds light on the suitability of different non-probability-based samples for cross sectional attitude tracking.

Polarized Social Media Users or Polarized Ads? Mechanisms That Shape the Sample Composition in Social Media Recruited Surveys.

Zaza Zindel, *Universit _ t Bielefeld*
Simon Kühne, *Bielefeld University*
André Grow, *Max Planck Institute for Demographic Research*

A growing number of research projects are turning to social networks such as Facebook to recruit (rare) populations for online survey participation. These social media platforms are based on an advertising revenue model that researchers can leverage by purchasing advertisements to recruit survey participants. However, researchers have only limited control over the algorithms that allocate ads to users. Moreover, the goals of survey researchers (e.g., obtain representative survey data) do not necessarily match those of social media platforms or advertisement management systems (to maximize interaction and profit). Consequently, ad allocation algorithms may lead to polarized response patterns that favor users with very strong opinions on certain survey topics. Thus far, survey research lacks insights into 1) how the advertising algorithms allocate ads to users and 2) how different ad designs affect the composition of survey participants. Our presentation covers multiple Facebook advertising experiments conducted as part of various web survey projects in 2020-2021. We contrast different advertising texts, images, and campaign options. In doing so, we examine whether polarizing features of ads, such as unmoderated discussions and negative reactions, are also reflected in the survey data. This talk thus provides valuable insights into data quality hurdles such as sampling errors in survey data generated in this way. Moreover, our examples can serve as a starting point for further research.

Retiring Overrepresented Panelists to Improve Web Panel Composition

Nicholas Hatley, *Pew Research Center*
Courtney Kennedy, *Pew Research Center*
Arnold Lau, *Pew Research Center*
Nick Bertoni, *Pew Research Center*

The American Trends Panel (ATP) is a probability-based, web panel that represents the primary platform for domestic polling by the Pew Research Center. In raw, unadjusted numbers, the ATP contains proportionately too many college graduates, registered voters, and Democratic-leaning adults. While this is not a challenge unique to the ATP, Center researchers developed a strategy to retire a set of overrepresented panelists in order to align the composition of the panel more closely to the US population. A key goal of this exercise was to improve the ATP's partisan balance. Researchers identified Democratic-leaning panelists that did not possess any hard-to-reach demographic characteristics and flagged them for potential retirement from the ATP. After simulating many different approaches, a strategy of retiring 2,602 out of roughly 13,000 active panelists was chosen. This approach resulted in an unweighted distribution on self-reported 2020 presidential election vote of 51% Biden and 47% Trump. This presentation will detail the different approaches that were considered for structuring the retirement and how certain decisions were made on which panelists to consider. This presentation will also leverage the amount of survey data collected on the ATP to examine the impact retiring these panelists had on different survey estimates

The National Cancer Institute's Cannabis Supplement Grant: Twelve Surveys of Cannabis Among Cancer Patients

Digital Literacy: A New Lens Needed to Increase Inclusion and Diversity in Research

Brooke Worster, *Thomas Jefferson University*
Preethi Selvan, MPH, *Thomas Jefferson University Hospital/Sidney Kimmel Cancer Center*
Gregory Garber, MSW, *Thomas Jefferson University Hospital/Sidney Kimmel Cancer Center*
Ayako Shimada, *Thomas Jefferson University Hospital/Sidney Kimmel Cancer Center*
Amy Leader, DrPH,MPH, *Thomas Jefferson University Hospital/Sidney Kimmel Cancer Center*

Background : The COVID-19 pandemic revealed deep concerns about digital literacy and the need to foster research methods that do not exacerbate the 'digital divide' and exclude vulnerable populations from participating in research. While we increasingly use technology for survey-based research, this creates biases that threaten validity and generalizability of this research.   Objective: To ensure an adequate response rate and increase the diversity of respondents to a cross-sectional survey about cannabis use among cancer patients at a large, urban academic cancer center.   We aimed to engage older patients, those with low digital literacy, and those from a racial or ethnic minority background to understand the full scope of cannabis use among cancer patients.   Methods: We identified patients who received treatment at our cancer center over the past 3 years from electronic health records. We randomly sampled patients to receive the survey with a goal of 1,000 completed surveys. We mailed a paper copy with a self-addressed stamped envelope to potential participants, with an option to complete online. After 13 weeks, we sent

an email with a web link to non-responders. Participants who completed the survey were mailed a $5 gift card to a local convenience store. We analyzed response and demographic data by frequencies, mean scores, and bivariate calculations.   Results: Of the 18137 cancer patients identified, 10050 were included in our study sample. In total, 1558 surveys (16%) were completed: 976 (63%) by paper and 582 (37%) electronically. Older patients, those of a minority race or ethnicity, and those who had never used cannabis were more likely to complete the survey on paper.   Conclusion: Electronic surveys are popular for ease and cost but solely relying on them as a distribution mechanism excludes vulnerable populations. Traditional mechanisms of mailed surveys may overcome potential biases and create a more representative sample.

Case Studies on Potential Challenges and Outcomes of Cannabis and Cancer Survey Research Methods

Nicolas Hernandez Ortega, *University of Miami Sylvester Comprehensive Cancer Center*
Corrine McDaniels-Davidson, *San Diego State University Institute for Public Health*
Richard T. Lee, *Case Comprehensive Cancer Center*
Talya Salz, *Memorial Sloan Kettering Cancer Center*

  Introduction The National Cancer Institute's Cannabis supplement tasks 12 cancer centers across the United States with surveying a minimum of 1000 cancer patients to better understand the prevalence of cannabis use among patients. Survey methodologies and receptiveness to the survey have differed widely across cancer centers, as have recruitment outcomes. Enrolling cancer patients in studies generally presents challenges; however, additional recruitment difficulties during this study are specific to the sensitive survey topic and the timing of the survey during the Covid-19 pandemic. In this discussion, four cancer centers (Sylvester Comprehensive Cancer Center in Miami, Moores Cancer Center at UC San Diego Health in San Diego, Memorial Sloan Kettering Cancer Center in New York City, and Case Comprehensive Cancer Center in Cleveland) present their methods for identifying and recruiting patients, modes of survey administration offered, incentives, institutional constraints, and other challenges and facilitators relevant to survey completion. Methods Cancer centers administered surveys through REDCap and Qualtrics softwares and physical modalities to measure center-specific cancer and cannabis questions as well as a set of core questions determined by NCI. In this presentation, each site will highlight factors that may affect survey response rates at their center, including regional legal issues concerning Cannabis, languages spoken, the digital divide, and specific patient populations sampled. Results Data collection efforts were met with difficulties ranging from additional institutional review board requirements, limitations on participant consent to contact methods, participant concerns with personal identifying information. Difficulties warranted adjustments to study methods as well as cancer center REDCap surveys. These mid-survey adjustments and their effects on recruitment and response will be discussed.   Conclusions Discussions will focus on the potential for bias in survey response and how to interpret our findings given the specific recruitment challenges of this multi-site survey effort.

Survey: Pattern of Cannabis Use Among Cancer Patients

Kathy Helzlsouer, *The National Institutes of Health*
Kathy J. Helzlsouer, *National Cancer Institute*
Andrew Freedman, *National Cancer Institute*
Sonia Rosenfield, *National Cancer Institute*
Zoe Padgett, *ICF*
Deirdre Middleton, *ICF*

The legal landscape of medical and recreational cannabis use in the United States is rapidly evolving with wide variation in state policies. The available delivery methods of cannabis have also undergone dramatic changes and include edibles, oils, tinctures, topicals, and inhaled forms. State-based policy changes are taking place at a time when research on the potential beneficial or adverse health effects of cannabis use among cancer patients is limited. Common conditions for which it has been used among cancer patients include anorexia, nausea, and pain. The extent of use, perceived and real benefits and risks of use, potential interactions with cancer treatment and other medications, and impact on comorbid conditions are uncertain. Surveys indicate that while US medical oncologists discuss cannabis use with patients, few feel informed to make specific recommendations. To understand the patterns of use among cancer patients living in states with varying cannabis legalization policies, the National Cancer Institute, Division of Cancer Control and Populations Sciences funded twelve comprehensive cancer centers to conduct surveys of 1000 or

more cancer patients undergoing or recently completing treatment. Funded centers used a variety of sampling designs and patient recruitment methods. A core set of questions were developed to be used by all sites that covered the following topics: perceived benefits and harms or risks; sources of information on cannabis use; discussion with or recommendations from health care providers; current and past use of cannabis products; reasons for use or not use of cannabis during treatment; among users, the mode, frequency, duration of use, and symptom response. Sites included additional questions depending on their specific research interests. Survey response rates varied across sites. The challenges, opportunities and need for conducting future research on patterns of cannabis use will be discussed.

Same Survey, Different Sample: Weighting and Harmonizing Data Across Twelve Surveys of Cancer Patients
Richard Harding, *ICF*
Yun Kim, *ICF*
Zoe Padgett, *ICF*
Deirdre Middleton, *ICF*

The National Cancer Institute (NCI) awarded grants to twelve Cancer Centers across the United States to conduct surveys about attitudes, beliefs, and behaviors related to cannabis use among cancer patients. Each cancer center was responsible for designing and conducting their survey to meet the needs of their unique populations and research goals. Therefore, the surveys used a variety of approaches to recruitment, sampling, and other aspects of methodology. In collaboration with NCI and the grantees, ICF weighted each cancer center's data to the patient population that each cancer center drew their sample from and harmonized the data across the 12 grantees. Weights were created for each cancer center accounting for different sample designs, adjusted for nonresponse and raked to the appropriate patient population. To harmonize the data from the grantees, we developed core measures, standardized data structures and variable formats, and produced a harmonized codebook. This process includes assessing missing core elements, checking consistency in variable names and response options between the data, and recoding variables. Harmonization of the grantee data is the first step in assessing the possibility of pooling the cancer center's data for combined analysis. The weighted harmonized data will promote future data analysis when appropriate on cannabis use with larger population coverage and comparison between cancer centers. This presentation describes the issues with data weighting and harmonization across the 12 cancer centers in different catchment areas. Key lessons include best practices to develop efficient process for data harmonization and benefits of having weighted harmonized data.

The National Cancer Institute's Cannabis Supplement Grant: Twelve Surveys of Cannabis Use Among Cancer Patients
Zoe Padgett, *ICF*
Deirdre Middleton, *ICF*
R. Lee Harding, *ICF*
Yun Kim, *ICF*
Kathy Helzlsouer, *National Cancer Institute*
Andrew Freedman, *National Cancer Institute*
Sonia Rosenfield, *National Cancer Institute*
Gary Ellison, National Cancer Institute

The National Cancer Institute's Cannabis Supplement Grant funds twelve cancer centers across the United States to conduct surveys of their cancer patient populations about their use of cannabis. This panel will showcase five of those surveys, discuss the purpose of the grant, and talk about the difficulty of developing measures and harmonizing data across twelve unique surveys. First, representatives from NCI will discuss the motivation behind the grant and the need to study medical cannabis use among cancer patients. Research in this area is limited, but we know that cannabis has been used among cancer patients to address various cancer symptoms and cancer treatment side effects. More information is needed in order for states to make informed policy decisions and medical oncologists to make specific recommendations to patients. We will then discuss the process for developing a set of core measures for each cancer center to include on their surveys and how data were weighted and harmonized. This process involved listening to stakeholders from NCI and each of the cancer centers, and developing survey items on a topic with little prior survey research. Each survey had its own sampling and data collection methodology, impacting the weighting process. Finally,

several of the cancer centers will discuss challenges and successes in their data collection approaches, and how the context of each survey impacted response rates and potentially biased survey results. Topics discussed will include digital literacy, incentives, IRB considerations, and the sensitivity of the survey topic based on the legal status of cannabis in different states. This panel will contribute research on a newer topic with a unique population and discuss multiple modes of contact and data collection with varying outcomes.

"Did You Ever, Even Once, Use Cannabis for Any Reason?": Designing Questions on Cancer Patients' Cannabis Use for Twelve Different Surveys

Zoe Padgett, *ICF*
Deirdre Middleton, *ICF*
R. Lee Harding, *ICF*
Kathy Helzlsouer, *National Cancer Institute*
Andrew Freedman, *National Cancer Institute*
Sonia Rosenfield, *National Cancer Institute*
Gary Ellison, *National Cancer Institute*

Designing a questionnaire for a single survey is challenging. Designing a questionnaire to be agreed upon by twelve different organizations is even more difficult! The National Cancer Institute's Cannabis Supplement Grant presented exactly this challenge. Working in collaboration with NCI, we needed to design a set of core measures to be collected by twelve different cancer centers across the United States in twelve different surveys. To make things even more complicated, medical cannabis use in cancer patients is a topic with few existing, pre-tested questions available. A core set of about 40 questions was developed to be used by all sites, covering the following topics: perceived benefits and harms or risks; sources of information on cannabis use; discussion with or recommendations from health care providers; current and past use of cannabis products; reasons for use or not use of cannabis during treatment; and among users, the mode, frequency, duration of use, and symptom response. All twelve sites used the core measures in their surveys and some included extra questions to expand the scope. They utilized varying contact and data collection modes including web, paper, and phone. Sites were located in states with varying laws about medical and recreational cannabis use and differing population characteristics. This presentation describes the process for gathering input from stakeholders, synthesizing the available existing questions, and designing a questionnaire that could stand alone or be easily adapted to fit additional questions. We present the flow of core processes and key decision points along the way. Lessons learned include starting questionnaire development early, providing a forum for feedback, utilizing subject-matter expertise, and being flexible and adaptable throughout. We present a selection of the final core measures and discuss best practices for designing medical cannabis-related questions and changes we would make in future surveys to improve measurement.

Improving Representation Among Traditionally Underrepresented Groups in Surveys

Examining Patterns of Missingness in American Community Survey Across Question Types with a Multilingual Perspective

Heather Smalley, *Willamette University*

Language use trends in the United States across space and time tell a rich story of immigrant communities, assimilation, and accessibility. Historically, language use has been studied through the census every ten years; however, annually since 2006, a sequence of three language questions has been included in the American Community Survey (ACS). These data have been used by demographers and sociolinguists alike to gauge the need for accessible material in minority languages. As a result, the ACS is offered in sequential modes (internet and mail, CATI, and CAPI) with varying levels of support for non-English speakers. Previous researchers have used these data to assess patterns in missingness across modes related to household language and English proficiency in a single iteration of the ACS (McGovern and Griffin 2003) and over time (Smalley 2019). A natural extension of this work is to consider how patterns of missingness may vary across types of questions, such as radio button, check box, open-ended response, when considering the household language and level of English proficiency. This may also affect the ability to follow skip logic; particularly in self-administered modes. We use big data files from both the household and personal level ACS

between 2006 and 2019. The person level data provides more detail on certain variables; for instance, language spoken is specified for individuals, but when aggregated to the household level, it is summarized using five language classes.

A Meta-Analysis of Ethnographic Studies of Undercounted Populations in the U.S. Census
Joanne Pascale, *U.S. Census Bureau*

The goal of the census is to count every person in the country once and in the right place. Inevitably, some individuals are omitted from the form (an undercount), and others are listed on more than one form (an overcount). If the undercounts outnumber the overcounts, a net undercount results. Research on the 2010 Census found there was a net undercount of about one million young children (under age 5), resulting in a net undercount rate of 4.6% overall for this age group – higher than any other age group by far. The rate was higher for Hispanic and Black children (7.5% and 6.3%, respectively), and in larger counties, presumably because these counties had higher concentrations of minorities and possibly other hard-to-count characteristics. In 2014 the Census Bureau released a task force report recommending research into the causes of the undercount of young children. One study aimed to conduct interviews in 2020 in households at risk for undercounting young children, but is on hold due to the pandemic. The hiatus in face-to-face interviewing has provided an opportunity to revisit prior research on hard-to-count populations in general, given the likelihood that characteristics of undercounted young children intersect with other undercounted sub-groups. The aim of this meta-analysis is to explore whether there are data collection issues that cut across these sub-groups and present potential avenues for reducing the undercount. Data sources include a series of ethnographic studies, mostly from the 1990s, with a variety of hard-to-count groups (e.g., American Indians, migrant farm workers, recent immigrants, racial and ethnic enclaves, and Black males), and more recent studies on the undercount of young children in particular. Themes uncovered include family, household and community structure (e.g., complex and extended families; tenuously attached household members; crowded housing), linguistic isolation, and inter-household and cross-border mobility.

Examining the Use of Appended Emails to the Sample Frame During Recruitment to a Probability-Based Panel
Lindsay Liebert, *NORC at the University of Chicago*
Ipek Bilgen, *NORC at the University of Chicago*
David Dutwin, *NORC at the University of Chicago*
Erlina Hendarwan, *NORC at the University of Chicago*
Roopam Singh, *NORC at the University of Chicago*
J. Michael Dennis, *NORC at the University of Chicago*

This paper examines the impact of using email during recruitment to a probability-based mostly web panel. During the 2021 recruitment effort for NORC's AmeriSpeak panel, potential emails obtained from a third-party vendor for invited households were utilized as a method of recruitment, in addition to mail and phone during the initial recruitment prior to the nonresponse recruitment efforts which include FedEx and in person recruitment. A randomized experiment was implemented for invited panelists with emails, with half participating in an email recruitment campaign and the other half serving as the control group. The email recruitment campaign involved sending a series of email invitations and reminders to join the AmeriSpeak panel by completing the recruitment survey, using up to six identified emails for each invited panelist. This paper will analyze the differences in recruitment yield rates resulting from the email campaign, optimal number of emails to use, and how to identify these emails using accuracy metrics provided by the vendor with emails. Additionally, we will present the effect of using email contact in addition to other traditional recruitment modes on targeted hard-to-reach subgroups, including but not limited to Spanish-dominant households, young adults, and Non-Hispanic Black/African American persons. Preliminary results find that recruitment yield rates for targeted Spanish-dominant households are higher for those sent panel recruitment information via email. In addition to findings from the randomized experiment, we will also present findings on the success of the email campaign using metadata from the email delivery platform. In particular, we will examine at which email reminder invited panelists enter the recruitment survey, how many emails invited panelists receive before starting recruitment, email open rate, and click-to-open rate.

Improving Equity and Inclusion in Travel Surveys through Differential Incentive Offerings
Michelle Lee, *N/A*

Amid the current trend of decreasing survey response rates, it is even more critical for survey researchers to reach sufficient and representative sample sizes among hard-to-survey populations or other groups of interest. Offering incentives has been shown to increase survey response rates in general, with variations of incentive amounts and structures yielding differing results based on survey design. In an effort to improve response from People of Color, a historically underrepresented group in household travel surveys, two recent household travel surveys, the Travel Behavior Inventory in the Twin Cities, Minnesota region, and the Puget Sound Regional Travel Study in the Puget Sound, Washington region, tested the effectiveness of differential incentive structures for improving response rates. Both studies recruited participants through address-based sampling. In the mailed invitation letters, each study offered higher, differential incentives to households in census block groups that contained high proportions of People of Color. The Travel Behavior Inventory also offered a higher, differential incentive to all households in the sample who qualified as hard-to-survey based on demographic information provided during a recruitment survey, including if the household member completing the recruitment survey reported being Hispanic or a Person of Color, if the annual household income was less than $35,000, or if the household had four or more members. The goal of these approaches was to increase the proportion of historically underrepresented demographic groups in the final sample. This presentation will review the effectiveness of these approaches across both studies.

All In: An Examination of Predictors of Survey Participation for Sub-groups
Mina Muller, *Ipsos Public Affairs*
Randall Thomas, *Ipsos Public Affairs*
Frances Barlas, *Ipsos Public Affairs*

In general population surveys, researchers face several challenges in maintaining representativeness, especially from underrepresented racial and ethnic groups. While the existing literature offers several best practice methods for increasing the overall response rates, few studies exist pertaining to how various survey features affect the participation of people overall and people of color specifically. In our previous work we found that what motivated survey participation across opt-in samples and probability-based samples was similar for different race and ethnicity subgroups. In this study conducted in April 2021, we had over 3,000 respondents from a nationally representative probability-based panel participate in an online study to explore the motivational factors of survey participation. We asked questions on prior survey participation and experiences, specifically focusing on factors affecting their survey participation, including what topics were appealing, reasons for participating and types of information that are helpful to know before participating, and whether these factors would increase or decrease their participation. We used multiple regression models to better understand predictors of survey participation overall, and for underrepresented racial and ethnic groups by looking at their reasons for participation and their previous experiences with surveys. While some factors like incentives were more strongly predictive of participation for all participants, we also found that other features like survey topic can be differentially influential for participation. We summarize how the salience of survey topic, information provided about the research and survey length motivates people differently and will provide researchers additional considerations to motivate more participation from diverse participants.

The Demise of Democracy? Eroding Trust, Voter Fraud, & Voter Restrictions
Suppression or Depression: Will New Restrictive Voting Laws Discourage Women from Casting Ballots in 2022?
Debbie Borie-Holtz, *Rutgers University*
Alexa Kurtz, MPP, *Rutgers University*
Lauren Ellis, *Rutgers University*

In the aftermath of the 2020 presidential election, legislatures in several battleground states adopted restrictive voting laws to limit the number of dropboxes, require photo identification for absentee ballots, and shorten the time to request mail-in ballots. We first surveyed roughly 670 women in five of these battleground states (Florida, Georgia, Nevada, North Carolina and Pennsylvania) in the 10 days leading up to the 2020 election as women were expected to participate at levels higher than men in these polities. Among the questions we asked women, 7 out of 10 told us they were more concerned that eligible voters would be prevented from voting in the presidential election as compared to ineligible voters casting fraudulent ballots. However, this rate varied by state and race/ethnicity. In Nevada, only 57 percent were concerned about voter suppression as compared to 91 percent of women in Georgia.   To build on these findings and

the implications these new laws may portend for some voters (e.g., Grofman, 2016; Tan & Grofman, 2015; cf. Birch, 2007, 2011; Schedler, 2002, 2006), we plan to survey our 2020 panel of women voters in January/February 2022 to first measure their awareness of recent changes to these state electoral laws. We will also assess the extent to which these changes may impact their likelihood to vote as well as their enthusiasm to participate in the 2022 midterm election. Further, we will measure whether these new voting provisions will limit their ability to encourage the same family members and friends to again cast ballots in 2022 as compared to the last cycle. Finally, we will present a nested model analysis examining hypothetical changes to the likely voter estimates due to decreased voter participation opportunities in the 2022 midterm elections within these battleground states.

Partisan De-Linking of Survey Cooperation and Turnout in the 2020 Presidential Election

Marc Trussler, *University of Pennsylvania*
Joshua D. Clinton, *Vanderbilt University*
John Lapinski, *University of Pennsylvania*
Sina Shaikh, *University of Pennsylvania*

It is often assumed that survey cooperation and turning out to vote are intrinsically related due to a common cause in political interest. If this is the case, the problem of non-response (even if correlated with an outcome variable) is mitigated because those who refuse to take an election survey are less likely to be part of the population of interest: the electorate. This relationship is disturbed, however, if a factor unrelated to turnout causes survey non-response among a particular group. To investigate this, we pair together the survey disposition files from nearly 1 million phone calls made for the National Exit Poll with verified voter turnout for the 2020 election. Consistent with partisan messaging about "Fake News" and "Fake Polls" the relationship between survey cooperation and turnout was severed in the 2020 election for likely Republican respondents. Republicans refusing to cooperate with polls were just as likely to turn out to vote as those who cooperated. The relationship between survey cooperation and reduced turnout was still present for likely Democrat and independent respondents. This finding is crucial to understanding the bias in the 2020 polls, and to mitigating polling bias going forward. It makes it clear that there is a group of highly motivated voters that lack a specific motivation to respond to polls. Given that these individuals are motivated by a political consideration likely correlated with vote choice, simply giving more weight to Republicans who do respond cannot make up for their exclusion. Without interventions to reach these individuals election polling will continue to miss the mark.

Californians' Confidence in the Electoral Process

Rachel Lawler, *Public Policy Institute of California*

Prior to 2020, California enacted a number of reforms that expanded voter registration, vote-by-mail, and in-person early voting options in an effort to increase voter participation. However, after the onset of the COVID-19 pandemic, election officials made the decision to send mail ballots to all California voters – regardless of whether they requested it or not – to allow people to vote from the safety of their home to prevent the spread of the virus at polling stations. Governor Newsom recently announced that this system of mailing ballots to all voters is now permanent in California, but that voters will have the option to go to the polls if they choose to do so. Using survey data collected through the Public Policy Institute of California's Statewide Survey series, we track Californians' confidence in and concerns about the state electoral process, with particular focus on the role of party registration, political ideology, and race/ethnicity in shaping residents' views. Findings show that, up until the months preceding the 2020 presidential election, the share of likely voters saying they have a "great deal" or "quite a lot" of confidence in California's electoral system had been steadily and significantly declining over the past 15 years. With more voters casting a ballot than ever before in November 2020 – and with a never tested fully vote-by-mail system in place in the most populated state in the union – it is reasonable to expect fear of voter fraud to increase. However, PPIC's surveys show that voter confidence has rebounded to an all-time high. Bivariate and multivariate analyses are conducted to evaluate the influence of demographic and attitudinal variables on confidence levels and concerns about the electoral system, and time trends are examined to determine shifts in public opinion.

Expressive Responding and Trump's Big Lie

Matthew Graham, *The George Washington University*
Omer Yair, *Reichman University*

Do surveys measure sincere belief in Donald Trump's claims that fraud decided the 2020 election? We apply a multi-method approach to detecting expressive responding to the case of Trump's "big lie." Our evidence includes two versions of an honesty encouragement design, a list experiment, two opportunities to express related sentiments, and two opportunities to bet on prominent predictions about the future that were spawned by the big lie. We find minimal evidence of expressive responding. Nearly all survey respondents who directly endorse the big lie appear to genuinely believe it. These "believers" are about evenly split between those who confidently accept the big lie and those who find it to be plausible but are not deeply convinced. Similarly, large majorities of those who predicted that evidence of fraud would enable Trump to retain power in January 2021 or be reinstated in August 2021 appear to have been sincere.

Tell Me More About That: Findings from Qualitative Research

Lessons Learned From Conducting Online Focus Groups of Transgender and Nonbinary Adults
Anna Brown, *Pew Research Center*

The public is becoming more aware of issues related to gender identity, and more Americans than ever before report knowing people who are transgender or nonbinary. Still, fewer than half of U.S. adults say they personally know someone who is transgender (42%) or who uses gender-neutral pronouns (26%). A healthy public dialogue on issues of gender identity must include the voices of transgender and nonbinary Americans, but representative surveys of this population are often cost prohibitive because less than 1% of U.S. adults are transgender. Focus groups are one way to learn more about the nuanced experiences and views of transgender and nonbinary people. This presentation will share lessons learned from a robust online focus group study of adults who told us they are transgender or nonbinary, from the recruitment process to facilitation to analyzing the findings. Because this is a group that faces unique challenges in terms of legal recognition, discrimination, and social acceptance, special care must be taken to protect the participants' identities and to ensure that they feel comfortable sharing openly and honestly. The presentation will also detail the steps taken to create an atmosphere that is respectful and sensitive to the participants' needs while producing research that is rich in detail and insight.

Have We Heard This Before? A Comparison of Findings Across In-person and Remote Cognitive Interviews.
Chris Stringer, *U.S. Census Bureau*
Courtney Reiser, *US Census Bureau*

The social-distancing requirements of the COVID-19 pandemic have resulted in the exclusive use of remote communications technologies (phone, videoconferencing, etc.) to conduct cognitive interviews at the Census Bureau. While remote testing seems adequate for accomplishing the goals of our pretesting research, we do not have a definitive understanding of the potential differences in data quality and participant experience compared to in-person interviews. The research question we seek to answer is whether remote cognitive interviews and those done in-person yield analogous findings and lead to parallel recommendations. In this study, we compare results from cognitive interviews conducted in person with findings obtained from remote testing. We identified survey questions and probes from a prior cognitive interviewing study conducted in-person and included them in a recent study done remotely. We examined the findings relevant to the four stages of the cognitive model to see whether there are differences and if so, in which dimension the differences may be meaningful. This is an initial step in ongoing research to understand mode differences in cognitive interviewing. The results will feed into the larger discussion of the cost-error tradeoffs of these modes and the how we may approach optimizing remote interviews when given a choice of mode.

Every Voice Matters: Promoting Inclusion of Persons With Disabilities in Qualitative Research Methods
Holly Matulewicz, *Mathematica*
Karen CyBulski, *Mathematica*

Focus groups are a useful and effective means of obtaining information on how groups of individuals think or feel about a specific issue. They have been used broadly for several decades to understand factors underlying attitudes, perceptions, and experiences. These attitudes and perceptions are more likely to be exposed through a social

gathering, where many different views and beliefs can be elicited. Yet, people with disabilities are at greater risk for exclusion from participation in focus groups. This exclusion can stem from institutional review boards (or others) recommending caution over research with people with disabilities, researcher or funder misperceptions of an individual's ability to meaningfully participate and engage in focus groups, or other intrinsic factors such as stigma and implicit bias. When conducting research that directly impacts people with disabilities, it is critical that their voices be directly heard. Disability is not homogeneous, and the same type of disability can impact individuals differently. This presentation will focus on strategies make focus groups inclusive and equitable to people with disabilities. By respecting individual differences, working to overcome the emotional and/ or physical challenges that may be encountered, and being thoughtful in your design, focus groups can be more accessible and inclusive.

Can you Hear Me Now? Data Collection Challenges In the Midst of a Pandemic and Tips for Success
Jennifer McNulty, *Fors Marsh Group*
Amy Swallow, *Fors Marsh Group*
Natalie Morrissey, *Fors Marsh Group*
Alicia Garza, *Fors Marsh Group*
Nicole Togno, *Fors Marsh Group*

In March of 2020, the COVID-19 pandemic hit the United States, forcing researchers to abruptly adapt data collection efforts for the foreseeable future. To ensure the safety of staff, participants, and clients, researchers at Fors Marsh Group quickly moved all of our data collection efforts to a remote, virtual environment. Conducting research virtually poses challenges, including technological obstacles and an inability to control the participant's environment. However, this approach also has a variety of benefits, including better geographic reach, cost-effectiveness, and decreased participation barriers, such as a lack of childcare or transportation. Given the numerous benefits of this approach, as well as pandemic-related changes in societal expectations, researchers will continue to leverage virtual data collection in a post-pandemic world. The Fors Marsh Group (FMG) customer experience team tests a variety of products, including surveys, dashboards, paper forms, and websites in the public and private sectors. We employ an array of qualitative and quantitative research methods using a variety of data collection platforms and videoconferencing software to understand the entire user experience. To name a few, we use Zoom, WebEx, Adobe Connect, Optimal Workshop, and Sticky by Tobii Pro to conduct surveys, eye tracking, card sorting, tree testing, focus groups, in-depth interviews, and concept testing. Our experience with diverse methods and virtual platforms, in a rapidly changing digital and social world, has taught us how to account for and address a plethora of obstacles specific to virtual data collection. In this paper, we will explore the challenges that COVID-19 has posed for our virtual data collections in terms of planning and implementation. We will also discuss the solutions we have learned along the way. Finally, we will provide some recommendations for success in implementing virtual data collections moving forward.

Data Visualizations – Maps, Dashboards, and Infographics … Changes Implemented to Improve Data Presentation, on a Very Small State-College Budget, With Limited Staff and Time
Joel LaLone, *Center for Community Studies at Jefferson Community College*
Larry Danforth, *Center for Community Studies at SUNY Jefferson*

This presentation will demonstrate how research agencies who work with large volumes of data while having limited in-house resources, including staffing, none of whom may be programming or data visualization app experts, can utilize accessible, free software to improve deliverables and improve ease of access and use of research results for both clients and consumers.   The Center for Community Studies at SUNY Jefferson in Watertown, NY has completed over 150 community-based studies with a 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 Center is comprised of three statistics faculty on reduced teaching loads, with no additional support staff.   These three individuals recruit, train, hire, and supervise approximately 100 students/year for data collection. Typically, a mixed mode sampling design is employed with 50% of surveys completed via live telephone interviews, and 50% completed via email invitations to complete the survey online. Annual quality-of-life community indicator surveys are our flagship studies; these studies have been completed continuously for 22 years since 2000.   With each passing year we continue to add to the volume of rich survey data: 22 years, three separate counties, 25 core community indicator questions, weighted by five post-stratification nonresponse factors, cross-tabbed typically by at least 6 potential demographic explanatory variables.   In

the past year it has become clear that we could improve our data presentation (at the time of our first year's annual county-wide survey in April 2000 there was no such thing as an app on a smart phone!).   We decided to tackle these three visualization improvements (geospatial maps, dashboards, and infographics) in house.

A Mixed-Methods Approach to Evaluating a National Training Program for Local Staff Who Conduct Interviews With People Living With HIV

Mohua Basu, *Centers for Disease Control and Prevention*
Catherine Espinosa, *Centers for Disease Control and Prevention*
Jennifer Fagan, *Centers for Disease Control and Prevention*

The Medical Monitoring Project (MMP) Interview Workgroup provides national training for state and local health department staff who interview people with HIV. Adequately trained interviewers are paramount to field safety, information security, respect for respondents, linkage to services, and high-quality data. The workgroup conducted a mixed-methods needs assessment to identify future training needs and interviewers' preferred learning styles. The assessment included 6 in-depth key-informant interviews (KIIs) with Project Coordinators (PCs), who train interviewers locally, and 87 structured surveys with interviewers and PCs. At least one respondent from each of the 23 MMP project areas participated in the survey. Sixty-five (75%) respondents were interviewers, and 22 (25%) were PCs. Most staff worked with MMP for <5 years (52; 60%). While 16% of respondents preferred individual learning and 3% preferred group learning, 81% preferred a combination of individual and group learning. Most respondents (83%) reported being visual learners, 36% auditory learners, and 56% kinesthetic learners. Among 65 interviewers, shadowing was ranked first among learning methods, videos second, and role-playing third; however, newer interviewers (<5 years with MMP) ranked shadowing and role-playing highest while veteran staff (>5 years with MMP) ranked videos and conference calls highest. During KIIs, PCs described group training activities (e.g., discussions, mock interviews, brainstorming, team presentations) as popular among interviewers, but did not mention individual activities. The KIIs described how the pandemic impacted learning methods. In-person activities were perceived as more helpful than web-conference activities, the primary training mode during COVID-19. Results demonstrate the importance of incorporating group training activities in virtual settings and providing a variety of techniques to keep the interest of both newer and veteran staff and those who learn best through different learning styles. A mixed-methods needs assessment can inform national training and provide details that cannot be attained from a single method.

Care Experience Surveys: What Motivates Response?

Rachael Allen, *RTI International*
Lisa M. Lines, PhD, MPH, *RTI International*

As survey response rates fall, CAHPS ® survey users, including clinicians, healthcare managers and systems, accreditation organizations, professional associations, vendors, and researchers, have expressed concern about how declining response rates may affect the surveys' ability to represent the population of people receiving care and reflect people's experiences with healthcare. Strategies to increase response rates have included prenotification and follow-up attempts, shorter surveys, multiple modes of survey administration, official survey sponsorship, and communicating the salience of the survey to potential respondents. Between February 2019 and November 2020 we conducted 38 in-person and telephone interviews with people receiving home health care and their family members as part of a redesign of the Home Health Care CAHPS Survey. ® To assess factors that affect people's likelihood of responding to experience of care surveys, we asked participants if they had received similar surveys in the mail before, if they usually answer them, and what might make them more or less likely to fill out a survey that did not include an incentive payment.   Participants said they were more likely to respond to experience of care surveys if the subject matter was something that they cared about or were familiar with and if they knew it was from an official or legitimate source, which is consistent with prior research. However, some reported that deciding to complete the survey also depended on their satisfaction with the care they received and whether the results would improve the care and services that others might receive in the future. Thus, the quality of care provided, and respondents' altruism may also influence response rates on care experience surveys and other surveys of patients.

Let Me Put It This Way: Sensitive Topics in Questionnaire Design & Interviewing

Interviewer-Respondent Interactional Patterns across Question Sensitivity and Respondent Race

Tiffany Neman, *UW MADISON*
Jennifer Dykema, *University of Wisconsin-Madison*
Dana Garbarski, *Loyola University Chicago*
Cameron Jones, *University of Wisconsin-Madison*
Nora Cate Schaeffer, *University of Wisconsin-Madison*
Dorothy Farrar Edwards, *University of Wisconsin-Madison*

Sensitive survey questions—that is, questions that ask about seemingly personal or intrusive topics, like personal income or substance abuse—have long been associated with poorer data quality outcomes, including greater response bias and item and unit nonresponse rates (Schaeffer 2000; Tourangeau and Yan 2007). Less understood, however, are the precise features of the interviewer-respondent interaction that contribute to these outcomes. What common behavioral patterns do interviewers or respondents exhibit when sensitive questions are asked, and how do interviewers deal with problems that arise when respondents do not answer sensitive questions? The current study addresses these gaps in the literature by examining several interactional patterns that occur when sensitive questions are broached. Given recent evidence of racial differences in respondent interactional behaviors (Holbrook et al. 2006; Johnson et al. 2015), we further extend our analysis to investigate how interactional patterns vary across the interplay of question sensitivity and respondent race/ethnicity. Using transcriptions of telephone interviews (n =375) from the Voices Heard Survey, we systematically code respondent verbal behaviors associated with comprehension and mapping difficulties, including tokens ("um" or "uh"), non-paradigmatic question-answer sequences (non-codable answers), expressions of uncertainty (e.g., requests for repetition), and response times and latencies, among others. We rely on a set of mixed-effects models to test the associations between these behaviors and question sensitivity, measured using mean sensitivity ratings from an expert review panel. Preliminary findings suggest that, as question sensitivity increases, respondents are significantly less likely to express behaviors—specifically verbal tokens, expressions of uncertainty, and non-codable answers—that are indicative of cognitive processing problems, with no significant differences in results across respondent race or ethnicity. Our analysis highlights the importance of the tools survey researchers use to evaluate data quality and the complexity of their interpretation, especially when sensitive questions are asked.

Who Are You? Check All That Apply: Measuring Race, Ethnicity & Self Identification

Questioning Identity: How a Diverse Group of Respondents Answer Standard Questions about Ethnicity and Race

Dana Garbarski, *Loyola University*
Jennifer Dykema, *University of Wisconsin-Madison*
Tiffany Neman, *University of Wisconsin-Madison*
Cameron Jones, *University of Wisconsin-Madison*
Nora Cate Schaeffer, *University of Wisconsin-Madison*
Dorothy Farrar Edwards, *University of Wisconsin-Madison*

How ethnoracial identity is measured in surveys has implications for understandings of disparities and inequities. However, individual and societal understandings of ethnicity and race are socially constructed, contextually contingent, fluid, and not mutually exclusive. Relatedly, the ways in which people identify themselves may not conform to the standardized survey questions used to measure ethnicity and race, leading to potential measurement error for researchers and othering of respondents. In interviewer administered surveys, answers to survey questions are an interactional achievement, accomplished through interviewer-respondent interaction. We know very little about the response process that produces the survey measurement of ethnicity and race. Examining the interviewer-respondent interaction illustrates whether, when, how, and why issues with these survey questions emerge for respondents. We examine the following research questions in an ethnoracially diverse sample of respondents: 1)   Are there behavioral indicators of interactional problems when respondents answer questions about ethnicity and race? Do these vary across ethnoracial groups? 2)   How do interviewers respond in the presence of these behavioral indicators of interactional problems? In what ways are they (non)responsive? Data are provided by the 2013-2014 "Voices Heard" survey, a computer-assisted telephone survey designed to measure respondents' perceptions of barriers and facilitators to participating in medical research. Interviews (n=410) were conducted with a quota sample of respondents

in the following groups: White, Black, Latinx/Latine, and American Indian. We examine features of the interaction for two questions about Hispanic/Latino ethnicity and race written to align with common survey practice. The features of the interaction examined include paradigmatic question-answer sequences, indicators of respondents' difficulty or concerns, and interviewers' rapport-building behaviors in the presence of respondents' difficulty or concerns.

Response Formats for Race and Ethnicity Measurement in the U.S.
Frances Barlas, *Ipsos*
Kip Davis, *Ipsos*
Randall K. Thomas, *Ipsos*

Much has been made about the large change in the racial composition of the United States as indicated by the latest Decennial Census figures with many more Americans identifying as more than one race compared to 10 years ago. In the United States, race and ethnicity are related to many life experiences and outcomes, yet conceptual and measurement problems in classification abound. For measuring Latino origin, a common question format asks 'Are you of Latino, Latino, or Spanish origin?' and provides 5 responses, 4 of which indicate 'Yes' with different countries of origin. Sometimes such surveys do not allow for the possibility that a person has more than one country of origin. Other surveys ask a simpler 'Do you consider yourself to be Latino or Latino?' and provide a yes or no response. When asking about race, some ask 'What is your race?' while others ask 'What race or races do you consider yourself to be?' The item stem asking about a singular race may be more likely to prime people to think of a single race than the multiple response format. An often-discussed modification is a merging of the two questions (race and Latino ethnicity) into a single question. We designed an experiment where respondents were randomly assigned to the traditional two question format or an alternative one question format that combines race and ethnicity measurement. The experiment we report on here randomly assigned people to 1 of 4 conditions – 2 items with singular race referenced in the question stem, 2 items asking race or races, one question with singular race in the question stem, and two questions with multiple races. We had over 2000 respondents from Ipsos' KnowledgePanel, a probability-based online panel, complete this web-based experiment. We report on the significant differences we obtained between formats.

Using Race Data Upcoding to Advance Equity and Data Disaggregation: Reporting Multi-Racial Identities and Increased Granularity of Racial/Ethnic Origins
Andrew Juhnke, *UCLA Center for Health Policy Research*
Jiyoun Yoo, *UCLA Center for Health Policy Research*

For the first time in the 2021 version of the California Health Interview Survey (CHIS), white and black race origin write-in questions were included on the questionnaire. CHIS 2021 respondents who self-identified as white or black in the initial race question were then asked a follow-up question about their origins. There are multiple goals of this new addition to CHIS, including being able to identify Middle East and North Africa (MENA) respondents, as well as providing the same level of origin granularity for all races as was already included in the survey for those who self-identify as Asian, Native Hawaiian or Pacific Islander, and American Indian/Alaska Native. For years, CHIS has maintained a rigorous race upcoding process, through which we work in conjunction with our data collection vendor to create as complete of a race/ethnicity variable set for each respondent as possible. We use the information provided in the various race/ethnicity and related questions, including those with other write-in response options. This race upcoding process allows for us to accurately capture multi-racial identities, and the new inclusion of the white and black origin write-ins enhances this further. This level of detail helps serve CHIS' goals of advancing equity and increasing broader subgroup representation in our data. Additionally, it provides researchers the ability to do deeper analyses of racial/ethnic minorities that may be underrepresented, underserved, and in need of evidence-based research to advance programs and policies to support them. We examine: (1) the general CHIS race upcoding process, detailing how the process helps identify multi-racial identities, and (2) CHIS' new white and black origin questions and the potential uses for this additional layer of data gathered by the survey.

Measuring Caste in India: Reconciling Measurement Differences Between Official Statistics and Survey Results
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*

India's caste system is an ancient social hierarchy based on occupation and economic status. People in India are born into a particular caste and tend to keep many aspects of their social life within its boundaries, including whom they marry and whom they choose to count as their close friends. Despite the caste system's significance in Indian society, there is no consensus on what proportion of Indians belong to each caste category. One commonly cited source for this information is the nation's 2011 census, but its methods for determining caste have been criticized for anomalies and errors. For example, only certain religious groups can report being "Scheduled Caste" (commonly known as Dalits) on the census, despite the fact that Indians from all religions do personally consider themselves members of this group. Surveys have also measured caste in India, but they tend to use different methodologies and generally produce different estimates from those produced by the census. This paper explores the methodological differences between the country's census and national surveys — including a Pew Research Center survey of roughly 30,000 adults in India— when it comes to measuring caste in India. The paper aims to help researchers identify several potential reasons why official statistics may differ from survey results—such as question wording, sample universe, weights and more. Deviations between the census and surveys provide important findings about caste identity in India.No Concurrent Session X

Sampling Frames, Panel Construction, & Weighting
The Challenge of Recruiting Rare and Physically Hard-To-Reach Populations: Using RDS To Recruit Overseas Labor Migrants

Holly Koogler, *ICF*
Yun Kim, *ICF*
Ronaldo Iachan, *ICF*
Suteera Nagavajara, *ICF*
Kato Francis, *Makerere University*
Eddy Walakira, *Makerere University*
Bart Robertson, *Global Fund to End Modern Slavery*

Collecting representative data from respondents located in areas inaccessible to researchers has long been a significant challenge across many fields. One potential solution is to use multi-modal respondent driven sampling (RDS). RDS is a popular and growing method of recruiting rare populations; however, there has been limited discussion about the use of RDS for physically hard-to-reach populations. In this presentation we will share details of the design and implementation of a recently completed study of 400 labor migrants from Uganda to the Middle East. Collecting labor condition data among migrants working in the Middle East is challenging due to the inaccessibility of the workers. The most common solution is to collect data among returned migrants; however, this solution limits the utility of the data since the method does not allow for the collection of current working conditions. We addressed this challenge by using returned migrants as seeds (initial respondents) who then recruited fellow migrants, many of whom were still working abroad. Interviews with overseas migrants took place by voice call. Using this method, we were able to estimate the prevalence of forced labor among Ugandan migrants working in the Middle East without the lag normally affecting this kind of research. The presentation will share technical and methodological challenges in undertaking this methodology as well as the ultimate successes. Challenges include the representation of a multifaceted population in terms of countries of and sector of work, both with more than 8 meaningful categories, and the balance between diverse seeds and longer chains. It will draw comparisons with a face-to-face RDS study completed among 200 sex workers in Uganda during the same year. The study offers insights into recruiting physically hard-to-reach populations with diverse future applications, including research in conflict zones and in countries with repressive governments.

Building a Panel to Better Understand the Experiences of Diverse Populations in Canada
Cilanne Boulet, *Statistics Canada*

In 2021, Statistics Canada initiated the Disaggregated Data Action Plan, a multi-year initiative to support more representative data collection methods, enhance statistics on diverse populations to allow for intersectional analyses,

and support government and societal efforts to address known inequalities and bring considerations of fairness and inclusion into decision making. As part of this initiative, Statistics Canada is building a new probabilistic panel specifically designed to collect of data that can be disaggregated according to visible minority group. This new tool will allow us to address data gaps and emerging questions related to diversity. This talk will give an overview of the design of this panel.

Constructing and Evaluating a Sample Frame Built From Disparate Sources: a Survey of City, State and Organization Leaders

Martha Kovac, *Mathematica*
Shauna Robinson, *Mathematica*
Emily Hoe, *Mathematica*
Eva Ward, *Mathematica*
Jared Coopersmith, *Mathematica*

To a large extent the quality of the sample frame dictates whether any probability sample can be described as representative of the target population. The quality of a sample frame can be defined by population coverage, but also by availability of auxiliary information like covariates or contact information for the respondents. In some sectors, like public education, sample frames with high coverage and auxiliary information are available from federal agencies, while in other sectors, like for-profit businesses, sample frames are available from vendors for a fee. In yet other sectors, such as economic and community development organizations, frames with little auxiliary information or no frames at all are available. Mathematica fielded a probability survey to a multi-sector sample of city, state and organization leaders and influencers on perceptions surrounding the social determinants of health and health equity, sponsored by the Robert Wood Johnson Foundation. Sectors in the study included economic and community development, education, employment, housing, healthcare services, parks and recreation, physical safety, public health, and transportation. No single sample frame covered this diverse set of organizational leaders, so we used a variety of sources to construct the sample frame of leaders, including high- and low-quality government databases, sample vendors, and manually creating list frames through internet searches. We also solicited respondent referrals from the sampled respondents to help fill holes in the frame coverage. In this presentation we'll describe the data sources we used, the challenges we faced, and the results of the survey fielding in terms of available contact information, eligibility, and response rates, by source and by sector. We'll describe the particular challenges in frame construction to support an inclusive survey fielding.

The Drop Point Address Challenge for a Mail Survey, Experiences with the 2021 NYC Community Health Survey and 2021 NYC KIDS Survey

John Sokolowski, *Abt Associates*
Nicholas Ruther, *Abt Associates*
Theresa Stroble, *Abt Associates*
Wyatt Bush, *Abt Associates*
Michael Witt, *Abt Associates*
Martha McRoy, *Abt Associates*
Amber Levanon Seligson, *NYC Department of Health and Mental Hygiene (DOHMH)*
Steven Fernandez, *NYC Department of Health and Mental Hygiene (DOHMH)*
Michael L. Sanderson, *Utah Department of Health*

A challenge of designing and collecting data via a mail survey is addresses serving multiple housing units, otherwise known as drop points (DPs), which make the selection of a unique housing unit difficult. DP addresses are common in NYC, where recent data suggest they account for 18% of addresses representing 20% of the adult population. In 2021, the NYC Community Health Survey (CHS) and KIDS Survey, conducted by the NYC Department of Health and Mental Hygiene (DOHMH), transitioned from telephone-based sample designs to multimode list frame/ABS designs. The CHS is an annual survey that collects data from 10,000 randomly selected adults that enables researchers to track the health of New Yorkers and measure the extent to which diseases and risk factors occur at the neighborhood, borough, and city levels. The KIDS Survey collects data every other year on the health of 7,500 children in NYC between the ages of 1 and 13 and provides data on the health care needs of children. The 2021 data collection protocol for both surveys involved a

combination of multiple mail solicitations asking respondents to complete the survey via web, with a paper survey mailing or a telephone option for nonrespondents. This presentation will discuss how the sample design, sample selection and mailout strategies were optimized for DPs. Key points in the presentation will include utilizing a multistage sample selection approach for DPs and supplementing standard ABS frame information to facilitate mailings. The supplementation process included expanding DPs from the single address obtained in the ABS frame to the unit-level by editing DP addresses and conducting additional research to create appropriate apartment/unit number for mailing. The presentation will include a comparison of survey estimates between DPs and non-DPs as well as a discussion of how this information will be used to optimize future surveys

Sometimes, It's Worth the Weight: Complexity of Combining ABS and RDD Frames

Kathleen Call, *University of Minnesota*
Giovann Alarcon, *UMN SHADAC*
Alisha Simon, *MN Department of Health*
Sarah Hagge, *MN Department of Health*
Karen Turner, *UMN, SHADAC*
Kathleen Panas, *MN Department of Health*

The Minnesota Health Access (MNHA) survey represents the primary source of uninsurance and other policy relevant data in Minnesota. The MNHA consisted of a dual Random Digit Dialing frame (RDD) that drew sample from landline and cellphone frames. In 2019, the MNHA introduced an Address-Based Sampling frame (ABS) as a significant portion of its sample. RDD frames were retained in 2019 to better assess the continuation of trends in key indicators. The main challenge of introducing an ABS frame was that this frame could yield a sample with different characteristics compared to the RDD sample, which oversampled prepaid cellphones. Previous research shows that oversampling prepaid cellphones is a cost-efficient way to increase respondents who are more likely to be BIPOC, have lower income, have lower levels of educational attainment, rent their home, and lack health insurance. This paper describes the experience of making different weighting adjustments as we appended the samples from two RDD (landline/cellphone) and one ABS frames to obtain good representation of the Minnesota population. In addition to specific demographic characteristics, these different frames could also differ in terms of health insurance status, leading to potential unintentional bias of results. Compared to the cellphone frame, the ABS frame provided a sample that has higher educational attainment, higher income, larger households (3+ members), owns their home, has internet access, and identifies themselves as White. The MNHA weighting strategy also had to take into account that the oversampled prepaid portion of the cellphone frame produced a sample with an uninsurance rate (11%) that more than tripled that of the ABS (3.2%) and landline (3.7%) frames. Overall, we assessed nine different weighting approaches, which included adjustments to both appended and post-stratification weights. The design effects showed high variation between approaches, however, the uninsurance rate only ranged between 5.2% and 4.7%.

Accumulating Retrospective Data Across Surveys

Mahmoud Elkasabi, *ICF*

Although the Demographic and Health Surveys (DHS) are cross-sectional surveys that aim at producing estimates based on the "current" situation, some data, such as birth histories, are collected based on reports of retrospective events or actions. In countries where several DHS surveys were collected in a relatively short-time span, such retrospective data might be available for the same calendar year or period from several surveys. Such overlap in the retrospective data across DHS surveys can be exploited for the benefit of the survey indicators. In this research, we propose the pooled approach to accumulate retrospective data across consecutive DHS surveys into one pooled dataset. Such an accumulated dataset can be used for different purposes such as to 1) produce national and regional indicators with better precision and/or based on shorter reference periods; 2) produce sub-regional indicators with good precision. In this research, we accumulate birth histories across different DHS surveys, and examine the performance of childhood mortality indicators based on the accumulated data. Initial results indicate that the estimated standard errors of the cumulated indicators are less than their counterparts from separate surveys, and accumulating birth histories smooths fluctuations in time series for national and subnational indicators. In addition, results indicate that cumulating data establishes time trends that are more stable and reliable than those based on data from single survey data.

Building a Geographically-Stratified Internet-Based Panel

Matin Mirramezani, *The Generation Lab*
Rebecca Oh, *The Generation Lab*
Cyrus Beschloss, *The Generation Lab*

Compared to traditional methods, emerging online survey methods have had equal - and usually higher - access to young respondents. According to the Pew Research Center, among young people, 100% are internet users, 96% own a smartphone, and 90% use at least one social media site. Guided by these statistics and borrowing from high-quality survey methods, the Generation Lab has developed a geographically-stratified internet-based methodology to survey young people across the United States. In our patent-pending approach, we leverage the near-universal reach of the internet to create a non-probability panel with the accuracy of a probability-based panel at a lower cost and higher speed. Although this approach does not yield a probability-based panel, we utilize a set of constraints and settings to reduce targeting and enhance randomness. We target a set of randomized geographic units, selected by applying probability proportional to size (PPS) sampling, to recruit participants and use location targeting based on physical location. We also narrow our target population to people aged 18-29. We further validate the demographic information through our intake process. The resulting panel is monitored across a variety of demographic variables to ensure representativeness compared to the population statistics. We employ a stratified random method of sampling. Surveys are deployed and sent out to the respondents who satisfy the qualification of our intended population. Throughout the polling process, we utilize a variety of features to ensure the integrity of the data that we collect, including survey protocols preventing multiple responses and flagging invalid inputs. An honorarium is given to the sampled respondents who complete the surveys.

How are you Counting All Eyeballs? Review of Audience Duplication Estimation Techniques in Cross-Media Measurement

Kumar Rao, *Comscore, Inc*
Michael Vinson, *Comscore*

It is no surprise that technology is radically changing the media industry. Watching video content on platforms other than traditional television sets, such as computers, smartphones, tablets, and over-the-top devices has become ubiquitous in our digital lives. Measuring audiences accurately across all these platforms has made one particular task of cross-media measurement – estimation of audience duplication – more complex than ever. The topic of audience duplication has received research attention in the past for estimating duplication between media vehicles such as television and magazines (Rust, Zimmer, & Leone, 1986), websites (Danaher, 2007; Danaher, Lee, & Kerbache, 2010), and programs on television (Abe, 1997). However, to date, there is little to none published research that has examined audience duplication in a cross-media content measurement context involving television and digital media vehicles. In this study, we bring to light five different audience duplication estimation techniques in cross-media measurement via a real-life measurement case study. These techniques are designed to produce an estimate of audience duplication behind a pair of media vehicles with certain sampling and measurement properties and distributional assumptions about the underlying data. The paper will describe the complex nuances in real-life audience duplication examples in television and digital measurement, and how those estimates vary as a function of content hierarchy (ex: Program-level, Episode-level, etc.). Moreover, we will examine the possibility of estimates that use as little information as possible, as a Bayesian update on a low-information prior. By calculating duplication, which therefore enables the estimate of a deduplicated total reach metric, from only a small amount of information, we enable a variety of cross-platform estimates that would otherwise be unavailable. We discuss the audience duplication results from all estimation techniques and suggest avenues for future research

Mixed Mode Surveys: Implementation & Evaluation
Using an Adaptive Mode Adjustment to Improve Accuracy in a National Web and Paper Survey

Courtney Kennedy, *Pew Research Center*
Arnold Lau, *Pew Research Center*

Andrew Mercer, *Pew Research Center*
J. Michael Brick, *Westat*
Ismael Cervantes-Flores, *Westat*

While multimode designs are increasingly popular in the U.S., it is not entirely clear how best to statistically combine cases from the various modes. The typical approach assumes that all respondents should be treated the same regardless of mode, even though it is well known that the response patterns by mode vary substantially. For example, adults responding online tend to be younger, more politically progressive and have more formal education than those responding by paper. Brick and colleagues (2021) propose an adaptive mode adjustment to address these differences and reduce bias in combined study estimates. This project explores how an adaptive mode adjustment performs in Pew Research Center's 2021 National Public Opinion Reference Survey (NPORS). Functionally, this adjustment increases the influence of mail respondents relative to web respondents because mail respondents are more representative of the entire U.S. adult public on most demographic dimensions (age is an exception). The 2021 NPORS featured a national sample from the U.S.P.S. Computerized Delivery Sequence File and a 29% AAPOR RR1 response rate. Respondents were first invited to respond online, and nonrespondents were later invited to respond via paper. Most (61%) of the n=3,937 respondents answered by mail while 39% did so online. This presentation examines how the adaptive mode adjustment affects weighted survey estimates and their precision. We focus in particular on whether the adjustment helps to improve representation of politically conservative adults and/or those with lower levels of technology usage.

Methods and Improvements for Yield Rate Estimation in Survey Planning

Joseph McMichael, *RTI International*
Joe Murphy, *RTI International*
Patrick Chen, *RTI International*
Rachel Harter, *RTI International*
Paul Biemer, *RTI International*
S. Grace Deng, *U.S. Energy Information Administration*

The 2020 Residential Energy Consumption Survey (RECS), conducted by the U.S. Energy Information Administration, included over 50,000 sampled households from the 50 US states and the District of Columbia. The survey was designed to provide precise estimates of residential energy consumption and household characteristics at the state level. The survey contacted sampled households by mail, inviting them to participate in either a web or paper survey. Stratified by state, the RECS sample was allocated based on the number of expected completes divided by the estimated yield rate for each state. The state-level yield rate estimates were based on results from the 2015 RECS and the 2015 RECS National Pilot. In states where the past survey sample was deemed too small to be reliable for a RECS yield-rate estimate, a linear model was fitted that predicted the yield rate using American Community Survey estimates at the block-group level. Although historical and model-based estimates worked better than assuming no yield differential among states, RECS still required adjustments to the size of planned sample releases to accommodate differences in expected and realized yield rates. Estimating a yield rate for sample allocation is critical for self-administered surveys with mail contact. Inaccurate yield estimates can lead to misapplication of survey resources with budget overruns, failure to achieve sample targets, or waste when too many cases are released. This paper discusses the yield-rate estimation methods used on RECS and proposes potential improvements that can leverage some general use, address-level, yield prediction models to help survey researchers improve sample planning and allocation. We present models using both public and privately available data comparing the strength and weaknesses of each. In addition, we test the ability of address-level models to predict web vs. paper mode choice.

Overcoming Challenges of Conducting a Non-Probability Survey on the Impact of COVID19 on Ethnic Minorities in the United Kingdom

Natalie Shlomo, *The University of Manchester*
Nissa Finney, *St. Andrews*
James Nazroo, *University of Manchester*
Laia Becares, *The University of Sussex*
Dharmi Kapadia, *University of Manchester*
Harry Taylor, *University of Manchester*

Daniel Ellingworth, *University of Manchester*
Neema Begum, *University of Nottingham*
Rebecca Wilson, *Ipsos*

The Centre on Dynamics of Ethnicity (CoDE) led from the University of Manchester in the United Kingdom, in partnership with Ipsos-MORI, designed the Evidence for Equality National Survey (EVENS) in 2021 to understand the impact of the COVID19 pandemic on ethnic and religious minority groups. This paper documents the methodological and practical challenges, and the innovations to overcome them, to implement the survey, the largest and most comprehensive survey of ethnic and religious minorities undertaken in the UK. Standard probability-based surveys, even with ethnic minority group boosts, do not have the sample sizes required to obtain reliable estimates for small group statistics, and typically adopt sampling methods that exclude smaller ethnic groups or those living in predominantly White areas. We discuss the challenges encountered on setting up and administering a non-probability based sample to overcome these limitations, using innovative recruitment strategies which included strategic partnerships with Voluntary Community and Social Enterprises, media partners, social media channels, and sampling existing panel members in non-probability and probability-based panels. Accessibility was enabled by translation of EVENS into 14 languages and web and CATI options. Participants first went through a registration survey to test eligibility and were then supplied a link into the main survey. Some were also given further links to snowball to family and friends. Daily monitoring of the data collection against desired sample sizes and R-indicator calculations allowed the team to focus attention on recruitment of specific groups   in a responsive data collection mode. 15,000 responses, which included a general population sample for comparative analyses, were achieved. To our knowledge, this is the first survey of its kind that relied on an opt-in non-probability sample with an incentive and we describe the measures applied to mitigate erroneous responses, the quality checks on the data and the resulting quality of the sample.

Impact of Questionnaire Mode on Estimates of Health Insurance Coverage and Vaccines for Children Program Eligibility in the National Immunization Survey

Rachel Francis, *NORC at the University Of Chicago*
Zachary Seeskin, *NORC at the University of Chicago*
Amie Conley, *NORC at the University of Chicago*
Laurie D. Elam-Evans, *Centers for Disease Control and Prevention*
Chalanda S. Smith, *Centers for Disease Control and Prevention*
Holly A. Hill, *Centers for Disease Control and Prevention*
Michael Chen, *Centers for Disease Control and Prevention*

The National Immunization Surveys (NIS) monitor vaccination coverage and access in the United States among children ages 19-35 months (NIS-Child), adolescents ages 13-17 years (NIS-Teen), and for influenza vaccination, children ages 6 months-17 years (NIS-Flu). The NIS currently operates as a random-digit-dial (RDD) mobile telephone survey. However, from January 19 through July 9, 2021, a pilot was conducted to assess the viability of using an address-based sampling (ABS) approach with multiple modes of data collection, including a self-administered web survey and self-administered paper-and-pencil interviewing (PAPI) in addition to telephone interviewing. The ABS pilot used modified versions of the current NIS RDD questionnaires to accommodate multiple modes of data collection, with one of the most significant modifications related to the way health insurance coverage is assessed. In particular, the addition of self-administered questionnaire formats and space limitations of the PAPI component necessitated the consolidation of multiple health insurance questions from the RDD into a single health insurance question, and removal of interviewer help text and state-specific text that is automatically populated when using a computer assisted telephone interviewing (CATI) system. We compared estimates of health insurance coverage and eligibility for the Vaccines for Children (VFC) program obtained from the RDD and ABS versions of the NIS questionnaires and found estimates from the ABS NIS pilot indicated higher rates of private insurance coverage and lower rates of coverage with Medicaid and other public insurance types compared to the RDD NIS. Additionally, VFC eligibility was higher in the RDD NIS compared with the ABS NIS version. These differences were most prominent from web interviews and appear to be attributable to a combination of demographic and questionnaire differences. We will discuss how questionnaire design and questionnaire administration mode may influence health insurance coverage reporting and estimation and VFC eligibility.

Improving Inferences Based on Survey Data Collected Using Mixed-Mode Designs

Wenshan Yu, *University of Michigan, Ann Arbor*

Trivellore E. Raghunathan, *Professor of Biostatistics, School of Public Health; Research Professor, Institute for Social Research, University of Michigan*

Michael R. Elliott, *Professor of Biostatistics, Department of Biostatistics, University of Michigan; Research Professor of Survey Methodology, Survey Methodology Program, University of Michigan*

Mixed-mode designs have become increasingly common in survey data collection. Although different modes often have different measurement properties, standard practice is to treat mixed-mode data as if they had been collected with a single mode, neglecting the potential impact of mode effects. To account for potential mode effects when making inference for mixed-mode samples, we propose 1) a "testimator" approach, 2) a Bayesian approach, and 3) a model averaging extension of the Bayesian approach. In the "testimator" approach, we test whether the means and the variances of mixed-mode samples are the same using t-tests and F-tests respectively. If the means are the same, we take the average of mode-specific estimates. If the means are different, we take the average when we have no prior information about preferred modes and take the smaller (or larger) estimate when we have prior information about preferred modes (e.g., a smaller estimate is better). In the Bayesian approach, we use a data-driven method to determine whether there are mode effects. If there are no mode effects, we draw inference using a common mean model. If there are mode effects, we draw inference using the data collected with the mode that produces smaller estimates. In the model averaging approach, we combine estimates of different models (characterized by whether assume same means and variances across modes) using marginal posteriors as weights. We use propensity score stratification to control for selection effects when auxiliary variables are available. Compared to existing approaches, our approaches can account for model uncertainties, thus achieving robust inference. We evaluate the approaches in simulation studies and find that they largely reduce bias and have good coverage rates in certain scenarios. We will apply the methods on Arab Barometer survey data, where face-to-face is the benchmark mode, and telephone is the comparison mode.

Now You're Speaking My Language: Advances in Survey Translation

Multilingual Communities Require Multilingual Surveys: A Language Justice-Informed Approach to the NYCHVS

Caitlin Waickman, *New York City Department of Housing Preservation & Development*
Allison Corbett, *New York City Department of Housing Preservation & Development*

New York City has a population of 3.2 million residents born outside of the United States. Nearly half of them speak a language other than English at home, and almost a quarter have limited English proficiency. For surveys fielded in diverse areas, language access is an issue not only of equality, but also data quality. Common approaches, such as relying on interpreters or deferring to household members who speak English, can lead to the mis-phrasing of questions, discomfort of the respondent, or the collection of incomplete information. To address this issue, sponsors of the New York City Housing and Vacancy Survey (NYCHVS) worked with experts to develop and implement a language justice approach to data collection. Language justice aims to decenter English in an effort to enable all involved to equally participate fully, regardless of the language they speak. In 2021 the NYCHVS was fielded in seven languages. To apply a language justice framework, the sponsor team considered all components of the survey that respondents may interact with—materials for outreach, recruitment, the interview, and leave-behind resources. The survey itself was developed using an Advance Translation approach and reviewed by local bilingual New Yorkers for cultural appropriateness. Final translations were programmed for easy access for bilingual field representatives. All field representatives received extensive training on the importance of language access and the related procedures. A holistic approach to language justice is essential for an equitable approach to surveys, especially in diverse areas. The sponsor team for the NYCHVS will share their experience in developing their approach and best practices for any researchers working in a multi-lingual context

How Much Does Translation Matter? Effects of a Language Match in the Questionnaire on Response Rates in Online Surveys

Madeleine Siegel, *DeZIM*
Almuth Lietz, *DeZIM*

For international, multicultural and multiregional surveys, a high-quality translation of the questionnaire is of central importance. In addition to a general increase in the response rate, questionnaire translation also aims in particular at minimizing response bias, e.g., caused by incorrect answers and reduced comparability between the different respondents' answers due to a misunderstanding of survey content. Here, we are mainly interested in different variants of non-response affecting data quality. In our analysis, we therefore investigate which positive aspects a high-quality translation entails for survey completion, response rate and participants' satisfaction. We assume that respondents who are confronted with a questionnaire in their everyday or mainly used language will be more willing to complete the survey and particularly satisfied with it. As an indicator for survey satisfaction, we consider the participants' overall survey evaluation at the end of the questionnaire, which might also influence their willingness to participate in future waves. However, since translation is very time-, personnel-, and resource-intensive, we want to examine to what degree and under what circumstances the effort is worthwhile. The analysis is based on the first wave of the online access panel DeZIM.panel, which is conducted by the German Center for Integration and Migration Research (DeZIM) and goes into the field in November 2021. As part of the DeZIM.panel, approximately 7,000 individuals have agreed to participate in repeated surveys on various topics. The recruitment sample of the panel is representative of the population with an oversampling of certain subgroups. Based on this oversampling, the questionnaire was translated into five different languages (Turkish, Russian, Arabic, Portuguese, English), which allows us to examine the effects of a language match in more detail for different languages. Our article presents first results and gives hints for practical survey research concerning questionnaire translation.

Su Experiencia nos Importa: An Experiment in Conducting a Dual-Language Household Survey

Scott Leary, *Internal Revenue Service*
Rizwan Javaid, *Internal Revenue Service*
Kris Pate, *Internal Revenue Service*

Brenda Schafer, *Internal Revenue Service*
Patrick Langetieg, *Internal Revenue Service*

Spanish is the second most prevalent language spoken in the United States after English, with an estimated 13.5% of individuals age 5 and older speaking Spanish at home. Of those, 39% speak English less than "very well" (U.S. Census Bureau, 2019). It is vital to obtain survey responses from Spanish speaking individuals, but studies have shown that the survey language impacts response. (U.S. Census Bureau, 2019). The Individual Taxpayer Burden (ITB) Survey, which gathers data about the time and money taxpayers spend complying with their tax reporting responsibilities is a multi-mode survey offered in both English and Spanish. While both versions are available on the web, only English surveys are mailed automatically. Participants must request a Spanish survey. Historically, less than 1% of respondents completed the survey in Spanish. In an attempt to increase Spanish-language completes in 2020, we conducted a dual-language experiment where half of survey participants (n=15,000) received all communication and survey materials in English and Spanish. This experiment is a continuation of research by Brick and Colleagues (Brick et al 2012; 2019) that found that dual-language surveys increase response rates in areas high in Hispanic populations while having little to no effect on the remaining population. Our study expands on this research by comparing a full dual-language mode to a multi-mode administration.   We will examine the effect of a dual-language administration on overall response rates and frequency of Spanish-language completes. This analysis will include a comparison of respondents in communities with a high and low concentrations of Spanish speaking residents based on Census data. In addition, response rate by other demographic variables (e.g., strata, age, and income) will be considered. Lastly, we will include a cost-benefit discussion of administering a dual-language multi-mode survey.

Lost in Translation: Can Google Translate Provide Researchers With an Effective Way to Code Non-English Open-Ends?

Laura Silver, *Pew Research Center*
Patrick van Kessel, *Pew Research Center*
Christine, *Huang*
J.J. Moncus, *Pew Research Center*

Working with open-ended data is no easy feat at the best of times, but it is particularly difficult when multiple languages and cultures are involved. How should one explore how people talk about an issue in their own words while balancing potentially conflicting goals related to accuracy, money and time? We explore this using a recent Pew Research Center project where respondents were asked an open-ended question about the meaning of life. Center researchers received nearly 18,000 open-ended responses spanning 17 societies and 14 languages. They developed a codebook and trained a group of translators to apply the codebook to the non-English responses. The translators' coding performance was evaluated against a baseline of sample responses that were translated into English and coded by Center researchers. We compare the results of this method against an alternative approach: translating the non-English responses into English using Google Translate and coding them in-house. Each approach necessitates a trade-off. While translators are more likely to grasp the correct meaning of a response by evaluating it in its native language, they may have a less nuanced understanding of the codebook than the researchers themselves. On the other hand, while in-house researchers may apply the codebook more accurately, doing so requires either costly professional translation or machine translation, the latter of which may fail to capture the correct meaning of certain responses. Results indicate that while Google Translate may be sufficient for simplistic codes in a codebook, it does not function as an effective alternative to professional translation with more nuanced codes – and it is significantly weaker in certain languages (e.g. Greek) than others (e.g. Spanish). We evaluate the performance of Google Translate as applied to this corpus of text. We also discuss the relative pros and cons of these approaches when considering time and money.

Adapting the Advance Translation Method for Use at Large U.S. Survey Organizations

Patricia Goerman, *US Census Bureau*
Mikelyn Meyers, *U.S. Census Bureau*
Brita Dorer, *GESIS*
Elyzabeth Gaumer, *New York City Department of Housing Preservation and Development*

Working towards comparability across translations of survey instruments is a complex and challenging task faced by survey organizations throughout the world. The European Social Survey (ESS) has implemented a method called Advance Translation, which strives to simplify and improve the source questionnaire wording as a part of the source questionnaire design process, with a view to facilitate final translation (Dorer 2011; Fitzgerald 2015; Dorer 2020). Survey methodologists typically start with a draft source version of a questionnaire which is sent to translators for an initial "quick draft" translation. The translation is done independently by teams of translators, ideally in each language. The translators then review their first draft translation and code each question to indicate whether the source version was difficult to translate and whether they recommend any changes to the source wording prior to working on a final translation. In the final translation phase, the team members use the committee approach, consisting of translators and survey experts creating a single translation together with comments on the source text within each language. This consensus version includes the comments on which the translation team has agreed in their Review meeting. Implementing this method in the context of different survey organizations can require some adaptation, for instance, surveys run by Eurofound have made some adaptations to the method. In this talk we focus on modifications and lessons learned from using the method in the U.S. context, in a recent iteration of the New York City Housing and Vacancy Survey. For example, in this project we faced challenges staffing in-house translators with the appropriate expertise, and ultimately used a combination of contracted translators and subject matter experts recruited from the local community. We conclude with future research proposals to improve the ease of implementation of the method for different surveys in the U.S.

Using Social Media & Big Data to Predict COVID Beliefs

The Impact of Lifting Mask Mandates in the U.S.: A Study on Sentiment and Usage of Masks During the COVID-19 Pandemic

Brianna Richardson, *Momentive*
Sam Gutierrez, *Momentive*

The COVID-19 pandemic has introduced an array of new public health policies, mandates, and recommendations, including the wearing of face masks in public to help control the spread of the virus. Beginning in the spring of 2021, many of these mandates were lifted or relaxed. The details around the mandates varied widely by state, greatly impacting the usage and sentiment around their requirement. Previous studies have examined public opinion and sentiment on personal face mask usage during the pandemic. However, researchers have yet to determine how both face mask usage and sentiment around mask-wearing change together after states began to lift or relax mask mandates. Using a longitudinal panel of over two million online survey respondents collected by Momentive and Outbreaks Near Me from June 2020 to the present in conjunction with sentiment analysis performed on mask-related Tweets, this paper aims to measure if, and to what extent, adherence to mask-wearing and sentiment around face mask usage changed as states began to lift mask mandates. Findings from this study will contribute to the increasing body of research around adherence to, and sentiment around, public health mandates.

COVID-19 & BLM Movement Impacts on Resident Sentiment

Parker Quinn, *Polco*
Nick Mastronardi, *Polco*

We all know Covid-19 and the BLM movement of 2020 were big events, but how big were they?! This research uses a panel data set of cities across the country over a couple decades to ascertain time trends of resident sentiment along several dimensions of community livability and public safety. Disparities in resident opinions based on socio demographics also will be presented. Then, relative to that trend, the researchers use a regression discontinuity model to assess the specific impacts of these events for each demographic, in each city, along each quality of life dimension. The national database includes the assessments of community livability and government performance from more than 850 probability samples and 100+ non-probability paired samples.

Modeling COVID Concern in the United States

Samuel Solomon, *D3 Systems, Inc*
Jeffrey Yan, *D3 Systems*

As the COVID-19 pandemic spread across the globe during the first half of 2020, the fight against the disease quickly became an all-of-society effort. Individuals' willingness to wear masks and limit social interactions proved critical in slowing the spread. To that end, individuals' levels of concern about the pandemic were instrumental in the containment, or lack of containment, of the virus in their country. This paper explores which factors were associated with concern about the pandemic in the United States. Using data from a self-administered online survey completed by 1,094 American adults using SurveyMonkey's opt-in Audience Panel in June 2020, this paper explores different approaches to modeling Americans' concern about COVID, operationalized as agreement with the statement, "I am afraid that either myself or someone in my family may actually catch the Coronavirus." The paper will explore which demographic and attitudinal variables were the best at predicting whether someone is concerned about COVID. Because the pandemic was set against the backdrop of the 2020 US presidential election, the paper explores whether and how political identities and attitudes intersected with ideas about the pandemic. The paper will provide an example of how survey data can be used to understand correlates of concern about the pandemic, providing more texture to the question of how Americans responded to this once-in-a-century crisis.

A Partisan Pandemic: How COVID-19 Was Primed for Polarization

Josh Pasek, *University of Michigan*
Austin Hegland, *University of Michigan*
Annie Zhang, *University of Michigan*
Brianna Zichettella, *University of Michigan*

Despite the universal impact of COVID-19, Americans in different political parties rapidly formed diverging attitudes about the pandemic. Although other matters of scientific concern had previously elicited partisan reactions, Americans had diverged on those issues over decades rather than months. In this paper, we review evidence on factors that led to polarization of previous scientific issues to explain why reactions diverged so quickly this time around. Secondary analyses of publicly available survey data reveal that partisan reactions to the pandemic were closely associated with trust in public health institutions and that the association between partisanship and trust increased over time. In contrast, an alternative hypothesis that conservative media use might explain polarization was not supported. The conflation between trust and partisanship appears to largely explain polarized reactions to COVID-19 over time.

The Disparate Impacts of the COVID-19 Pandemic

"Protecting the Tenant" or "An Enormous Burden" : Public Perception of the Eviction Moratorium

Claire Kelley, *American Institutes for Research*
Sarah Kelley, *Child Trends*

The COVID-19 eviction moratorium has sparked fervent debate and extensive news coverage . While proponents (including the CDC and housing advocacy groups) point towards its role in reducing homelessness during the pandemic and preventing additional spread of COVID, opponents worry about the impact on landlords and the potential for abuse by tenants. Substantial research has attempted to quantify the impact of the moratorium on COVID incidence (e.g. Leifheit etal 2020; Nande 2021) or housing stability (Acosta etal 2020; Tsai 2021); less work has focused on public sentiment towards the eviction moratorium. Survey research suggests that during the peak of the pandemic, the eviction moratorium was supported by a majority of Americans (Pollack 2021), however very little work has examined the attitudes that underlie support or opposition for the eviction moratorium.   In this paper, we focus on public perception of the eviction moratorium measured through news coverage and social media mentions. To do this, we scraped news, Reddit and Twitter data and analyzed them using natural language processing techniques.   Our results suggest that public opinion was far more varied than simply "for" or "against". News coverage in favor of the moratorium focused on the plight of tenants impacted by COVID or the economic downturn, while anti-moratorium news outlets focused on the plight of small local landlords with reduced income. Social media presents an even more varied view: significant polarization exists in both directions with some advocates arguing for extension of the eviction ban and stricter pro-tenant policies, while across the aisle anti-moratorium advocates decried abuse of the legislation. These conversations also shifted as the pandemic and economic downturn progressed. Understanding these public

sentiment toward the eviction moratoria can help policy makers to better understand what policy alterations might make future housing policies more popular and more responsive to the public's needs.

When COVID-19 Hit a Fragmented Society, Ethnic and Religiosity Discrepancies - Examination of the Israeli Case 2020–2021

or anabi, *Israel Democracy Institute*
Tamar Hermann, *The Israel Democracy Institute and the Open University of Israel*

When hit by the COVID 19 pandemic in March 2020, Israel was in the midst of a prolonged electoral crisis, resulting from a stalemate between the two main political blocs. Nevertheless, the national public mood at the time was quite untroubled as the national economy and security seemed stable. However, after a few weeks the system was shattered. The competing partisan elites failed to join forces in combating the pandemic; large sectors of the economy failed to survive the government's imposed lockdowns and social distancing regulations, and the budget starved national health services were overstretched in attempting to meet the health challenge. This presentation will describe and analyze how Israeli public opinion responded to this "perfect storm" and its repercussions, based on a series of 15 surveys conducted by the Israeli Democracy Institute between March 2020 to December 2021. In particular it will focus on the deepening of the pre-existing, sometimes overlapping socio-political cleavages between: Jews and Arabs, secular and religious, rich and poor, and the voters of the political parties of the Left, Center and Right.

Differential Impacts of the Covid-19 Pandemic on Migrant Workers in the Gulf

John Lee Holmes, *The Social and Economic Survey Research Institute (SESRI)*
Abdoulaye Diop, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Michael Ewers, *Dept. of Geography and Earth Sciences, Univ. of North Carolina at Charlotte*
John Lee P. Holmes, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Lina Bader, *Qatar University Social and Economic Survey Research Institute (SESRI)*
Engi Almaghraby, *Qatar University Social and Economic Survey Research Institute (SESRI)*

The Covid-19 pandemic has affected everyone everywhere yet there remains a curious and significant variation in its impact across population groups and regions. Nowhere is this variation clearer than in the small Arabian Gulf country of Qatar, home to one of the highest proportions of migrant laborers relative to the total population of any country (87%, UN 2019). And yet Qatar retains among the highest GDP per capita ($66,000) rates in the world. Blessed with a robust health care system, nevertheless Qataris face high rates of chronic health problems known to be risk factors for the disease. Finally, the emirate is preparing to host the FIFA World Cup in 2022. So it is no surprise that the Qataris responded with strict measures to tackle the pandemic and bring it under control. These controls impacted all residents, from the minority population of wealthy Qatari citizens to the diverse array of migrant workers and their families. Among the migrants there is a range from higher-income workers living in households, high rises and gated communities to lower-income construction and service workers living in labor accommodations and working under conditions that make social distancing difficult. The study sought to better understand how and to what degree these different populations experienced the impact of the pandemic. It assesses which factors most shape this variation with a focus on the migrant labor population which hail from many countries. Drawing on a nationally representative telephone survey of 2,653 respondents conducted in nine languages between April and June 2021, we provide a clear picture of the differential impact of the pandemic across Qatari nationals, higher-income and lower-income migrant workers. This survey is also among the first to provide reliable information on the differential impact of the pandemic across groups by nationality, income, gender, and education.

COVID Changed Our Minds About Inequality: The Pandemic and Preferences for the Free Market vs the Welfare State

MDR Evans, *University of Nevada, Reno*
Jonathan Kelley, *International Survey Center*
SMC Kelley, *International Survey Center*

European publics' preferences have long endorsed a benign and collectivist welfare state with extensive government spending on health, welfare, and old age support; a generous social safety net, and an egalitarian income distribution. By contrast, Americans' attitudes tilt toward individualism and a market economy, with less government spending, a

less generous safety net, and a less egalitarian income distribution, but also higher wages and lower taxes. To dig deeper into Americans' ideals, we developed visualizations as abstract images of ideal-typical future worlds of welfare states and free markets and asked people about their preferences. Americans' pre-COVID preferences tilted towards a free-market future (41%), but with substantial minorities preferring a heavily welfare state future (31%), or liking both equally, 28%. But then, during the pandemic, the balance of opinion flipped: Preference for the free-market society dropped to 30% while more people came to endorse the welfare state, 39%, with the rest, 31%, liking both equally. At both times, the opinions of Americans spanned the full spectrum with is no polarization (no peaks at the extremes). Besides the descriptive statistics, our multivariate analyses explore the extent to which this shift reflects increasing poverty and other socioeconomic characteristics and the extent to which it is independent of them, reflecting instead a specific response to the pandemic. Data. Cross-cultural data are from our World Inequality Study (WIS) version 2, which pools surveys from several major international survey projects with large, representative national samples. There are 172 surveys in 45 nations (30 of which have the variables required here) with over 500,000 respondents. US data are from our International Social Science Survey (ISSS) Rounds 19, 20, and 21. In all there are 20 surveys, 12 of which include the questions analyzed here

Discrimination and Disparities in Mental Health Outcomes: Using a Population Health Survey to Capture Instances and Effects of Racial Discrimination

Todd Hughes, *UCLA Center for Health Policy Research*
Sean Tan, *UCLA Center for Health Policy Research*
Ninez Ponce, *UCLA Center for Health Policy Research*
Kathy Langdale, *SSRS*

In response to the urgent and critical need for accurate health data amid the COVID-19 pandemic, new data modules were included in the 2020 and 2021 California Health Interview Survey (CHIS), a population-representative state health survey. The broader visibility of racial injustice particularly among African Americans and increased frequency of hate crimes aimed at Asian Americans coupled with worsening racial health inequities due to the COVID-19 pandemic warranted the inclusion of discrimination-related questions in the modules added in the 2020 and 2021 CHIS. Variations of questions seeking to determine instances of discrimination have been sporadically included in the CHIS or asked of a random sample of CHIS respondents since 2003. Along with measuring the health and socioeconomic effects of COVID-19, the new modules capture instances of discrimination such as perceived unfair treatment due to race/ethnicity as well as personal experiences of a hate incident among Asian Americans or Native Hawaiian and Pacific Islanders (NHPIs). This paper will present and discuss the inclusion and results of the new discrimination modules in the CHIS, and compare key mental health outcomes for those who reported experiencing unfair treatment due to their race or ethnicity.

Design & Rewind: Evaluating Costs & Quality
Proxy Survey Cost Indicators: Are they Actually Correlated with Costs?

James Wagner, *University of Michigan*
Lena Centeno, *Westat*
Rick Dulaney, *Westat*
Brad Edwards, *Westat*
Tuba Suzer-Gurtekin, *University of Michigan*
Stephanie Coffey, *US Census Bureau*

Survey design decisions are – by their very nature – tradeoffs between costs and errors. However, measuring costs is often difficult. Further, surveys are growing more complex. Many surveys now include multiple modes or responsive and adaptive design elements that require cost information be available in order to make decisions during data collection. These complexities create new challenges for monitoring and understanding survey costs. Often, survey cost information lags behind reporting of paradata. Expenses may need to be routed through financial systems and may not show up in project accounting only after a significant amount of time has passed. Further, in some situations, the measurement of costs at the case level is difficult. For example, field interviewers may report hours worked each day, but do not report hours worked on each case. Given the time lag in reporting cost information and the difficulty of

assigning costs directly to cases, survey designers and managers have frequently turned to proxy indicators for cost. These proxy measures are often based upon level-of-effort paradata available for specific data collection designs and survey modes and which may also vary across organizations. Unfortunately, little is known about how accurately these proxy indicators actually mirror the true costs of the survey. In this presentation, we examine a set of these proxy indicators across several surveys with different designs, including different modes of interview. We examine the strength of correlation between these indicators and the total project cost and a related, but different category of costs, total interviewer hours. This provides some initial evidence about the quality of these proxies as surrogates for the true costs using data from several different surveys with different modes (telephone, face-to-face) across three organizations (University of Michigan's Survey Research Center, Westat, US Census Bureau).

Is it Worth It? Cost-Effectiveness and Quality Implications of a Final Mailing to Survey Nonrespondents
Joe Murphy, *RTI International*
Saki Kinney, *RTI International*
D. Ross Beall, *U.S. Energy Information Administration*
Katie Lewis, *U.S. Energy Information Administration*

Decisions on when to stop pursuing survey completes can have implications on data quality, representativeness, response rates, schedule, and cost. In the context of a self-administered survey, is the final mailing to nonrespondents worth the time and resources? If so, is it worth it only for certain types of cases or areas? We examine data from the 2020 Residential Energy Consumption Survey (RECS) comparing responses received prior to the last of three mailings to responses received after the last mailing. The 2020 RECS, conducted by the U.S. Energy Information Administration, sampled over 50,000 households in all 50 US states and the District of Columbia. The survey contacted sampled households by mail, inviting them to participate in either a web or paper survey. The first questionnaire package was sent via first-class USPS mail. Consistent with best practices, the second and third questionnaire packages each used distinct packaging and special postage (UPS Mail Innovations and UPS second-day for the third), in an attempt to reach respondents who may have ignored earlier requests. Specifically, we examine rates of item-level nonresponse, differences in administration time, breakoffs, and edit failure rates between early and late responders. We compare levels of representativeness in the respondent sample before and after the final mailing using American Community Survey data. We also compare response rates and relative costs per complete. Using evidence from these comparisons, we identify areas and types of cases for which the last mailing was most and least effective. This enables planning for future data collections that may offer the final mailing only for a subset of cases to optimize quality and costs. We conclude with findings applicable to other surveys evaluating the cost-effectiveness of their mailing strategies.

I'm Just Here for the Popcorn: An Experiment Using a Marketing Forward Mail Strategy in a Push-to-Web Approach
Stephanie Melton, *Nielsen*
Hani Zainulbhai, *Nielsen*
Randy Brooks, *Nielsen*
Megan Walsh, *Nielsen*

For any push-to-web strategy, mailed materials are critical to getting participants to your web survey. The look, feel, and messaging all play a role in whether someone takes the next step. Nielsen currently employs a push-to-web strategy to recruit homes into its TV panel. The materials mailed to the sample addresses adhere to standard best practices in the survey literature: multiple sequential mailings, cash incentives and a professional, research-oriented tone. However, decreasing response rates across the survey industry combined with dwindling brand recognition of Nielsen prompted us to rethink our approach. How do we get people to pay attention to our mailers and drive excitement to participate in our surveys - especially younger and hard to reach demographics? In this study, we explored a radical departure and evaluated the effectiveness of a "marketing forward" look and feel in effort to stand out and motivate participation. Specifically, this test evaluated our production materials with test materials with   bright imagery and a friendly tone, and a re-designed brand awareness piece: a bag of microwave popcorn in a box as a nod to Nielsen's research on TV and media. Our evaluation consisted of qualitative testing and a quantitative A/B test of response rates. For each group, the number of total pieces (a brand awareness piece, an invitation letter, a reminder letter) and the incentives provided were kept constant. Our preliminary results suggest this marketing forward approach may not be any more effective at survey log-ins compared to a traditional professional research tone. The

qualitative findings provide context on how friendlier and warmer messaging can resonate and where the marketing approach fails to pop. In this paper/presentation, we will share the results of the completed test and provide insights into the kinds of elements that resonate as part of a push-to-web strategy.

Too Much of a Good Thing? An Evaluation of Reminder Amount and Mode with Escalating Incentives

Jill Carle, *Westat*
Mike Brick, *Westat*
Michelle Amsbary, *Westat*

As response rates continue to decline and survey costs increase (Dillman 2017), survey researchers have produced a broad body of literature seeking to optimize response rates in web surveys. Researchers generally agree that some elements increase response rates, including pre-survey notifications and requests from authority. While survey completion reminders generally increase response rates (Van Mol 2017), less is known about the impact of the number of reminders and reminder mode e.g. mail, email, on response rates. Through household- and individual-level contact data in the American National Election Studies (ANES) 2020 Time Series Study, we will test the impact of the number of reminders and reminder modes received by sampled participants on screener click-through and completion, and pre-election survey completion. We will evaluate the extent to which survey incentive escalation interacts with reminder amount and mode to maximize response rates for specific demographic and geographic areas.

How Good is Your Data? Web Survey Data Quality

Racial Equity in Private: CATI versus Online Mode Effects on Attitudes Towards Police Reform

Ethan Santangelo, *Emma White Research, LLC*
Chris Campbell, *Emma White Research*
Emma White, *Emma White Research*

The tendency for individuals to misreport their attitudes, behaviors, and traits in response to sensitive questions is well-documented (Tourangeau and Yan 2007), with this phenomenon shaped by the mode for which a survey is administered and the privacy it affords (Groves et al. 2011; Kreuter et al. 2008). Racial attitudes are one area of public opinion where scholars have long explored the potential effects of desirable response editing, or social desirability bias. The literature has found that attitudes on public policy to enforce racial equality and promote equity are vulnerable to answer manipulation (Krysan and Couper 2003; Berinsky 1999; Krysan 1998). However, survey context is crucial for understanding the nature of question sensitivity and any resulting bias (Hopkins 2009; Johnson and van de Vijver 2003). As Moberg et al. (2019) note in their discussion of trends in American public opinion on racial attitudes and related public policy, well-known instruments such as the GSS and ANES were created in the twentieth century for a different context. Our understanding of social desirability bias in contemporary public policy attitudes that evoke racial norms, such as police reform, needs to be updated. How might social norms for supporting law enforcement and survey mode shape responses to police reform, trust, and favorability? Are participants sticking to locally well-accepted, socially desirable views of police when prompted by an interviewer? To explore these questions, we use data from a 2021 mayoral survey in the predominantly white Michigan suburb of Taylor (n=335) where respondents participated in either a telephone interview or online survey. Building upon the well-established literature on social desirability bias, we explore if and to what degree police reform is susceptible to mode-based response effects in one American context.

Do Response Effects Change Over Time? Experimental Results From Six Waves of a German Online Panel Survey

Fabienne Kraemer, *GESIS - Leibniz-Institute for the Social Sciences*
Henning Silber, *GESIS - Leibniz Institute for the Social Sciences*
Bella Struminskaya, *Utrecht University*
Bernd Weiß, *GESIS - Leibniz Institute for the Social Sciences*
Michael Bosnjak, *RKI - Robert Koch Institute*
Joanna Koßmann, *ZPID - Leibniz Institute for Psychology*

Satisficing response behavior (i.e., taking shortcuts in the response process) is a threat to data quality. Previous research (e.g., Schonlau & Toepoel, 2015; Sun et al., 2019) provides mixed evidence on whether satisficing increases

across time in panel studies, which would impair the quality of survey responses in later waves. Moreover, as previous studies were non-experimental, little is known about the mechanisms behind possible increases.   Against this background, our experimental study (data collection: 2020-2021) investigates satisficing in a panel study and its underlying mechanisms. Participants of a German non-probability online panel (n=1,397) were randomly assigned to two groups and either received target question design experiments in all six panel waves (high conditioning) or only in the last of the six waves (control condition). This design allows us to study satisficing over time and examine its causes: either an increase in general survey experience (process learning – control group) or growing familiarity with specific questions (content learning – high conditioning group). The target question design experiments comprised six between-subject experiments, which manipulated (1) the response order of answer categories, (2) whether a question included a 'don't know' option, and (3) whether a question was displayed in the agree/disagree or the construct-specific response format. First results reveal the existence of all three forms of satisficing (i.e., primacy effects, acquiescence, and saying 'don't know') in each of the six waves of the study. Overall, response effects are considerably large ranging up to 27.7 percentage points. However, there seems to be no clear pattern of increase or decrease in satisficing over time. In addition, there is no significant difference in satisficing between the high conditioning and the control group, which indicates that neither process learning, nor content learning leads to increases in satisficing over time and a decline in data quality in later waves.

If You Can't Say Something Nice, Don't Say Anything at All: Variations in Disclosing Undesirable Behaviors by Mode and Sample Type

Ki Park, *University of Northern Iowa, Center for Social & Behavioral Research*
Kyle Endres, *CSBR-UNI*
Eva Aizpurua, *Trinity College Dublin*
Erin Heiden, *CSBR-UNI*
Mary E. Losch, *CSBR-UNI*

Previous research suggests that mode of survey administration affects the disclosure of undesirable behaviors such as substance use. Thus, self-administrated surveys such as web or mailback questionnaires may report a higher incidence of undesirable behaviors compared to telephone mode. In this study, we examined the differences in the level of disclosure of undesirable behaviors (e.g. substance use and gambling) by mode of administration (web, mailback, and phone) and sample type (probability and non-probability). The probability-based data were collected using a probability sample of dual-frame random digit dial (DF-RDD) and an address-based sample (ABS) for a statewide general population survey. The survey was fielded in the fall of 2021 and yielded more than 3,000 completes across the modes (web, mail back, and phone). At the same time, a non-probability sample of adults in the state completed the same online questionnaire (n=700). The participants were asked about their attitudes and behaviors related to substance use, gambling, and health-related behaviors. We analyze multiple indicators of socially desirable responding, including responses to behavioral questions by mode and sample type. We hypothesize that respondents in self-administered surveys disclose more often socially undesirable information compared to telephone interviews. The implications of these findings and how they fit in with previous research will be discussed

Evaluating Consistency in Survey Device Type Choice Among Web Panelists in Self-Administered Surveys

Kyle Berta, *SSRS*
Chintan Turakhia, *SSRS*
James Noack, *SSRS*

With increasing ownership and usage of smartphones among Americans, the share of respondents who choose to take self-administered web surveys on mobile devices has matched, if not exceeded, the share of those who complete them on desktop computers. In this methodological brief we explore how consistent web panelists are in their choice of device type. We examine whether those who take surveys on mobile devices consistently take them on mobile devices, or if they switch back and forth between mobile devices and desktop computers. We explore the demographic differences among those who are consistently mobile, consistently desktop, and those who are a mix of both. For this analysis, we utilize data from the probability based SSRS Opinion Panel.

Welcome to the Machine: Integrating Automation Into Research

Nathan Wiggin, *COMENGAGE.US*

The market research industry is in a bind. Though survey data is incredibly important in customer-focused decision making, as a whole we are taking too long to get results. We cannot keep up with big data and the lag time is harming our ability to inform decision makers. Survey automation is not a new idea. We have all gotten the emails, text messages, or phone calls after taking our car to the dealer for service. What is new is: · You can develop automation without tens of thousands of dollars by integrating systems. · You can automate even complex studies including regression, discrete choice modeling (e.g. conjoint). · You can build tools that will 'see' the data for us and literally write the report for us. This presentation is a synthesis of nearly two years work with over a dozen clients. It covers our realization that automation is necessary and it can be done, even by an extremely small firm. It will give a brief overview of the process from start to finish and attendees will be able to experience the process from the eyes of the participants and the researchers It will touch on known quality issues in current research, especially in report writing where the dreaded 'fat finger' is all too common. And explore ways to greatly reduce errors by streamlining and automating the back end of survey research; data cleaning, analyzing and reporting. The presentation will conclude by providing ideas for when to and when not to use automation, problems with automation, such as identifying bad respondents, and specific things to consider that can greatly reduce the time needed between collecting and reporting the data

Shall You Choose, or Should We? Testing Parallel vs. Sequential Multimode Designs

Remember, You Can Complete This Survey Online! Adding a Web Survey URL to the Mail Questionnaire Cover in a Mixed-Mode General Population Survey

Amanda Ganshert, *University of Nebraska-Lincoln BOSR*
Lindsey Witt-Swanson, *University of Nebraska-Lincoln BOSR*
Kristen Olson, *University of Nebraska-Lincoln*

Push-to-web surveys are 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 rates among younger people by providing a web option (Smyth, Olson, & Millar, 2014), a group typically with low response rates in general population surveys, but allowing participation of other groups that may not have easy internet access or web skills through inclusion of a mail survey, including older adults and those without internet access. A common push-to-web methodology only offers the web survey URL in the cover letter. However, when asking respondents to respond via web, researchers should decrease burden as much as possible to increase response rates (Dillman et al., 2014). Thus, providing the web survey URL and login information on the paper questionnaire as well as the cover letter may increase the proportion of people who participate via the web. In this study, we report results of the 2021 Workforce Survey of the Greater Omaha Area (n=8,500 households sampled, data collection from November 2021 to January 2022), a concurrent mixed-mode mail survey of Nebraska adults. All sample members were sent a paper survey and a shortened URL to access the web survey in the cover letter in the first and reminder mailings. Half of the sample members were randomly assigned to receive either the questionnaire with the web link and login information on the questionnaire cover and half received the questionnaire without the web link and login information. We will examine response rates, selection rates of the two modes (mail and web), speed of response, and the composition of the respondent pool across the questionnaire cover conditions. We will conclude with recommendations for general population push-to-web methodology.

Using Cluster Analysis to Develop a Tailored Contacting Strategy

Jamie Wescott, *RTI International*
Michael Duprey, *RTI International*
Jerry Timbrook, *RTI International*

Given the ongoing trend of declining survey response rates, researchers are challenged to find new methods for increasing response. One option is to increase the data collection effort applied to each case (e.g., additional reminders/contacts); however, this effort is constrained by project resources and limits on the total number of contacts for each sample member. Therefore, it is essential that researchers maximize the efficacy of each contact attempt. Contacting sample members using different modes (e.g., telephone, mail) is common practice, but all contact modes are not equally likely to result in response for all sample members (de Leeuw 2005). Tailoring the mode and timing of a

contact based on sample members' preferences can increase the effectiveness of contacting efforts. In this paper, we use machine learning to develop a tailored strategy for contacting sample members and prompting them to complete a survey. Our approach uses data from a nationally representative mixed-mode (telephone and web) study of approximately 170,000 individuals conducted from March 2020 to January 2021. Sample members were contacted by a mix of mail, email, telephone, and text messages at multiple points during the field period. K-prototypes and gradient boosting are employed to cluster sample members using their characteristics (e.g., education level, sex from administrative data) and response behavior (e.g., mode and timing of response) to suggest the most successful contacting strategy for each characteristics class. We use this framework to tailor the contacting strategy for a new follow-up study of the same cohort that will be fielded in early 2022. We offer sample members in this new study the contact mode and timing that was most successful for similar sample members in the base study. Our presentation will describe the methodology used to conduct the clustering, the tailored contacting strategy implemented in the follow-up study, and preliminary results.

Has Anything Changed in a Decade? An Experimental Evaluation of Mail/Web Mixed-Mode General Population Surveys

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

Mixed-mode web and mail surveys are commonly used to survey the general adult population. Although early research indicated that concurrent mixed-mode designs (offering both web and mail simultaneously) decreased response rates compared to surveys that offered mail only (Medway and Fulton 2012), recent reviews indicate that the difference in response rates for these designs may be shrinking (Olson, et al. 2021). Furthermore, many surveys are under pressure to turn to web-only designs (Wells, et al. 2019), which may attract individuals who prefer to participate by web, obtain data quickly, and reduce data entry costs, but risk missing important subgroups and reducing response rates. Surprisingly few recent mode comparisons simultaneously evaluate single mode (mail-only and web-only) versus concurrent and sequential mixed-mode designs. In this paper we evaluate mail, web, and mixed-mode surveys of the general population to see if early findings continue to hold over decade later or if the mixed-mode landscape has changed. We do this using the Nebraska 2020 survey, a mixed-mode web and mail survey conducted during Fall 2020. A simple random sample of 10,000 households were randomly assigned to one of four mode conditions (n=2500 in each condition): mail-only, web-only, concurrent mixed mode, and sequential mixed mode. We examine response rates, timing of completion, mode selection, sample composition, and key survey estimates across the experimental conditions. We also examine separately whether characteristics of the mail and web respondents vary depending on the mode conditions to which they are assigned. Initial analyses indicate that response rates were lowest in the web-only condition (RR1=22.1%) and were equivalent in the other three conditions (mail-only: 32.7%, concurrent: 33.4%, sequential: 31.6%, p=0.39). Implications for the design of self-administered general population surveys will be discussed.

Concurrent and Sequential Mixed-Mode Experiments Before and During the COVID Pandemic

Lindsey Witt-Swanson, *University of Nebraska-Lincoln*
Amanda Ganshert, *University of Nebraska-Lincoln, BOSR*
Kristen Olson, *University of Nebraska-Lincoln, BOSR*

Surveys increasingly use self-administered and mixed-mode approaches to collect data from the general adult population (Olson et al. 2019; Dillman et al. 2014). Yet sequential mixed mode designs are not uniform – some switch modes immediately after the first contact, and others wait until later contact attempts to maximize the number of respondents in the initial (and usually cheaper) mode. Much of the work on mode sequences combines mail and email contacts (Millar and Dillman 2011), but how the timing of sending a paper questionnaire affects response rates, uptake of different modes, the respondent pool, and costs in a general population sequential mixed-mode design where email addresses are not available is an open question. Furthermore, whether the COVID pandemic changed householder's willingness to participate or their mode choice is unknown. We address these questions using 2021 Nebraska Annual Social Indicators Survey (8,000 households sampled, data collection from August to November 2021) data. In the 2021 NASIS, sampled addresses were randomly assigned to one of six mode conditions: (1) a web-only condition, (2) a mail-only condition, (3) a web and mail concurrent mixed-mode condition, or one of three sequential mixed-mode conditions that varied whether the mail is offered at the (4) second, (5) third, or (6) both second and third follow-up

mailings. All conditions used a four-contact approach, including a survey invitation, postcard reminder, and two survey reminder mailings. We examine differences in response rates, selection of mail versus web modes, respondent characteristics, and costs between conditions. We compare the experimental conditions in the 2021 NASIS to data collected immediately before the COVID pandemic in the Winter 2020 NASIS (n=1,584; AAPOR RR2=26.4%; data collection from January to April 2020) in which almost identical experimental conditions were fielded, allowing a direct pre-post pandemic comparison. We conclude with implications for fielding self-administered surveys.

Pick Me! Incentivizing Web Response in a Concurrent Mixed-Mode Design
Rebecca Medway, *American Institutes for Research*
Mahi Megra, *American Institutes for Research*

Offering both web and paper modes of response is an attractive – and increasingly employed – design feature for maximizing response rates, minimizing bias, and controlling costs. However, there are challenges to employing such designs. While researchers often prefer web response, sample members may prefer paper (Shih & Fan 2008). Offering both modes concurrently may reduce the response rate (Medway & Fulton 2012) but offering them sequentially may prematurely dissuade sample members from responding who would have done so if they knew the second mode would be offered. Biemer et al. (2018) reported promising results of a "choice plus" protocol that addresses some of these challenges; it offers both response modes concurrently but incentivizes web response by offering a larger contingent incentive for web response than for paper response. This presentation reports the results of an experiment that builds on these findings. Households sampled for the 2019 National Household Education Survey (NHES:2019) were randomly assigned to: (1) choice plus, (2) web-push (sequential web-then-paper), or (3) paper-only. Within the choice plus condition, households were randomly assigned to receive a $10 or $20 contingent incentive for web response. Compared to web-push and paper-only, choice plus resulted in a higher response rate, both overall and among subgroups that typically have lower-than-average NHES response rates. Compared to web-push, choice plus also increased response to early mailings but did not increase the percentage of responses by web – the increase in response was almost entirely by paper. Among those in the choice plus condition, $20 led to more response to early mailings and more response by web than did $10. However, choice plus also was more expensive than the other conditions, particularly when the $20 incentive was used. This presentation will be of interest to practitioners interested in maximizing response rates in mixed-mode surveys.

Illustrations of Passive Data Collection
Would You Recommend Usability to a Friend: A Deep Dive Into NPS, UMUX-Lite, and Goal Completion
William Dusch, *IBM*
Mika Jugovich, *IBM*
Jon Temple, *IBM*
Claude Elle, *IBM*

Many corporations employ Net Promoter Score (NPS) as a customer experience metric, primarily as a measure of customer loyalty. With the practice of using intercept surveys, a small survey that appears during a users interaction with a website or system, to capture NPS becoming widely adopted there is the potential to use this data for other purposes, such as conducting user experience research. Within user experience research the UMUX-Lite Score, a metric derived from a two question survey on usability and ease-of-use, provides a measure of users perceived ease of using a system and has been shown in previous research to be highly correlated with NPS. Another important aspect of user experience research is determining if a user is able to complete their desired task or goal when interacting with the system. Goal completion rates can be measured utilizing click stream analysis, or through small intercept surveys. While asking for goal completion data on intercept surveys provides great information, it increases the respondent burden if you are also asking NPS, UMUX-Lite, or other questions. To what degree is UMUX-Lite and NPS related to goal completion, can NPS or UMUX-Lite be a proxy for them. To understand this relationship our team has analyzed three years of intercept survey data where NPS, UMUX-Lite, and goal completion were all assessed. This session will discuss how the three measures are related, focusing primarily on the relationship between NPS and UMUX-Lite with goal completion. We'll also discuss the predictive qualities of the component questions of the UMUX-Lite and show that UMUX-LITE is a significantly better predictor of goal completion than NPS.

Person-Generated Health Data From Fitbits and ePROs to Mitigate Biases in Artificial Intelligence-Enabled Precision Public Health

Ritika Chaturvedi, *University of Southern California*
Tania Gutsche, *Center for Economic and Social Research and Department of Economics University of Southern California*
Eva Guadalupe Ortega, *Center for Economic and Social Research and Department of Economics University of Southern California*

Racial-ethnic minorities and other underserved populations experience disproportionate adverse health outcomes despite decades of research correlating social factors to health disparities. Social and structural determinants of health (SDoH)—including economic/educational inequalities, healthcare access, systemic racism, and lifetime stress—account for ~80% of modifiable risk factors that contribute to disparities. Many public health approaches use population averages to create "one-size-fits-all" interventions, but are limited by individual heterogeneity in number, magnitude, interplay, and amplification of SDoH. Person-generated health data (PGHD) from consumer digital technologies are invaluable tools for artificial intelligence-enabled "precision public health," a field that aims to develop digital interventions targeting unique needs of specific populations. PGHD can continuously measure everyday lived experiences outside of intermittent clinical settings. However, PGHD are typically captured from convenience samples through "bring-your-own-device" designs, and are vulnerable to input biases that have systematically underrepresented vulnerable populations. Here, we describe the creation of a first-in-kind holistic and sociodemographically representative sample of Fitbit and health data collected continuously from 1000 individuals called American Life in Realtime (ALiR). Participants were recruited from existing members of a nationally-representative survey panel of American adults from the USC's Understanding America Study and were provided with a Fitbit and a custom smartphone app. Vulnerable and underserved individuals from the panel were oversampled, including a majority sample of individuals without a bachelors' degree, census-matched samples of non-white, hispanic, and older adults (age>60), and a large sample of non-white men, a substantially underrepresented population in PGHD research. Participation was incentivized using a gamified approach leading to overall response rates of 87%. ALiR establishes a generalizable research infrastructure to use PGHD to explore the population-specific influence of SDoH in virtually any population and on any health outcome, to mitigate the influence of widespread methodological biases that result in systemic underrepresentation of the most vulnerable.

Is Tracking All That It Takes? An Exploration of the Validity of Metered Data Measurements

Oriol Bosch, *The London School of Economics and Political Science*
Melanie Revilla, *Research and Expertise Centre for Survey Methodology (RECSM), Universitat Pompeu Fabra*

Metered data, also called "web-tracking data", is generally collected from a sample of participants who willingly install or configure, into their devices, technologies that track digital traces left when people go online (e.g. URLs visited). Since metered data allows observation of web browsing unobtrusively and without relying on self-reports, it has been proposed as a potentially useful tool to measure online behaviours. As survey data, metered data is normally used to measure specific constructs, and then make inferences for a target population. To operationalize these constructs into measurements, researchers must specify the pieces of information from participants' tracked online behaviors that will be used. For instance, to measure the "number of visits to political articles", a list of visits to URLs considered as containing "political articles" must be defined and added up. While previous research shows that most survey design decisions can affect the validity of survey questions, little is known for metered data. Using a three-wave online survey conducted in opt-in panels in 2021 in Spain, Portugal and Italy, matched at the individual level with metered data , we first discuss the design decision that should be considered when creating metered data measures (e.g. how much time should a visit take to be considered?). Second, by operationalizing a set of specific constructs (e.g. time spent visiting political articles) using different combinations of design decisions, we estimate the convergent and discriminant validity of an extensive range of measurements (up to 432 per concept). Though the application of random forests of regression trees, finally, we estimate the influence of the different design choices on the validity of measurements. Our results can help researchers deciding what to consider when operationalizing constructs into metered data measurements, by shedding light on the mechanisms in which different design decisions can affect their validity.

Using Google Searches & Tweets for Social Research

A Review of the Use of Google Trends Data in Survey and Public Opinion Research

Florian Keusch, *University of Mannheim*
Johanna Mehltretter, *University of Mannheim*

Aggregated Internet search data, for example, from Google Trends, are increasingly used as a supplement or alternative to survey data. Particularly for sensitive topics, where surveys suffer from social desirability bias, the expectation is that the anonymous search queries of a large number of Internet users are a good reflection of true behaviors and attitudes. In addition, Google Trends data allow researchers to study changes in attitudes and behaviors across time and geographic areas at much finer granularity than possible in surveys. Besides problems of representativity and technical limitations of the data, the use of Google Trends data as an indicator of public opinion raises concerns of validity. It is often difficult to determine the original intent behind a search. For example, if someone uses an anti-immigrant term in their Internet search is this because of a true anti-immigrant attitude or for some other reason (e.g., trying to understand what the term means)? This study summarizes and structures the existing research on the use of Google Trends data in survey and public opinion research. We conduct a systematic literature review of published research that has used Google Trends data. Studies are classified based on (1) the substantive area and type of research questions Google Trends has been used for so far, (2) how researchers have identified keywords for search terms, and (3) what methods have been used to validate the results of Google Trends data. Our review identifies the promises and shortcomings of the use of Google Trends data and where more methodological research is needed to better understand under what conditions survey and public opinion researchers can use these data.

New Thinking About When Social Media and Survey Responses May Align

Frederick Conrad, *University of Michigan*
Michael F. Schober, *The New School*
Johann A. Gagnon-Bartsch, *University of Michigan*
Robyn A. Ferg, *Westat*
Mao Li, *University of Michigan*
Paul C. Beatty, *US Census Bureau*

The benefits of tracking public opinion by analyzing social media posts are potentially great, especially if such measures might sometimes stand in for certain survey measures. An important step in establishing the plausibility of this approach is demonstrating that patterns of survey responses and social media posts change over time in much the same way, i.e., their movement is aligned . Some early promising findings, e.g., a correspondence between consumers' responses to survey questions about their economic expectations and the sentiment of tweets containing the word "jobs" (O'Connor, et al., 2010), failed to replicate more recently (e.g., Conrad et al., 2019; Pasek et al., 2018) despite substantial effort to extend the pattern. In this talk we present our recent theorizing about when alignment is more likely than not: (1) survey responses vary enough so there is something to align with, i.e., signal-to-noise ratio is high; (2) the two time-series are numerically comparable, e.g., neither is cumulative; (3) the social media corpus includes content semantically related to particular survey questions; (4) the social media metric makes sense for the survey question, e.g., volume of posts when the question concerns awareness of a topic but sentiment of posts when it concerns the positivity of respondents' opinions. To test these proposals, we assess alignment between responses (n=76,919) to each of the 24 questions in the US Census Tracking Survey from January to September 2020 and a corpus of tweets (n=3,499,628) from the same time-period containing text relevant to the 2020 US decennial census. We observe alignment (comovement between the two time-series) for more than 1/3 of the questions and are largely accurate in predicting when comovement will be observed. These results point to a way forward in understanding when and how social media can credibly be used to measure public opinion and thinking.

Efficient and Reliable Geocoding of Twitter Data to Enable Spatial Data Linkage with Official Statistics and Other Data Sources

H. Long Nguyen, *Universitaet Bielefeld*

Dorian Tsolak, *Bielefeld University*
Anna Karmann, *Bielefeld University*
Stefan Knauff, *Bielefeld University*
Simon Kühne, *Bielefeld University*

Ready-to-use spatial information in the form of GPS coordinates is only available for a tiny fraction of Twitter data, making it difficult to link them to external geographic information, e.g., for the study of regional differences in public opinion. We address this problem by using the free text locations provided by users in their Twitter profiles to determine the corresponding geolocations. Since users can enter any text as profile location, automated identification of real-world locations based on user-provided text strings is highly complicated. Using two relational databases — one housing a growing corpus of over 1 billion German tweets, the other used for geocoding location text strings — hosted on an in-house high-performance computing server, we were able to assign real locations to the tweets sent by nearly half of the users in our corpus. This represents a vast improvement over the 0.2 percent of tweets in our corpus to which Twitter assigns geographic coordinates. Based on the geocoding results, we are not only able to determine a corresponding place for users with valid profile locations, but also the administrative level to which the place belongs. Enriching Twitter data with this information ensures that they can be directly linked to external data sources at different levels of aggregation, including aggregated regional statistics and survey data. To evaluate the performance of our method, we compare the spatial distribution of our geolocated tweets with the distributions of tweets geotagged by Twitter and the general population in Germany. Our procedure provides a solution to a common problem in using big online data for public opinion research and is made open-source for reuse by the scientific community.

Google This then Tweet That! An Exploration of How Google Search Trend and Twitter Data Correlate Over Time with each other and with COVID-19 Disease Incidence and Mortality

Trent Buskirk, *University of Massachusetts - Boston*
Youzhi Yu, *Bowling Green State University*

Rises in survey costs and declining response rates combined with increases in the number and types of access methods for alternate big data sources have allowed survey researchers a wider pallet of information to draw upon for creating estimates of public opinion. Social Media Data like Twitter and Digital Trace Data like Google Search information have emerged as two commonly accessed and used sources of alternate data, in part because of their cost and ease of acquisition. In this research we collected COVID related search keyword incidence rates from 9 different keyword-categories over a 30-week time period using both the Google Health API and from the Twitter firehose API via a third-party vendor. Data were separately accessed from 8 Metropolitan Statistical Areas (MSAs). We also collected COVID-19 disease incidence and mortality from the repository maintained by Johns Hopkins University. In this paper we will evaluate the relationship between the Google Health Trend incidence data with Twitter keyword-category incidence over time and across the 8 MSA's. We will also compare how these social media and digital trace data can predict COVID-19 disease incidence and mortality per 100,000 residents. Specifically, we investigate the partial auto-correlation between the respective Google and Twitter time series data as well as change-points in these series as compared to those of the COVID-19 data. Our preliminary work suggests that a definitive lag exists between the Google and Twitter data, consistent across nearly all MSAs and keyword-categories whereby Google activity in a prior week is more highly correlated with Twitter activity in the week directly following. This finding could have implications for how frequently each type of data may need to be accessed in surveillance activities. Additional relationships between these respective time series and the actual COVID-19 outcomes will also be presented.

Small Area Estimation of State-Level Twitter Populations

James McKinstry, *SSRS/Social Science Research Solutions*
Mickey Jackson, *SSRS*

Data scraped from social media sites, such as Twitter, holds promise for providing a richer understanding of trends in public opinion, supplementing information collected from traditional surveys. But to draw inferences from scraped social media data, we need to understand how the people on these sites differ from the general population. For example, prior survey research has found that Twitter users are generally younger, higher-income, and more likely to hold a college degree. Another important dimension is geography—are residents of certain states or regions

disproportionately likely to be on Twitter? Because geography in the United States is correlated with a wide range of political and social attitudes, understanding geographic variation in Twitter usage is important for understanding the types of voices that may be over- or underrepresented in data collected from Twitter. Survey data can be used to estimate the prevalence and characteristics of Twitter users at the state level. However, the standard approach to producing population estimates from probability-based surveys, direct estimation using weights reflecting each respondent's probability of selection for the survey, is often poorly suited for producing estimates within small geographic areas, such as states. This is because, even with a large national sample, many states will have small sample sizes. For all but the largest states, direct estimates of state-level Twitter use would be very imprecise, limiting the ability of researchers to draw conclusions about between-state differences in Twitter usage. In this paper, we apply a model-based small area estimation approach, Multilevel Regression and Poststratification, to a nationally representative survey to estimate the prevalence and characteristics of Twitter users at the state level. Researchers may find these estimates useful as benchmarks when analyzing Twitter data for state-level public opinion research, or as explanatory variables in analyzing cross-state variation in other attitudes and behaviors.

Understanding Nonresponse Bias in a Population-Based Surveillance System

Measuring Nonresponse Bias in the Pregnancy Risk Assessment Monitoring System (PRAMS), 2019

Holly Shulman, *CDC*
Lee Warner, *CDC*
Philip Hastings, *Far Harbor*
Ruben A. Smith, *CDC*

As response rates for health surveys decline, it is important to assess the impact of declining response rates on the risk of bias. The availability of birth certificate data for the PRAMS sampling frame provides a unique opportunity to examine nonresponse bias in a population-based survey. PRAMS is a jurisdiction-specific, population-based survey established in 1987 of women who have recently delivered a live born infant. We examined 2019 birth certificate indicators from 47 of 50 (94%) PRAMS sites. Response rates ranged from 40-80% allowing assessment of any impact of response rate on bias. The birth certificate includes indicators of health behaviors, medical conditions, and socio-economic variables enabling examination of the impact of topic sensitivity on bias.   We compared the prevalence of indicators from the birth certificate for the full sampling frame with the corresponding sample estimates obtained by applying the PRAMS weights to respondents. The difference is an estimate of true bias. We examine health behaviors (breastfeeding in hospital, smoking before and during pregnancy, adequacy of prenatal care), medical variables (delivery method, gestational diabetes, infertility treatment, gestational hypertension, Body Mass Index (BMI), number of previous live births) and socio-economic variables (delivery payment source, WIC participant).   Mean absolute differences were computed for each indicator within each site to assess magnitude of bias. Instances where the full frame prevalence fell outside the 95% confidence interval for the sample estimate were identified as evidence of significant bias. Overall, there was little evidence of significant bias for the medical and socio-economic indicators examined. Two health behavior variables, breastfeeding and prenatal care adequacy, exhibited the most evidence of bias with 27% and 13% of sites exhibiting significant bias, respectively.   The direction of bias tended to overestimate positive behaviors and underestimate risky behaviors. We found little association between bias or topic sensitivity and response rates.

Understanding Nonresponse Bias in a Population-Based Surveillance System

Holly Shulman, *CDC*
Joseph Pirozzolo, *Far Harbor*
Lee Warner, *CDC*
Philip Hastings, *Far Harbor*

Declining survey response rates increase risk of nonresponse bias and highlight the importance of quantifying bias. We utilize data from the 2019 Pregnancy Risk Assessment Monitoring System (PRAMS), a jurisdictional-population-based survey conducted since 1987 that measures health-related behaviors and experiences among recently postpartum women. Birth certificate records serve as the sampling frame among 46 participating PRAMS jurisdictions, providing rich auxiliary data for modeling across both respondents and nonrespondents. The availability of frame data offers a unique opportunity to examine nonresponse bias in ways not typically available for population-based surveys. This session

highlights several approaches to assess bias in PRAMS. The first presentation will provide an overview of PRAMS methodology, response rate trends for PRAMS over time, and the protocol established by CDC for data release. This sets the context for the three analyses to follow, each of which presents a different approach to examining bias. The first analysis, a true bias analysis, compares population prevalence of health behaviors, medical conditions, and socioeconomic variables available on the birth certificate with estimates using the survey weights, enabling examination of the impact of topic sensitivity and response rates on bias. The second analysis uses a "level of effort" approach to compare survey estimates between easy-to-reach and hard-to-reach respondents. Response subgroups are created simulating 10% increments in response to assess incremental bias and validate assumptions about the similarity between late responders nonrespondents. The third analysis uses a model-based approach to conduct a sensitivity analysis. By varying assumptions about the nonresponse mechanism (ranging from fully random to fully non-random nonresponse) and simulating plausible values of the outcomes, boundaries are established for likely nonresponse bias. These findings strengthen our understanding of nonresponse bias in general and carry important implications for survey research on reproductive-health-related topics.

Level of Effort Approach to Assessing Incremental Nonresponse Bias in the Pregnancy Risk Assessment Monitoring System (PRAMS)

Holly Shulman, *CDC*
Philip Hastings, *Far Harbor*
Lee Warner, *CDC*
Ruben A. Smith, *CDC*

Level of effort approaches assess evidence of nonresponse bias by comparing survey estimates between easy-to-reach and hard-to-reach respondents. Many surveys incorporate extra efforts to increase response among hard-to-reach respondents in the hopes of minimizing nonresponse bias. We examined data from the 2019 Pregnancy Risk Assessment Monitoring System (PRAMS), a jurisdiction-specific population-based survey established in 1987 of women who have recently delivered a live born infant. PRAMS utilizes a mixed mode methodology (mail and telephone) for survey data collection. Using data from 20 PRAMS sites, we created response rate groupings simulating response rates of 10%, 20%, etc. up to the full response rate for each site (range 40% to 80%). Each survey respondent was ordered based on their survey completion date. The earliest 10% by each mode (mail and telephone) comprise the 10% grouping, the earliest 20% comprise the 20% group, and so on. All simulated response rate groups were reweighted as if they were the only respondents. The weighted estimates for each response group are compared to one another to assess incremental and overall bias. Twenty-one key survey indicators were examined covering sensitive and non-sensitive topics related to health behaviors and health conditions. Simulated estimates were determined to exhibit bias if they fell outside the 95% confidence interval for the full response survey estimate. Most indicators showed little evidence of bias, with current breastfeeding having the most instances of bias across the 20 sites. Incremental bias for each 10% decline in response rates ranged from about 1 percentage point for the current breastfeeding indicator to 0.3 percentage points for gestational diabetes indicator; most indicators exhibited incremental bias of about 0.5 percentage points per 10% decline in response rate. These results indicate the weighting adjustments for nonresponse are successfully accounting for variations in nonresponse.

Nonresponse Bias Sensitivity Analysis for the Pregnancy Risk Assessment Monitoring System (PRAMS), 2019

Holly Shulman, *CDC*
Patrick Malone, *Far Harbor*
Jie Min, *Far Harbor*
Eric Booth, *Far Harbor*
Joseph Pirozzolo, *Far Harbor*
Holly Shulman, *CDC*
Lee Warner, *CDC*
Ruben A. Smith, CDC

Declining survey response rates increase risk of nonresponse bias and highlight the importance of quantifying bias, which can vary by specific topics in a survey. Adjustments for nonresponse typically assume that non-response is missing at random (MAR) given the observed set of covariates used for adjustment. However, this assumption cannot

be conclusively tested because survey items remain unobserved for nonrespondents. To the extent that nonrespondents differ from respondents on unobserved factors related to survey topics, bias adjustments under the MAR assumption are compromised. Our research goal is to assess how increasingly severe violations of the MAR assumption could impact population survey estimates, by leveraging recent advances in binary nonresponse analysis (Andridge & Little, 2020). We focus on two research questions: 1) How sensitive is estimated nonresponse bias to untestable assumptions about the response process? 2) Can we establish an "upper bound" of potential bias for survey items? We examine these questions across five levels of response rates and a diverse set of survey topics. We utilize data from the 2019 Pregnancy Risk Assessment Monitoring System (PRAMS), a jurisdictional-population-based survey conducted since 1987 that measures health-related behaviors and experiences among postpartum women. Birth certificate records serve as the sampling frame among 46 participating PRAMS sites, providing us with rich auxiliary data for modeling across both respondents and nonrespondents. By varying assumptions about the nonresponse mechanism (ranging from fully random to fully non-random nonresponse) and simulating plausible values of the items, we establish the boundaries for likely nonresponse bias and provide evidence that the 'worst case scenario' of nonresponse bias occurs when (1) the auxiliary variables are less predictive of survey items, and (2) nonresponse deviates substantially from a MAR mechanism. These findings deepen our understanding of nonresponse bias in general and carry important implications for survey research on reproductive-health-related topics.

Overview of the Pregnancy Risk Assessment Monitoring System (PRAMS)

Holly B. Shulman, *Centers for Disease Control and Prevention*
Philip Hastings, *Far Harbor*
Ruben A. Smith, *Centers for Disease Control and Prevention*

The Pregnancy Risk Assessment Monitoring System (PRAMS) is a jurisdiction-specific population-based survey established in 1987 funded by the Centers for Disease Control and Prevention and currently conducted in 46 states, 2 territories, New York City, and Washington DC. PRAMS samples women about 2-6 months after they have given birth to a live born infant. The survey covers maternal behaviors, attitudes, and experiences before, during, and shortly after pregnancy. PRAMS utilizes a mixed mode methodology (mail and telephone) for survey data collection. Birth certificate records serve as the sampling frame providing contact information and rich auxiliary data on the population. Participating sites follow a standard protocol but have the flexibility to tailor certain aspects of the survey (e.g., stratification plan, incentives and rewards, survey modules) to their needs. We discuss PRAMS methodology, response rate trends for PRAMS over time, and the protocols established by CDC for data release. For many years PRAMS enjoyed high response rates (>70%) making nonresponse bias assessment less necessary. However, as response rates have declined in surveys over the last 2 decades (PRAMS mean response rate dropped from 77% in 2001 to 57% in 2019), assessment of nonresponse bias has become necessary. We describe past nonresponse bias analyses and the motivation behind the new comprehensive analyses currently underway. These new analyses will inform future decisions regarding data release protocols and will be discussed in other presentations as part of a contributed session.

The Horse Race: We Can't Tell Who Wins, But Know Pollsters Lose
The "Reddening" of Montana: Durable Changes, Ephemera, and Challenges for Polling

Eric Raile, *Montana State University*
David C. W. Parker, *Montana State University*
Greta Linse, *Montana State University*

Underestimation of support for Republican candidates was common during the 2020 election cycle in the United States. Post mortems have suggested a number of reasons for this underestimation, including a subpopulation that distrusts polling and associated organizations. Well-designed polls for the state of Montana similarly suffered from underestimation. Montana, a state previously characterized as purple, moved strongly into the red camp during this election cycle as Republican candidates swept races for national and statewide offices. Underestimation of support for Republican candidates was upwards of 10 percentage points in some races, despite the accuracy of 2018 estimates with similar underlying methods. We examine and test hypotheses about the underestimation. Beyond nonresponse error, we also investigate potential last-minute attitudinal change. The cycle featured a hotly contested race for the U.S.

Senate that had the potential to flip partisan control of the body, and advertising within the state was extraordinary. We have some pre-election vs. post-election observations at the individual level to help with assessing this hypothesis. We also investigate the potential influence of rapid population growth in the state, as well as changes to voting procedures. In particular, we see some evidence that previously disengaged Republican voters may have voted as a result of being mailed a ballot. Voter turnout statewide increased from 74% to 81% from 2016 to 2020. Initial analysis suggests that increased turnout in Republican-voting counties accounted for up to half of the error in projecting races. However, nonresponse error also shows up as a significant factor, complicating efforts to design reasonable post hoc weights. We shows the results of various weighting assumptions and their consequences for weight sizes. Finally, we consider the implications of this work for predicting the shape of future electorates and corresponding sampling approaches.

Using Experimental Vignettes to Study how Survey Methods and Results Influence the Public's Evaluation of Public Opinion Polls

Allyson Holbrook, *University of Illinois Chicago*
Paul J. Lavrakas, *University of Illinois at Chicago and NORC*
Timothy P. Johnson, *University of Illinois at Chicago and NORC*
Young Ik Cho, *University of Wisconsin Milwaukee*
Andrew Crosby, *Northwestern University*
Xiaoyan Hu, *University of Illinois at Chicago*
Evgenia Kapousouz, *University of Illinois at Chicago*
Polina Polskaia, Pace University
Henning Silber, GESIS

Public opinion polls are a major form of documenting "the will of the people" in contemporary American society (e.g., Gallup and Rae, 1940). AAPOR has identified aspects of survey methodology that researchers are required to report so that surveys can be evaluated (e.g., sample size, response rate, and sampling strategy), but little research has examined how members of the public use information about a survey's methodology to evaluate it. Understanding the processes that lead people to believe, or not believe, the results of opinion polls has many important implications for the opinion research community and it is very important to learn more about how people form judgments about poll accuracy. We examine these processes by analyzing combined data from vignettes about survey results from three telephone surveys examining respondent evaluations of hypothetical polls about gun control (n=582), school vouchers (n=403), and the state lottery (n=730). Each of these studies deployed a randomized/experimental factorial design whereby different respondents were read different vignettes describing a "hypothetical" poll and its findings. This was done in a way that randomly varied different aspects of the poll related to their quality, including its sample size, response rate, mode of data collection, and sponsor as well as the findings of the survey. After a respondent heard the poll described, s/he was asked two questions about their assessment of it. The focus of our analyses is 1) to examine how survey characteristics associated with quality and the result of the survey influenced evaluations of the survey results in the combined samples as a whole, and 2) to examine potential moderators of their influence, including respondent education, the order in which information was presented to respondents, and the topic of the survey.

Uses of Transactional Data in Election Administration Research

David Varas Alonso, *Fors Marsh Group*
Kysha Gregorowicz, *Fors Marsh Group*

Research on election administration generally uses data from national sample surveys (e.g., CES, Census' CPS voting and registration supplement) and census-type aggregate data sets (e.g., Election Administration and Voting Survey). A third, less used option is transactional or administrative data provided directly by the election offices. Fors Marsh Group (FMG), partnered with government and local agencies has been working with these data to analyze and identify voter behavior associated with voting success, and potential obstacles in the voting process. Transactional data provide information about how, when and where all transactions between a voter and the election office happened from the time of registration to the moment the voter casts their ballot. While these data are generally difficult to obtain, the number of states and localities that make voter records available is growing and opens new avenues for election administration research. Analysis with transactional data has allowed FMG to identify communities in the District of Columbia with higher challenges to register to vote remotely and cast ballots by mail, which was particularly

problematic for the 2020 general election that was conducted by-mail in DC because of the impact of COVID-19. Transactional data from voters living overseas has allowed FMG to find voter behaviors associated with voter success when casting a ballot from abroad and measure the impact of policies in the voting process for this population. This paper discusses the potential of transactional data in election administration research as well as other research fields, shows examples of uses and findings using these data, and discusses limitations and recommendations when using transactional data in applied research

'Them' Without 'Us': Measuring Negative Partisanship in a Multi-Party System

João Areal, *University of Mannheim*

Partisanship, voters' attachment to a political party, has been at the heart of the study of public opinion, consistently predicting outcomes such as participation, vote choice, information-seeking patterns, and out-group hostility. A rich methodological and theoretical literature has guided both the conceptualisation and measurement of partisanship across contexts, providing researchers with much needed clarity in operationalizing the concept for analyses and data collection. In contrast, much less attention has been paid to the concept of negative partisanship, voters' rejection of a particular party. Recent evidence from both the United States and multi-party systems attests to the growing relevance of negative partisanship in influencing political behaviour, yet we lack robust and comparable measures of the concept. In this methodological brief I attempt to fill this gap by running different iterations of a novel Negative Identity Scale (NID) across a 3-wave online election panel in Germany (N = ~2,000), a country with an institutionalised and increasingly fragmented multi-party system. Drawing on insights from both psychology and political behaviour, I devise a multi-item scale that captures the relevant underlying dimensions of negative partisanship as a meaningful form of social identity. Results indicate that negative partisanship is prevalent in the German electorate, far surpassing the rates of positive partisanship. Further, scalability analyses indicate that negative partisanship can be reliably measured as a form of social identity by the proposed items, independently from positive partisanship. The findings provide new avenues for data collection and further research on the concept of negative partisanship in multi-party systems and will be of interest to researchers in public opinion and political behaviour more broadly.

Compositional Comparisons of IVR, SMS-to-web, and Online Panels in 2020 U.S. Pre-Election Polling

Spencer Kimball, *Emerson College*
Isabel Holloway, *Emerson College*

 The 2020 U.S. election cycle witnessed an expansion into the world of SMS-to-web polling. This study analyzes the compositional makeups of different modes of data collection in the 2020 U.S. Pre-election polling. Using unweighted data counts of 24 polls, n=19,886, 49% was collected from IVR polls (n=9,795) , 25% was collected from online panels (n=5,039), and 25% was collected from SMS-to-web surveys (n=5,052). Data was collected between September-November 2020. This study utilizes a series of chi-square to identify differences in the compositions of these modes against demographic variables.

Teaching Survey Research: Challenges & Opportunities

Teaching Survey Research: Challenges and Opportunities

Chase Harrison, *Harvard University*

Improving SOGI Measurement

Measuring Sexual Orientation and Gender Identity: Findings from the Health Information National Trends Survey (HINTS) 2020

Ashley Murray, *National Cancer Institute | National Institutes of Health*
Richard P. Moser, *National Cancer Institute*
Robin C. Vanderpool, *National Cancer Institute*
Andrew Caporaso, *Westat*
Gordon Willis, *National Cancer Institute*
Kelly D. Blake, *National Cancer Institute*

An estimated 11.5 million (4.5%) sexual and gender minority adults live in the U.S. Though an estimated 1.4 million U.S. adults are gender minorities, including both transgender and gender non-conforming persons, there are still few nationally representative health surveys that include items to measure gender identity. It is vital for researchers to accurately and precisely assess sexual orientation and gender identity separately in nationally representative health surveys to understand the unique health risks and health care-related needs of these vulnerable populations. The current study presents findings from the Health Information National Trends Survey (HINTS) 5, Cycle 4 (2020). HINTS is a nationally representative, cross-sectional postal survey of non-institutionalized U.S. adults 18 years and older sponsored by the National Cancer Institute (N=3865). Weighted and unweighted frequencies and proportions were calculated to assess the number and proportion of sexual and gender minority respondents in the sample, and protocols were developed to minimize disclosure risk upon public release of the data. Following recoding for accuracy, such as removing erroneous responses, 171 respondents (weighted estimate = 5.99%, SE = .70) identified as lesbian, gay, bisexual, or other as their sexual orientation. Using new measures assessing both sex assigned at birth and gender identity to assess gender minority status, 13 respondents (.37%, unweighted) were transgender or gender non-conforming. Overall, 176 respondents (weighted estimate = 6.16%, SE = .71) were sexual or gender minority. Results indicated small, unweighted cell sizes of gender minority adults, similar to other nationally representative samples (e.g., CDC's Behavioral Risk Factor Surveillance Survey). Findings suggest there are implications for item development, understanding challenges with small sample sizes and potential disclosure risk, including the need to suppress certain data for public release, comparisons with other known studies in this area, and identifying areas for future research.

Improving Inclusivity When Asking Gender In Surveys: Evidence From Four Experiments

E.D. Bello-Pardo, *Yougov*
William Bannick, *YouGov*
Monika Nayak, *YouGov*

Binary gender questions in surveys are fundamentally flawed. While many institutions accept binary gender questions, those that only offer Male and Female or Man and Woman as response options, as standard, they breed data issues and subject many respondents to an uncomfortable experience. These ubiquitous questions conflate and confuse sex and gender and neglect to accurately represent non-cisgender survey respondents (including those who are transgender and non-binary). As a result, a survey respondent whose gender is not listed is often forced to decide between providing inaccurate data or exiting the survey. Such choices produce problematic patterns. They affect the quality of survey data and perpetuate damaging assumptions about gender identity. Some hesitant researchers voice concerns about how using an inclusive gender question could present new challenges in weighting (Kennedy et al pre-print, 2020) or lead to differential attrition and priming effects in conservative respondents. We address these concerns using data from four experiments conducted between 2019 and 2021. In these experiments, we tested a four-category gender question with the options "Woman", "Man", "Non-binary", and an open ended "Other". We also tested the effects of following that item with an optional transgender identity question. Our analysis reveals that using this inclusive gender measure does not result in the hypothesized weighting issues, differential attrition, or priming of conservative respondents. In fact, our approach increases inclusivity, provides a better experience to non-cisgender respondents, and generates more accurate gender data. Therefore, we recommend that survey researchers reject binary gender questions and adopt inclusive, non-binary options when asking for gender in surveys.

Examining Measurement Error in a Sexual Identity Question

Valerie Ryan, *National Center for Health Statistics*

Over ten years ago, researchers working in the Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER) at the National Center for Health Statistics used cognitive interview findings to develop a sexual identity question for the National Health Interview Survey (NHIS) that would reduce measurement error. The question asks, "Which of the following best represents how you think of yourself?" with the following response options: "Lesbian or Gay", "Straight, that is, not lesbian or gay", "Bisexual", "Something else", and "I don't know the answer". This presentation will explore how this question has performed by examining changes in response category frequencies over time and comparing the amount of missing data between the NHIS question and another sexual identity question. I use data collected between 2013 and 2020 from the NHIS to examine question performance over time. To evaluate missing data, I use data collected between 2015 and 2019 on the National Survey of Family Growth (NSFG), where a split sample

experiment was conducted to test the NHIS question against an older NSFG sexual identity question. Question performance on both surveys is also compared between demographic subgroups. The NHIS and NSFG are general population surveys conducted in the United States. Preliminary findings indicate an increase in respondents choosing "bisexual", "something else", and "I don't know the answer" over time on the NHIS, particularly among female respondents under the age of 30. For the NSFG, there were fewer missing responses for the NHIS question among respondents at each education level (less than high school, high school, and more than high school) and for those completing the interview in Spanish. The NHIS question exhibits less measurement error than the NSFG question, though response category frequencies are changing over time, particularly amongst younger respondents.

Are You Religious? What Does That Mean?

2020 Presidential Support among Old Testament and New Testament Christians and Pseudo-Christians

Paul Lavrakas, *Independent Consultant*

Scott Richards, *Reconnect Research*

Religious beliefs were highly correlated with whether American voters supported Donald Trump or Joe Biden in the 2020 Presidential election. Millions of self-professed Christians supported Donald Trump even though his words and behaviors often were at odds with many Christian ideals. This should not be a surprise as there is a great heterogeneity in beliefs as to what being a Christian means to those who self-reported themselves to be Christian. We conducted a national pre-election poll in Fall 2019 using a RICS-IVR sampling, recruiting, and data collection design (cf. Levine, Krotki, and Lavrakas 2019 POQ article). In the questionnaire, 853 respondents were asked a series of questions about the extent to which certain Christian precepts were supported/followed by "their best friend." Some of the precepts related to the concept of "love" (primarily a New Testament precept) and some were related to the concept of "vengeance" (primarily an Old Testament precept). We found that those whose best friend reportedly supported Trump scored higher on our Vengeance scale and those whose best friend reportedly supported Biden scored higher on our Love scale. We will discuss these findings and others in terms of how they demonstrate that the study of religion and politics needs to use more nuanced methods to better understand the complexities and apparent inconsistencies among various Christians as to which presidential candidate was supported in 2020, and why?

The 2021 General Social Survey Mode Shift and Religious Identification Among Protestant Christians

Eyob Moges, *NORC at the University Of Chicago*

Jodie Smylie, *NORC at University of Chicago*

Eyob Moges, *NORC at University of Chicago*

The General Social Survey (GSS) is a nationally representative survey that has historically been conducted face-to-face every two years to collect data on the attitudes and opinions of the general public of the United States. Beginning in 2020, the GSS was adapted from a face-to-face survey to a mixed-mode web and telephone survey in response to the COVID-19 pandemic. One impact of the mode shift was evidenced through differences in responses to questions about respondents' religious identification. On prior rounds of the GSS, interviews were primarily face-to-face, in which interviewers served as moderators, identifying and coding respondents' religious preferences into predetermined classifications through a set of mostly interviewer-directed survey prompts. However, in the 2021 GSS, web survey respondents did not have interviewers acting as mediators, and as such, were directed by a series of survey questions to self-identify their religious preferences. Respondents were initially asked to indicate their religious preference from broad categories including Protestant, Catholic, Orthodox, and several non-Christian religions. Respondents could also select "Other", after which they were prompted to answer additional questions to clarify and categorize their entry. In 2021, we saw an increase in the proportion of respondents selecting "Other" whose subsequent responses indicated they were a member of a Protestant denomination. In this presentation, we will explore differences between respondents that did not at first categorize themselves as Protestant versus other Christian respondents that initially identified themselves as such. We consider what respondent characteristics may correlate with respondents' unwillingness or inability to self-identify as Protestant and the language respondents use to locate themselves in the constellation of Christian identities.

Secularization Around the World

Tom Smith, *NORC at the University Of Chicago*

Benjamin Schapiro, *NORC at the University of Chicago*

Secularization theory posits that religion inevitably declines as societies modernize. This hypothesis is tested using data from the International Social Survey Program (ISSP) Religion studies of 1991, 1998, 2008, and 2018. In addition, to surveys by ISSP members, the 2018 study is supplemented by surveys in 14 nations in Africa and Asia supported by the Templeton Religion Trust. This paper examines how far secularization has advanced around the world. Unlike most studies of religion, it focuses on those at the secular or non-religious end of the spectrum. First, the size of the fully

secular share of the population is examined on 10 indicators of religious identifications, behaviors, and beliefs. For example, it looks at not identifying with a religion, never attending religious services, and not believing in God. Second, the rankings across the 10 indicators are combined to make an overall levels of secularism across countries. Third, reasons for the variation in secularism across countries are explored with particular focus on differences in the main religious traditions in each country. Finally, another method for assess the level of secularization is examined and compared to the 10-indicator method.

Abortion Attitudes in Crossnational Perspective: The Influence of Diversity in Religiosity

MDR Evans, *University of Nevada, Reno*
Jonathan Kelley, *International Survey Center*

Public opinion on abortion has once again skyrocketed up the international agenda with, for example, Poland and the EU central government locking horns over the issue. This analysis investigates cross-national and cross-temporal differences in individuals' attitudes towards abortion, in light of both contextual country-level influences (socioeconomic development, diversity in religiosity) and individuals' demographic characteristics, socioeconomic characteristics, and religious characteristics. Prior research has demonstrated large contextual effects of socioeconomic development, as well as substantial effects of several aspects of individuals' religion and, less clearly, effects of individuals' demographic and socioeconomic characteristics. Yet, all in, these explain little of the variance in attitudes towards abortion Societal diversity in religiosity is as important aspect of "cultural looseness". The consequences of societal diversity in religiosity (the centrality of religion to one's life)-- the social climate of encounters with other people who exhibit a variety of levels of religiosity-- for individuals' endorsement of abortion and other unconventional personal moral conduct have only recently begun to be explored. This paper shows that diversity in religiosity at the national level enhances individuals' tolerance of abortion and also other unconventional approaches to personal morality, net of an individual's own religiosity, net of the average levels of religiosity and socioeconomic development in the individual's society, and net of key individual-level controls. Data are from the pooled World Values Surveys/ European Values Surveys, 1981-2021. Analysis is by multilevel methods

Challenges & Methodological Innovations in International Surveys During the COVID-19 Pandemic

Doing Surveys in the Midst of a Pandemic: Lessons, Opportunities and Challenges from the Afrobarometer

Carolyn Logan, *Michigan State University*
Boniface Dulani, *Afrobarometer*

Afrobarometer, a network of research organizations in Africa, collects the opinions and experiences of ordinary African citizens on a range of governance, economic, and social issues in about 36 African countries. When the World Health Organization declared COVID-19 a global pandemic in March 2020, Afrobarometer was about halfway through its eighth round of surveys. As the number of COVID-19 cases across the continent began to grow rapidly, Afrobarometer responded by suspending fieldwork activities in April 2020. This period of fieldwork hiatus launched three major developments for Afrobarometer. First, the network developed a new set of survey questions on COVID-19. Secondly, Afrobarometer prepared a comprehensive set of strict protocols to guide resumption of fieldwork activities during the pandemic, including new guidelines for providing remote technical assistance to national research partners. Based on these guidelines, Afrobarometer was able to resume face-to-face fieldwork in October 2020 when the number of COVID-19 cases in most countries on the continent began to drop and governments relaxed some lockdown restrictions. Even so, the network still had to frequently adjust the fieldwork calendar to account for evolving COVID-19 domestic regulations and country case loads, and in one case had to cancel a survey altogether, but Round 8 fieldwork in other countries was completed in July 2021. Finally, the network also began to explore options for undertaking telephone surveys. This included developing new protocols for collecting telephone numbers from face-to-face survey respondents who are willing to be contacted in future surveys, as well as running a series of pilot telephone surveys using multiple sampling approaches. This paper documents how Afrobarometer responded and adapted to pandemic realities, and examines the lessons learned and opportunities gained from this experience, and the ways the pandemic has potentially reshaped the future of survey research on the African continent.

Challenges and Methodological Innovations in International Surveys During the COVID-19 Pandemic
Matthew "Clark" Letterman, *Pew Research Center*

Since March 2020, major multicultural, multilingual, and multinational (3MC) survey projects have had to adapt their research strategies in the face of the COVID-19 pandemic. This panel brings together representatives of some of the world's largest 3MC survey projects to discuss the challenges they've faced and the methodological innovations that they have implemented. The panel includes presenters representing the the European Social Survey (ESS), Arab Barometer, Afrobarometer and AmericasBarometer. The session will be moderated by Clark Letterman, a Senior Survey Manager on the Pew Research Center's International Research Methods team, who will also serve as the discussant. Topics include mode transitions from face-to-face to telephone surveys in multiple regions, transitioning from interviewer administered to self-completed surveys, the use of recontact panels recruited from prior face-to-face samples, designing and implementing health and safety protocols necessary to resume face-to-face data collection in the safest manner possible, coordinating across a diverse network of implementing partners, and more! With a diverse set of presenters, this panel provides a comparative perspective that incorporates lessons learned and future avenues for innovation in comparative 3MC projects across a wide range of survey environments and country specific contexts

New Approaches to Sampling for the 2021 AmericasBarometer
Noam Lupu, *Vanderbilt University - Nashville, TN*
J. Daniel Montalvo, *Vanderbilt University*
Elizabeth J. Zechmeister, *Vanderbilt University*

LAPOP Lab's sampling strategy for the AmericasBarometer, a region-wide public opinion survey, has changed significantly in 2021 due to the risks and obstacles presented by the COVID-19 pandemic. For the first time ever, the survey is being conducted via Computer-Assisted Telephone Interviewing (CATI) instead of traditional face-to-face interviews. This paper details how LAPOP has overcome various challenges associated with this transition by balancing state-of-the-art research principles with the practical limitations of surveying in the region. Informed by academic literature in survey methodology, consultations with local partners, and our own experience with several pilot studies, we outline best practices for sampling for telephone surveys in Latin America. For nationally representative samples, random digit dialing (RDD) is indispensable, as pre-packaged lists suffer from clear (and perhaps unknown) biases. We consider single frames (cell phones only) to be superior to dual frames (cell and landlines) due to the relatively high penetration of cell phones in Latin America and several complications associated with weighting dual sampling frames. Finally, the 2021 AmericasBarometer will, for the most part, not use any type of stratification, as it has in the past (for region), because the costs stemming from the technical hurdles to implementing them in the target countries outweigh any small gains in precision. We hope these guidelines will improve transparency about LAPOP's sampling strategy and serve as a model for others who conduct telephone surveys in Latin America.

Delivering the European Social Survey During COVID-19: Reflections and Future Implications
Tim Hanson, *City University London*
Professor Rory Fitzgerald, *City, University of London (European Social Survey)*
Niccolo Ghirelli, *City, University of London (European Social Survey)*
Siobhan O'Muircheartaigh, *City, University of London (European Social Survey)*

The European Social Survey (ESS) is an academically-driven cross-national survey that measures attitudes and behaviour across Europe. It has been conducted biennially since 2002 using face-to-face fieldwork and random probability sampling. The 10 th Round of ESS was due to start in September 2020, with more than 30 countries expected to participate. The COVID-19 pandemic led to numerous concerns about the feasibility of completing the round based on the usual face-to-face approach, which meant that contingency plans were needed. This included extending fieldwork timelines and allowing countries to carry out video interviews as a complementary method to in-person interviewing. However, it was clear that face-to-face interviewing would be unfeasible in some countries and an alternative data collection approach would be needed. This led to the development of a self-completion approach, which was adopted by several countries for their Round 10 fieldwork. This is the first time a self-completion approach has been used for ESS fieldwork. This paper reflects on the challenges and lessons from the experience of delivering ESS fieldwork during the pandemic. Particular attention is paid to the development of the self-completion approach,

including a summary of the development work and piloting to prepare for this and initial outcomes from the countries adopting this approach for their fieldwork. We consider the challenges arising from needing to quickly adapt an hour-long face-to-face survey to a self-completion format, assessing what worked well and what could be improved. There have been discussions regarding possible future mode switches for ESS for some time but the pandemic has heightened the focus on this debate and accelerated our thinking. We therefore also consider learning and implications from the experience of developing the self-completion approach during the pandemic regarding the future of ESS.

Attitudes Towards Vaccine Hesitancy & Uptake

Parents and the Pandemic: Parents' Vaccine Intentions And Concerns For Their Children

Lunna Lopes, *Public Policy Institute of California*
Liz Hamel, *Kaiser Family Foundation*
Grace Sparks, *Kaiser Family Foundation*
Mollyann Brodie, *Kaiser Family Foundation*

As COVID-19 vaccines have become available, first to adults, then to teenagers, and more recently to younger children ages 5-11, parents have not only had to make the personal medical decision about whether to get vaccinated, but have also faced the parenting decision of whether to vaccinate their children. Our analysis of a series of probability-based national public opinion surveys, including two with large oversamples of parents, looks at trends in vaccine uptake and intention, parents' concerns when it comes to vaccinating their children, and who they trust for information about the vaccines. We examine differences in parents' attitudes by age group of their children as well as by demographic characteristics like race and ethnicity, education, income, and partisan identification. Our surveys find that parents are more cautious when it comes to getting younger children vaccinated for COVID-19 and that even among vaccinated parents, notable shares express concerns about potential long term side effects the vaccine may have on children. This presentation will draw on the results of these 2021 surveys as well as forthcoming surveys to be conducted in early 2022 to summarize the current state of parental attitudes towards the COVID-19 vaccine and their vaccination intentions for their children in different age groups and explain how these have shifted over time.

Linking Intent and Behavior: Shifts in COVID-19 Vaccine Uptake Over Time in Detroit

Lydia Wileden, *University of Michigan*
Jeffrey Morenoff, *University of Michigan*
Susan Goold, *University of Michigan*
Elisabeth Gerber, *University of Michigan*
Caroline Egan, *University of Michigan*

Despite disparate rates of infection and mortality due to COVID-19, many communities of color reported lower intention to vaccinate and lower rates of vaccine receipt when vaccines became available in early 2021. This disparity in vaccination was often interpreted as evidence of vaccine resistance in Black and Latino communities, suggesting that those who had not yet vaccinated were opposed to ever vaccinating. In this paper, we draw on panel data from a representative study of Detroit residents (N=1,898) to highlight the shifting nature of vaccine intention and uptake over time. Our data shed light on the dynamic—rather than static—nature of individuals' intention to vaccinate and offer insights into who and why residents change their minds. Comparing vaccine intent and receipt between January and June of 2021, we find that a quarter (26%) of residents who said they were very unlikely to vaccinate against COVID in January had been vaccinated by June, while 14% of residents who reported they were very likely to vaccinate in January remained unvaccinated over that same period. Logistic regression models examining the relationship between respondents' assessments of their likelihood of getting vaccinated and their later reports of vaccine receipt show that the relationship between prior vaccine intent and later vaccine uptake varies significantly across demographic groups defined by race/ethnicity, age, education, and income. Preliminary analyses suggest that Black and Latino residents demonstrate some of the greatest willingness to change their minds on vaccination, in contrast to white populations. Additionally, lower-income residents are less likely to vaccinate despite their intentions. These results illustrate the importance of panel data in tracking intention and behavior related to public health, argue for an understanding of vaccine hesitancy as dynamic—especially among communities of color, and offer potential strategies for expanding public health vaccine campaigns.

Explaining Australia's COVID-19 Vaccine Success, and Ongoing Challenges, Using Survey and Linked Administrative Data
Nicholas Biddle, *Australian National University*

In late May 2021, as Canada, the United Kingdom and the USA all had exceeded 50 per cent first dose vaccination rates, only around one-in-ten Australians had received an approved vaccine. By mid-November, more than three-quarters of Australians had received at least one vaccination, exceeding the US and the UK, and converging on the Canadian vaccination rate. The aim of this paper is to use a combination of public opinion surveys, and linked administrative/survey data to explain this dramatic convergence, and to also help understand why there are still groups in Australia at the end of 2021 who have not been vaccinated. The paper will document the dramatic convergence in vaccination rates by race/ethnicity, but the widening gap by education and other socioeconomic measures. The paper will also show how trust in government and other institutions increased during the COVID-19 period in Australia, and how this explains in part the vaccination success, as well as ongoing disparities. The paper will conclude with implications for vaccination campaigns in other similar countries, as well as future vaccination campaigns in Australia.

Improving Survey Mail Materials to Optimize Self-Response

Improving Survey Mail Materials to Optimize Self-Response
Dorothy Barth, *U.S. Census Bureau*

Federal surveys often make first contact with units in sample through mail materials that provide survey information and instructions on how the recipients can respond. Survey sponsors often send multiple mailings to sampled units to encourage self-response. The design of the materials and the messaging in the materials affect survey response. In this panel, we will share the results of four controlled experiments that tested how changes to the mail contact materials in two prominent federal surveys, conducted by the U.S. Census Bureau, affected response.

Make Up Your Mind: The Emotional Turmoil of the Prenotice
Rachel Horwitz, *U.S. Census Bureau*
Sarah Heimel, *U.S. Census Bureau*
Beth Satisky, *U.S. Census Bureau*
Christine Bottini, *U.S. Census Bureau*

In the early 2000s, prenotice letters were commonplace. They informed sampled cases that a survey would be coming shortly, providing legitimacy, and increasing the chances the survey invitation would not be discarded. As web survey use increased, the prenotice letter seemed to be a waste of an opportunity – if someone was willing to do the survey right away, why not let them? Many studies were done that suggested final response rates were similar with and without the prenotice letter, but responses often came in faster when a prenotice was sent. We conducted one such study in 2017 – we excluded the prenotice from a survey with both a new and returning group of respondents and a 6-month data collection period. After finding there were no significant differences between the final response rates, the prenotice was removed from the mailing sequence. However, the 2019 survey cycle had depressed response rates, so the prenotice made a comeback in 2021 to help combat this downward trend. This study presents findings from another prenotice experiment fielded in the 2021 cycle of the same survey. While the stage is not different than past studies, we have the opportunity to see if anything has changed in regard to prenotice letters as survey response rates have continued to drop overall. Additionally, we can measure whether there is a differential impact on new sample cases compared to those that received the survey in prior cycles.

Testing Changes to the Initial Mailing of the American Community Survey
Lindsay Longsine, *U.S. Census Bureau*
Samantha Spiers, *U.S. Census Bureau*

The American Community Survey (ACS) is a national survey sent to a sample of U.S. addresses each month. To encourage self-response, the Census Bureau sends up to five mailings to each address in sample. Residents are asked to respond online or by mail. Due to negative impacts the COVID-19 pandemic had on Census Bureau operations, the ACS mailout strategy was changed from March 2020 through March 2021. Changes included: reducing the number of mailings, sending certain mailings to only a subsample of nonrespondents, and replacing some of the pre-assembled mailing packages with pressure seal letters. Response rates were continuously monitored during this time, and we saw nominally higher returns in months in which the initial mailing was a pressure seal letter compared to months in which the initial mailing was the normal, pre-assembled package. Since a pressure seal letter is less expensive than the mail package, we explored this result further by conducting a controlled field test after operations returned to normal. In May 2021 the Initial Mailing Pressure Seal Test was conducted to determine whether replacing the initial mail package with a pressure seal letter would affect response. This presentation will describe the experimental design and provide the results of the test.

Evaluating the American Community Survey Regional Office Internet Letter
Samantha Spiers, *U.S. Census Bureau*
Michael Risley, *U.S. Census Bureau*

During the American Community Survey (ACS) Computer-Assisted Personal Interviewing (CAPI) operation, the nonresponse followup operation for the ACS, there are several letters available that field representatives can provide to nonrespondent households to motivate them to respond. One of these letters, the Internet Letter, provides an internet user ID and instructions on how to respond online. In 2020, the Census Bureau decided to send the Internet Letter nationally to all mailable CAPI cases instead of regional offices sending it at the discretion of the field representative. We designed a new version of the Internet Letter to be consistent with the design of the other ACS mail materials, which are less text dense and use plain language to increase the likelihood of survey cooperation. We tested four variations of the letter, varying the address-side message ("Past Due" or "Required by Law") and the listed response options (internet only or internet and Telephone Questionnaire Assistance). In this talk, we discuss which content options were most effective at increasing self-response and combating decreasing CAPI response rates.

A Strategic Framework for Messaging in the American Community Survey
Samantha Spiers, *U.S. Census Bureau*
Broderick Oliver, *Census*

The American Community Survey (ACS) is a continuous national survey conducted by the U.S Census Bureau. The Census Bureau employs a mail contact strategy for the self-response phase of data collection. To motivate self-response, addresses in sample receive up to up to five mailings. The Census Bureau conducts considerable research and experimentation to improve the ACS mail materials, as the self-response mode of data collection is more cost effective than in-person follow-up interviews. In 2017, the Census Bureau took a radical approach to improving the ACS mail materials and embarked on a long-term project to redesign the materials from scratch. The focus of the redesign was new and improved messaging to motivate response. The Census Bureau conducted extensive research in different disciplines and learned what kinds of messages are most likely to gain survey cooperation. To improve how messages in the mail materials are received, understood, and acted upon by the recipients, the research suggested the use of plain language. Our research led to the development of a framework for ACS messaging. This presentation will discuss the development of this framework and how this framework and plain language principles were used to develop four new sets of ACS mail materials from the ground up.

Evaluating "Strategic Framework" Designs for American Community Survey Mail Materials
Lauren Contard, *N/A*
Broderick Oliver, *U.S. Census Bureau*
Dorothy Barth, *U.S. Census Bureau*

In September 2021, the U.S. Census Bureau conducted a field test of mail materials constructed using a strategic framework for messaging. The framework is discussed in the presentation "A Strategic Framework for Messaging in the American Community Survey." The new sets of materials used research-based messages and plain language

principles to motivate survey response. Some of the messaging themes included: establishing legitimacy and trust, conveying tangible benefits, and reducing the sense of survey burden. Three test treatments used color along with icons, columns and headers, and sidebars as design elements to draw attention to key information. The fourth test treatment used no color and used the least text for a minimalist design approach. This presentation will discuss how the different designs affected response to the survey.

Developing & Maintaining a COVID-19 Survey Panel During a Pandemic
Longitudinal Weighting Strategy for the Ohio COVID-19 Survey

Caroline Scruggs, *RTI International*
Marcus Berzofsky, *RTI International*
Naomi Freedner, *RTI International*
Robert Ashmead, *Ohio Colleges of Medicine Government Resource Center*
Bo Lu, *The Ohio State University, College of Public Health*
Amy Ferketich, *The Ohio State University, College of Public Health*
Timothy Sahr, *Ohio Colleges of Medicine Government Resource Center*

Tracking the impacts of the COVID-19 pandemic on a population is challenging given how fast the pandemic is changing. While cross-sectional surveys help explain the impact at the population level, longitudinal surveys can help explain the impact on the individual level. Both goals can be achieved through a survey with a panel design, but most states do not have a large enough panel to make representative estimates. To help inform COVID-19 policy and program decisions and understand how they impact Ohioans, in April 2020 the State of Ohio launched the Ohio COVID-19 Survey (OCS). The OCS leveraged a prior Ohio-wide population survey, the 2019 Ohio Medicaid Assessment Survey (OMAS), to build a panel from which survey participants could be obtained. The OCS is a weekly web- and telephone-based tracking survey which is representative of Ohio residents, and provides information on health measures such as the respondent's current health status and COVID-19 testing status, and economic measures such as employment status and consumer confidence. Using a rotating panel design, the OCS achieved 18 weekly, statewide, cross-sectional estimates and up to three repeated interviews with panel members. Under this design, the full panel was randomly split into six rotation groups which were sampled every six weeks over the three waves.   This paper will explore the methods used to create longitudinal weights for those who responded in the baseline and follow-up waves. The longitudinal weights used the baseline weights as the design weights and made adjustments to account for those who had become ineligible and for attrition/nonresponse. Since all respondents would have participated in a prior wave, we used prior items correlated with nonparticipation in the nonresponse calibration. This resulted in weights which represented the baseline population

Results of a Text Message Experiment on the Ohio COVID-19 Survey and Ohio Medicaid COVID-19 Survey

Meagan Brackin, *RTI International*
Naomi Freedner, *RTI International*
T.J. Nesius, *RTI International*
Padraic Burns, *RTI International*
Dain Palmer, *RTI International*
David Schultz, *RTI International*
Michael Nau, *Ohio Colleges of Medicine Government Resource Center*
Leyla Tosun, *Ohio Colleges of Medicine Government Resource Center*
Timothy Sahr, *Ohio Colleges of Medicine Government Resource Center*

As response rates have declined for traditional RDD surveys, finding more effective and cost-efficient modes of contact has become imperative, especially for surveys that rely on phone numbers as the primary mode of contact. Additionally, as call filtering/screening has becoming more prominent, getting respondents to answer their phones has become more difficult. Now, with the majority of phone users having cell phones, texting a survey invitation has become an increasingly used method of contact. However, there is little research as to what is the most effective time of day to text respondents. The Ohio COVID-19 Survey (OCS) and the Ohio Medicaid COVID-19 Survey (MCS) are biweekly surveys – which utilize a list frame where contact has been pre-approved – used to provide state and local health agencies

and officials with pandemic tracking data that can be trended to estimate an overall measurement of COVID-19 impact and citizenry response. The target sample size for each survey is 650 interviews per week. Because texting has proven effective on other listed sampled surveys and, with a short fielding period, a rapid protocol of texting, emailing (when email addresses were available) and calling was implemented. Survey invitations were sent on Monday via text and email, followed by text and email reminders on Tuesday, Thursday, Saturday, Tuesday, Thursday, and Saturday. CATI calling began on the first Wednesday of the two-week period and continued throughout. To examine if there was a more effective time of day to send text messages, both projects were split into two experiment groups, one that would be texted weekday mornings, at approximately 9:00 a.m., and one in the early evenings, at approximately 5:00 p.m. This paper will examine the outcomes of the texting experiment between the two groups and between the OCS and MCS since they make up different populations.

Quickly Fielding a Study During the Early Days of a Pandemic While Transitioning to a Virtual Call Center: Ohio COVID-19 Survey

Dain Palmer, *RTI International*
T.J. Nesius, *RTI International*
Meagan Brackin, *RTI International*
Naomi Freedner, *RTI International*
David Schultz, *RTI International*
Caroline Scruggs, *RTI International*
Padraic Burns, *RTI International*
Kurt Johnson, RTI International
Timothy Sahr, *Ohio Colleges of Medicine Government Resource Center*

Fielding a survey study is a challenge in the best of times, but in the beginning of a global pandemic understanding how populations were responding to the pandemic added to the urgency of starting data collection quickly. We began data collection for the Ohio COVID-19 Survey (OCS) on April 20, 2020 – 10 days after the request to initiate the project, with a protocol that includes text messages, email, and phone calls. The OCS began as a weekly survey, designed to quickly provide state and local health agencies and officials with pandemic tracking data that can be trended to estimate an overall measurement of COVID-19 impact and citizenry response. In addition to an extremely short window to prepare for data collection, RTI was simultaneously transitioning from an in-person to a virtual call center due to the raging COVID-19 global pandemic. This paper will outline how we managed an extremely aggressive launch date by compressing our normal pre-survey operations while maintaining necessary quality assurance processes. For example, the schedule for survey instrument programming and testing was broken into hourly increments. Because the schedule did not allow for a full integration testing of systems, we enacted modified systems testing and very carefully monitored once data collection began. Simultaneously, most RTI staff had recently shifted to remote work and RTI was building infrastructure to shift to a virtual call center. This included distributing equipment and providing technical support to interviewers and configuring systems and services for a remote environment while interview training and quality control were completely shifted to using online systems, including Zoom. These steps resulted in a successful and timely survey launch for a study that remains in the field.

Developing and Maintaining a COVID-19 Survey Panel During a Pandemic

Naomi freedner-Maguire, *RTI International*
Caroline Scruggs, *RTI*
Dain Palmer, *RTI*
Timothy Nesius, *RTI*
Padraic Burns, *RTI*
Robert Ashmead, *The Ohio State University*
Michael Nau, *The Ohio State University*
Timothy Sahr, *The Ohio State University*
Leyla Tosun, *The Ohio State University*

To help inform COVID-19 policy and program decisions and understand how they impact Ohioans, Ohio's health and human services agencies initiated two population-based COVID-19 tracking studies - the Ohio Covid Survey (OCS) in

April 2020 and the Ohio Medicaid COVID Survey (MCS) in September 2020 - to examine the impacts of the pandemic on the population's health, mental health, socioeconomic status (particularly employment and housing), COVID testing, and vaccine adoption. The OCS leveraged a prior Ohio-wide population survey, the 2019 Ohio Medicaid Assessment Survey (OMAS), to build a panel from which survey participants could be obtained. The MCS panel was developed from respondents from the 2019 and 2021 Ohio Medicaid Community Engagement Evaluation (OMCEE) and a random sample drawn from Ohio Medicaid's Affordable Care Act expansion population. Samples were selected to provide researchers with baseline information prior to the pandemic's arrival in Ohio; with the intention of examining Ohio's general and Medicaid populations pre-, during-, and post-pandemic (both surveys are currently planned to continue through June 2022). The OCS and MCS both have a two-week collection cycle targeting 650 completed surveys. Every other Monday a representative sub-sample is selected and fielded using a rapid protocol of texting, emailing and calling. Panel respondents are resurveyed every 8 (OCS) or 12 (MCS) weeks. The nature of these dynamic surveys has provided opportunities for experimentation and process evaluation. This panel will explore multiple topics, including launching the OCS in 10 days while transitioning to a remote work model; how the time of day that a text message is sent impacts the response; the mode of response (web or phone) by respondents over time and factors correlated with mode; the use of electronic gift cards and differences in redemption by survey mode and demographics; and methods for developing cross-sectional (population-level) and longitudinal (individual-level) weights.

Evaluating the Effectiveness of an Electronic Gift Card Incentive in the Ohio COVID-19 Survey and the Ohio Medicaid COVID Survey

Padraic Burns, *RTI International*
Caroline Scruggs, *RTI International*
Naomi Freedner, *RTI International*
Dain Palmer, *RTI International*
Mike Nau, *Ohio Colleges of Medicine Government Resource Center*
Tim Sahr, *Ohio Colleges of Medicine Government Resource Center*

In March of 2020, Ohio's health and human services agencies initiated two population-based COVID-19 tracking studies to examine physical symptoms, preventive practices (e.g., social distancing, mask wearing, hygiene), employment changes, and socioeconomic dynamics related to the pandemic within Ohio residential populations. These studies started in April 2020 and will continue through June 2022. The Ohio Covid Survey (OCS) sample consists of respondents from the 2019 Ohio Medicaid Assessment Survey (OMAS), a general population survey that examines healthcare access and use, insurance status, health behavioral risks, self-rated health status, and health demographics (OCS); the Ohio Medicaid Covid Survey (MCS) sample comprises respondents from the 2019 and 2021 Ohio Medicaid Community Engagement Evaluation (OMCEE) and a random sample drawn from Ohio Medicaid's Affordable Care Act expansion population. Both OCS and MCS have a two-week collection cycle where a representative sub-sample is selected, fielded for two weeks, and then resurveyed every 8 (OCS) or 12(MCS) weeks. Respondents can complete the survey by web or phone and are offered a $5 electronic Visa card incentive for each wave of the survey they complete. The benefits of using these electronic Visa cards are that the respondent receives the incentive instantly via e-mail upon survey completion, and delivering the incentive requires no fulfillment labor outside of initial set up. Some respondents may find them difficult to use, however, which may lead to attrition in a longitudinal study. The presentation will: 1) compare the likelihood of redeeming the Visa e-card across the two different survey populations and modes; 2) examine whether there are differences in socioeconomic characteristics correlated with gift card redemption; and 3) examine whether gift card redemption affected the likelihood of participating in a subsequent survey wave independent of sociodemographic characteristics. Qualitative data from respondent contact logs that tracked incentive issues will also be examined.

Fielding COVID-19 Surveys During a Pandemic Using an Innovative Operational Approach

Timothy Nesius, *RTI International*
Naomi Freedner, *RTI International*
Caroline Scruggs, *RTI International*
Dain Palmer, *RTI International*
Meagan Brackin, *RTI International*
Padraic Burns, *RTI International*

David Schultz, *RTI International*
Leyla Tosun, *The Ohio State University*
Timothy Sahr, *The Ohio State University*

In March of 2020, Ohio's health and human services agencies initiated two population-based COVID-19 tracking studies to examine physical symptoms, preventive practices (e.g., social distancing, mask wearing, hygiene), employment changes, and socioeconomic dynamics related to the pandemic within Ohio residential populations. These studies started in April 2020 and will continue through June 2022. The Ohio Covid Survey (OCS) sample consists of respondents from the 2019 Ohio Medicaid Assessment Survey (OMAS), a general population survey that examines healthcare access and use, insurance status, health behavioral risks, self-rated health status, and health demographics (OCS); the Ohio Medicaid Covid Survey (MCS) sample comprises respondents from the 2019 and 2021 Ohio Medicaid Community Engagement Evaluation (OMCEE) and a random sample drawn from Ohio Medicaid's Affordable Care Act expansion population. Both OCS and MCS have a two-week collection cycle where a representative sub-sample is selected, fielded for two weeks, and then resurveyed every 8 (OCS) or 12(MCS) weeks. To maximize contact rates, protocols for both studies include sending text messages and/or email (electronic), if available, as well as follow up phone calls. The study can be completed either by web or phone and respondents are offered a $5 incentive. The protocol consists of sending an electronic invitation on day one, an electronic reminder on day two, and follow up electronic reminders on alternating days for a total of seven electronic contacts. Phone follow up begins on day three of each collection cycle. This presentation will focus on exploring the mode of response (web or phone), whether respondents complete in one mode versus another over subsequent cycles, cost implications, and whether there are key differences in socioeconomic and/or demographic characteristics between predominantly web or predominantly phone responders and whether electronic and phone contacts support survey completion in the other mode by functioning as a prompt.

What to Do About Quality? Different Approaches to Investigate Quality in Online Surveys

Who Participates & Why? Correlates of Nonresponse

Survey Fatigue and Declining Response Rates: Evidence from the 2020 Census and Household Surveys

Jonathan Eggleston, *U.S. Census Bureau*

One hypothesis to explain declining survey response rates over time has been that individuals are fatigued from receiving multiple survey requests. However, there has been little prior investigation of this hypothesis due to a lack of pertinent data, such as not knowing whether nonrespondents have participated in prior surveys, or small sample sizes (when prior survey participation is known). In this paper, I investigate the survey fatigue hypothesis by looking at self-response in the 2020 Census for the United States. Specifically, I examine whether households who were sampled in the American Community Survey (ACS) or the Current Population Survey (CPS) from 2015 to 2019 have a lower self-response rate to the 2020 Census. The large sample sizes of the ACS and CPS provide a unique opportunity to examine how the effects on self-response change over a long time horizon and how they vary by demographics in the household. I find that households recently sampled for the ACS had lower self-response rates to the 2020 Census, with the magnitude varying from -1.5 percentage points for households sampled in January 2019 to -15.1 percentage points for households sampled in December 2019. Similar effects are found for the Current Population Survey (CPS) as well. In summary, although there might be multiple reasons for these effects, these results provide potential additional evidence that the proliferation of surveys may be leading to lower response rates across the entire industry.

Causes and Consequences of Survey Non-Response for Religious Activity: Insights from the Notre Dame Undergraduate Religion and Spirituality Longitudinal Study

Kraig Beyerlein, *University of Notre Dame*
Michael Wood, *Brigham Young University*
Audra Dugandzic, *University of Notre Dame*
Brianna McCaslin, *University of Notre Dame*

We bring novel data to bear on the issue of survey non-response for the social scientific of religion. In the fall of 2019, the Notre Dame Undergraduate Religion and Spirituality Longitudinal Study (ND-URSLS) sampled nearly 2,000 undergraduate students to participate in a year-long project on religious activity. Participating students took a baseline

survey, and were then invited to enroll in a weekly study that included brief, weekly surveys about their daily and weekly religious/spiritual (and other) activities. At the end of the school year, all students who participated in the baseline survey were invited to participate in an exit survey. Participation varied across the different stages of data collection: Not all sampled students completed the initial survey; not all who did enrolled in the weekly surveys; among those enrolled in the weekly surveys, the number of weekly surveys completed varied; and not all students who finished the initial survey took the end-of-the-year survey. We first use university administrative data to explain why certain sampled students but not others completed the baseline survey. Next, these data, along with information from the baseline survey, are employed to predict differential response to the weekly and exit surveys. Results from these analyses are then used to develop weights that adjust for non-response for each data collection phase. Last, we estimate models using weighted and unweighted data to examine the consequences of non-response for various religious activities across the different data collection points.

Social-Psychological Aspects of Probability-Based Online Panel Participation

Sebastian Kocar, *Australian National University*
Paul J. Lavrakas, *The Social Research Centre*

In this mixed-methods study, we aim to identify, discuss, and investigate the main motivational factors and barriers in all stages of the probability-based online panel lifecycle – recruitment to the panel, wave-by-wave data collection, and voluntary attrition. Our qualitative and quantitative data were gathered in four ways. First, in a 2019 wave of Life in Australia™ almost 1600 panellists provided a written verbatim to an open-ended question about their motivations for joining the panel. Between March 2020 and February 2021, fifteen of these panellists were classified into four distinct groups based on their panel response behavior and each participated in an in-depth interview (qualitative) and DiSC personality assessment (quantitative). We also carried out content analysis of panel recruitment material and communication. The results from the survey and in-depth interviews identified (1) contribution-focused, (2) survey-focused, (3) incentives-focused, and (4) self-expression-focused motivations as the most prevalent reasons for online panel participation. The most commonly reported barriers, some of which lead to voluntary attrition, were (1) major life changes, (2) lack of panel information, and (3) format, difficulty, or repetition of questions. Regarding personality types (DiSC), non-complying panellists scored lower on dominance, and higher on steadiness, than more frequently complying panellists. The study also used qualitative data to link the reported motivation and barriers with the existing survey participation theory. We presented evidence on how social-psychological theories can be applied to explaining panel/longitudinal survey participation; they are sufficiently robust to help understand the time dimension of survey participation and behavioral change.

On the Internet, No One Knows You're an Activist: Patterns of Participation and Non-Response in an Online, Opt-in Survey Panel

Tori Gorton, *University of Pennsylvania*
Daniel Hopkins, *University of Pennsylvania*

This paper examines non-response in an online panel of Pennsylvanian respondents. In recent years, social science research has come to rely heavily on surveys conducted with opt-in samples recruited online. We investigate the characteristics of one such panel of 11,000 Pennsylvania residents recruited in 2020-2021. By focusing on one large state whose demographics closely mirror those of the United States overall, we are able to benchmark key measures against files of registered voters and campaign contributors as well as U.S. Census data and other surveys. Using this data, we analyze panel attrition, types of non-response encountered, and predictors of non-response. Through a natural experiment in which matched respondents were surveyed immediately before or after an election, as well as longitudinal analyses, we are also able to investigate the extent of differential partisan response. We find that our sample is very highly politically engaged, with verified turnout in the 2018 midterm above 90 percent. We find that ethnic/racial minorities are under-represented among our panelists. Respondents' political attitudes show very high levels of stability over time, and various question-order experiments fail to move attitudes. There is limited evidence of differential partisan non-response over time. In sum, this evidence indicates that similar samples are effective ways to assess the views of highly engaged respondents, but disproportionately lack less engaged voters.

Identifying and Mitigating Nonresponse Bias in School Surveys During COVID-19: The 2021 Adolescent Behaviors and Experiences Survey (ABES) Case

Xiaoyi Deng, *ICF*
Ronaldo Iachan, *ICF*
Lee Harding, *ICF*
Alice Roberts, *ICF*

The Adolescent Behaviors and Experiences Survey (ABES) was a CDC-sponsored national student survey based on the National Youth Risk Behavior Survey that assessed risk behaviors and experiences during the COVID-19 pandemic. The ABES was administered online to a nationally representative sample of high school students in the spring of 2021 when COVID-19-related interruptions to the instructional environment were present across the country. Response rates were 38% at the school level and 48% at the student level. This presentation discusses potential nonresponse biases identified during analyses along with the strategies used to ensure that the impact of nonresponse bias in survey estimates is minimal. Because no survey data are available from non-participating students, we compared participating and non-participating schools' characteristics to evaluate potential nonresponse bias. In addition to geography, school type, school size, and socioeconomic-related factors, the analyses examined instructional models, i.e., whether students in participating schools attended class in-person, virtually, or both (hybrid). Variation in participation status was observed by Census region, city vs. non-city, Title 1 spending, poverty level indicators, and school instructional model. Participating schools were more likely to be from the Midwest, less urban, poorer, and provided hybrid or 100% virtual instruction. Logistic regression models identified the independent effects of four school characteristics on school participation: Census region, urban status, school type, and poverty level indicator. We used nonresponse weight adjustment methods to minimize potential nonresponse bias. Nonresponse adjustment cells were based on the four significant predictors identified in the models. The separate analyses of student-level participation within responding schools found no significant relationships with the selected school-level variables. The analyses demonstrate that even in the middle of a pandemic, a national school-based survey could provide valid student data. The validity is enhanced by nonresponse weight adjustments which minimize potential biases.

Extending Multilevel Regression with Poststratification to Incorporate Variables With Unknown Distributions

Brittany Alexander, *Texas A&M University*

Multilevel Regression with Poststratification (MRP) is a powerful tool to handle differential nonresponse in both probability and non-probability survey samples. MRP combines hierarchical regression with poststratification by estimating the support within many unique cells. However, previously for a variable to be included in the regression it's population distribution across the cells must be known so that in the poststratification step an appropriate cell weight could be calculated. While the US Census and other sources provides population level information on a variety of variables some variables that inform opinions such as political party do not have the necessary information to be used in the MRP framework due to a lack of population data. However, many variables (political party, psychometric indicators) can be estimated with a multilevel regression using demographic characteristics from Census data. In this presentation, we present an iterative approach. First, multilevel regressions for non-Census variables are fit. Then for each MCMC iteration, a sample of the non-Census variable is drawn from the posterior for each poststratification cell. Finally, the samples of the non-Census variable are included to estimate the main parameter of interest using standard MRP. This method is applied to a probability panel on public support of gene drive, where attitudes towards science and scientific knowledge are key predictors of gene drive support but the population distribution of the attitudes and knowledge measured in the survey is not known.

Using Text Messages to Invite Web Survey Participation: Experiments to Identify the Most Effective Message Content

Brenna Muldavin, *RTI International*
Stephanie Eckman, *RTI International*
Naomi Freedner, *RTI International*
Burton Levine, *RTI International*
Carol Pierannunzi, *Centers for Disease Control and Prevention*
William Garvin, *Centers for Disease Control and Prevention*
Machel Town, *Centers for Disease Control and Prevention*

Joe Murphy, *RTI International*

Due to the well-known problems with telephone and mail surveys, we tested the delivery of web survey invitations via text messages. As part of a pilot study for the Behavioral Risk Factor Surveillance System (BRFSS) in November of 2021, we sent text messages to more than 150,000 cell phone numbers, inviting the recipient to complete a brief health survey online. Each invitation included a link to the web survey. We manipulated the text message content in a series of experiments. The presentation will compare response rates, representativeness, and cost per completed interview between experimental groups to identify which text messages were most effective. It will also discuss the substantial logistical difficulties we encountered in sending a high volume of text messages. In addition to asking health questions, the web survey also asked survey participants how comfortable they were clicking on the link that led them to the survey. The answers to these questions provided insight into the reasons that some text messages performed better than others. Our results will guide use of text messages to deliver web survey invitations.   The findings and conclusions of this presentation are those of the authors and do not necessarily represent an official position of the Centers for Disease Control and Prevention

Data, Language, & Diversity: Every Research Study Is Cross-Cultural Today, presented by the Cross-Cultural & Multilingual Research Affinity Group

Data, Language, and Diversity: Every Research Study Is Cross-Cultural Today, Presented by the Cross-Cultural and Multilingual Research Affinity Group

Mandy Sha, *mandysha.com*

The Impact of the Supreme Court's 2001 Facebook v. Duguid Decision on Compliance with the Telephone Consumer Protection Act (TCPA)

Roundtable Discussion: The Impact of the Supreme Court's 2021 Facebook v. Duguid Decision on Compliance with the Telephone Consumer Protection Act (TCPA).

Dave Roe, *ICF International*

Measuring Sexual Orientation in Federal Surveys

Choices Matter: How Response Options for Survey Questions about Sexual Identity Affect Population Estimates of Its Association with Alcohol, Tobacco, and Other Drug Use

Brady West, *University of Michigan*
Sean Esteban McCabe, *University of Michigan*

This talk will present results from a randomized experiment in the 2015–2017 National Survey of Family Growth, where a large national sample of U.S. individuals aged 15–49 was randomly assigned to one of two different versions of a survey question about sexual identity (one with three response options, including heterosexual, gay/lesbian, and bisexual, and one adding the option "something else"). Analyses of changes in the associations of sexual identity with alcohol, tobacco, and other drug use across these treatments revealed evidence of significant differences in the associations that remained robust after adjusting for socio-demographics. The results suggest that when individuals choose their sexual identity from a more limited number of response options, the heterogeneity of the sexual identity subgroups increases, weakening estimated associations of sexual identity with these behaviors. Open-ended questions may therefore be necessary to measure sexual identity and estimate its associations with substance use behaviors accurately in surveys. We will discuss measurement implications for future SOGI-related research in conclusion.

The Impact of Sexual Orientation and Gender Identity and Disability Questions on Response Rates

Linda Young, *USDA NATIONAL AGRICULTURAL STATISTICS SERVICE*
Barbara R. Rater, *USDA National Agricultural Statistics Service*

The USDA National Agricultural Statistics Service (NASS) is studying the effect including sexual orientation and gender identity (SOGI) and disability questions might have on respondents' willingness to respond to the 2027 Census of Agriculture. As a first step in this process, NASS initiated an experimental study to assess the potential effect that the

presence of these questions might have on unit and item response rates. A random sample of about 75,000 was drawn from the NASS's list frame of all active U.S. farms. A survey was administered by mail, on-line and over-the-phone. The four treatment groups were (1) control (none of the test questions were included); (2) only disability questions; (3) only SOGI questions; and (4) both disability and SOGI questions. The two treatment groups with SOGI questions were each split into two subgroups: (1) respondents were asked a confirmation question if their recorded gender at birth differed from their present gender and (2) respondents were not be asked the confirmation question. A four-page questionnaire was used for all treatment groups, with some questions being deleted from the control questionnaire to provide space for the treatment (SOGI or disability) questions. Research questions considered in this paper are (1) whether each set of new questions (SOGI or disability) affects either unit or item response rates and (2) whether the presence of two sets of new questions has an additive impact on response rates or whether some interaction is present. Additional research questions that need to be addressed prior to asking SOGI or disability questions on the 2027 Census of Agriculture are discussed

Distribution of Spanish Responses to Open Ended Sexual Orientation Questions

Betsari Otero Class, *US Census Bureau*
Marcus Burger, *US Census Bureau*
Mikelyn Meyers, *US Census Bureau*

Sexual orientation is a topic gaining increasing interest in survey research, as some surveys are adding or updating sexual orientation and gender identity questions. Several surveys from the U.S. Census Bureau include a question on sexual orientation with response options for straight, lesbian or gay, and bisexual, followed by a write-in option for "something else". The wording of some of these survey questions varies, but little research has been done to date on the optimal question wording for these questions on surveys that are also translated into Spanish. Our findings indicate that Spanish speaking respondents often select the "something else" response option and then write in answers that align with the "straight" category. For example, we found that some Spanish speakers refer to themselves as "normal" when trying to convey that they are heterosexual due to a lexical gap in Spanish. In particular, Spanish has no neutral, accessible term for "heterosexual" that is similar to "straight" in English. We coded the Spanish write-in responses to sexual orientation questions from three different surveys (the Census Barriers, Attitudes and Motivators Survey Mindsets study, the Census Opinion Survey, and the Census Attitudes Survey) into hard-coded categories (along with a miscellaneous category). In this paper, we discuss the distribution of these categories. In addition, we examine whether the write-in options are capturing the intended responses and discuss what problems there might be with the Spanish translation. These findings provide valuable insight into how Spanish speakers are interpreting the question and how these interpretations affect the response distribution when answering sexual orientation questions. Preliminary findings suggests that the "something else" response category with the write-in option is not capturing the intended responses in Spanish.

Measuring Sexual Orientation In Federal Surveys

Karen Stein, *Abt Associates*

Accurate and inclusive data collection that includes measures of Sexual Orientation (SO) can help improve outcomes and reduce disparities by ensuring that marginalized populations are included in data-informed decisions (Jans, Wilson, and Herman, 2018; Suen et al., 2020). However, best practices on how, when, and where to ask SO questions are still emerging as measures are gradually added to instruments. Presentations in this session will address measurement issues associated with SO items in a range of federal surveys and discuss how design decisions about wording and context can impact respondent experiences and survey estimates. Two presentations will discuss the impact of a "something else" response option on substance abuse and privacy and confidentiality measures, respectively, while a third presentation will explore demographic differences in valid versus non-valid open-ended SO responses. The final two presentations will discuss findings from work fielding Spanish-language versions of an SO question and fielding the SO question in a survey of farmers. Data will be presented from a mix of medium-to-large-scale federal surveys, including the Census Barriers, Attitudes, and Motivators Study, the Census Opinion Survey, the Census Tracking Survey, the National Survey of Family Growth (CDC), and the Farm Producer Study (NASS).

Who Says That?: Demographic Differences in Open-Ended Responses to Survey Questions About Sexual Orientation

Shelley Feuer, *New School for Social Research*
Matt Virgile, *U.S. Census Bureau*
Jessica Holzberg, *U.S. Census Bureau*
Renee Ellis, *U.S. Census Bureau*

Reliable measurement of sexual orientation is crucial to ensuring that marginalized populations are accurately represented. To capture the nuances of sexual identity, sexual orientation questions in surveys often include an open-ended "something else" response option, after closed options for "lesbian or gay," "straight," and "bisexual." However, diversity in who selects "something else" can pose challenges to analyzing and understanding data on sexual orientation. In one Census Bureau survey, most respondents who selected this category used it to assert they are "normal" or object to the question, leading to their exclusion from analyses. Yet, analysis of a national health survey focusing on sexual minority women, showed that most respondents chose "something else" because their sexuality is fluid; these respondents were also younger with unique social and health indicators. Perhaps one way to reduce measurement error in "something else," therefore, is by examining differences in open-ended "something else" responses for different populations. This paper addresses the extent to which the type of "something else" responses vary by sociodemographic characteristics in three large-scale federal surveys. We first discuss the development of a coding scheme to categorize open-ended responses by content and sentiment; then we examine the relationship between respondent characteristics such as age, sex, and education and the distribution of categories. For instance, we analyze whether younger respondents are more likely to report an alternate sexual identity (and which terms are most frequently used) and whether older respondents are more likely to write "normal" or "heterosexual" instead of selecting "straight." We also examine the profile of respondents whose responses show disapproval for the question. The findings suggest potential strategies for measurement of sexual orientation by determining when and for whom "something else" responses can be back-coded into closed-ended sexual minority options, taken seriously as alternate sexual minority identities, or treated as item nonresponse.

Blending Data from Probability & Nonprobability Samples

Voices of Australia: Design, Results and Lessons Learned from an Attempt to Establish an Online Panel of First and Second Generation Australians
Benjamin Phillips, *Social Research Centre, The*
Anna Lethborg, *The Social Research Centre*
Dina Neiger, *The Social Research Centre*
Karly Day, *The Social Research Centre*
Diane Herz, *The Social Research Centre*

Voices of Australia was an attempt to establish a hybrid probability-nonprobability panel covering first and second generation migrants from countries outside Canada, Europe, N.Z., South Africa, and the U.S., funded by the Scanlon Foundation Research Institute and designed and implemented by the Social Research Centre. The effort was stopped after a large-scale pilot. We describe the study design, results of the pilot, and lessons learned with respect to the size of the pilot, communications testing, sample design, and survey methods. Contact material was tested in 12 online mini-focus groups (n = 44 participants total) of the target population, segmented by national origin and gender. The design used address-based sampling (ABS) with strata based on incidence of the target population as the 'spine' of the panel, to which subsequent nonprobability samples focused on specific migrant populations could be calibrated to. Households were approach via mail in a push-to-web design in multiple languages. Out of concern about topic-related nonresponse, a two-phase approach to screening (Brick, Andrews and Mathiowetz 2016; Brick, Williams and Montaquila 2011; Montaquila et al. 2012) was used, with the initial approach being blinded with respect to panel recruitment until after screening; the panel invitation was extended within the web survey to qualified households. We attempted to recruit multiple eligible household members. When it became clear that yields from ABS were extremely low (AAPOR RR3 = 3.8%; 137 panel members enrolled from 8,235 addresses sampled), we trialed nonprobability recruitment via community organizations working with the funder. This also yielded few (n = 144) panel members. Biases between the probability and nonprobability samples in many cases offset each other, suggesting that a hybrid approach may have had merit. This paper speaks to ABS designs outside the U.S., screening for hard-to-reach populations using ABS designs, and research methods for migrant populations.

Evaluating Coverage Error in Nonprobability Surveys Using Propensity Modeling

Burton Levine, *RTI International*
Michelle Cavazos, *RTI*

We present a novel way to evaluate if data from two different data sources represent the same population. This procedure is particularly useful to evaluate data quality from nonprobability surveys. A common practice in observational studies, where data is collected in both a treatment and a control group, is to use propensity scores to control for differences in the two groups. Where the distributions of propensity scores for the two groups overlap, the groups have common support . Where the distributions fail to overlap, there are members in one group without corresponding characteristics in the other group. The innovation we are highlighting is the application of this approach to evaluate the overlap of the demographic distributions from respondents sampled using two different survey methodologies. To illustrate this methodology, we evaluate two different combinations of data sources, one combination is address-based sample (ABS) data and random digit dialing (RDD) data, the other combination is ABS data with a nonprobability sample of individuals recruited from social media. For each combination of data sources, we calculate the propensity scores for being in one of the two surveys. The independent variables in the propensity models are the demographic variables used in the calibration procedure. We then compare the respondents' propensity scores across the two surveys. The areas where the propensity scores do not overlap indicate the presence of respondents in one of the surveys that contain combinations of demographic variables with no similar propensity score in the other survey. I conclude that the ABS and RDD data have common support. However, about 3% of the ABS respondents have no common support in the nonprobability social media survey, indicating some form of coverage bias, nonresponse bias, or selection bias in the nonprobability survey.

Measuring and Weighting Trended Data From Hybrid Samples. Or, What to Do When Your Data Falls Off a Cliff

Cameron McPhee, *SSRS*
Arifah Hasanbasri, *SSRS*
Mickey Jackson, *SSRS*

As the cost of fielding probability-based sample surveys continues to rise, nonprobability sample sources become an increasingly appealing alternative. However, it is well documented that drawing population inferences from nonprobability samples is challenging as the drivers of selection into nonprobability samples are unknown and have proven difficult to model. This is due, in part, to the fact that selection into nonprobability samples is related not only to demographics, but also to behavioral and attitudinal characteristics that are often unmeasured, lack existing benchmarks, and correlate to study outcomes. These challenges compound when measuring trends over time, as it is likely that the selection mechanisms at play are not only nonignorable but inconsistent. One potential solution, balancing the low cost of nonprobability sampling with the rigor of probability sampling, is to blend data from both sample sources. Such "hybrid" designs typically use the probability-based portion of the sample to provide benchmarks for non-demographic characteristics that may be correlated with selection into the nonprobability sample and study outcomes. This approach has been shown to reduce selection bias on outcome measures in some circumstances. However, in recurring hybrid studies, changes over time in the sourcing/composition of the nonprobability completes may still heavily distort trends. This paper presents an example of a trended research study facing this sort of challenge in which the bias introduced by the nonprobability portion of the sample increased significantly in the most recent study wave. We examine the possible sources of bias, including an investigation into the differences between "channels" through which the sample vendor recruits respondents. We also demonstrate the impact of a variety of calibration and weighting strategies, such as adding propensity weighting adjustments, which appear to mitigate some of the bias and may offer some insurance against unexpected changes in nonprobability sourcing, but not without costs.

Diving Deeper: Can Differential Weighting Improve Representativeness of Groups?

Kathleen Santos, *Ipsos Public Affairs*
Frances M. Barlas, *Ipsos Public Affairs*
Randall K. Thomas, *Ipsos Public Affairs*

Most samples intending to represent a general population have some divergence from the population demographics benchmarks. As a result, most researchers will compute weights that reduce or eliminate the divergences for key demographic factors. However, for racial or ethnic groups, the entire group is typically treated as one without regard to subgroups that might exist within the groupings (e.g., Latino overall versus Cuban, Mexican, Puerto Rican, or other Latino). We examined a study that had over 9,000 participants with probability-based KnowledgePanel and two opt-in non-probability samples. We explored using varying weighting algorithms to see if they could reduce bias overall or for specific groups. First, we established bias as average absolute divergence from federal benchmarks. With normal demographic weighting, we found that Latino in particular had higher rates of bias in both studies and for both KnowledgePanel and opt-in samples. We then used weighting targets for specific groups to see how they might be able to reduce bias. For Latino, we explored using weighting that incorporated Hispanic subgroup (Cuban, Mexican, Puerto Rican, Other) as well as using nested demographic groupings (age-gender within race-ethnicity, education within race-ethnicity, etc.). For Black respondents, we looked at regional distribution, also with age groupings and age-gender groupings. We then explored using other demographic variables beyond the more typical geodemographics in the weighting algorithms. We quantify each improvement in terms of bias reduction for each possible weighting algorithm we explored and determined replicability across studies. Additional weighting variables appeared to help reduce bias somewhat, but there may be other factors in recruitment and retention that are also contributing to the higher observed bias in specific groups.

Innovations in Hybrid Sampling Techniques - Improving Representation of Nonprobability Samples for Teens and Young Adults

Xiaolu Yan, *Ipsos - Seattle, WA*
Elisa Chan, *Ipsos*
Mansour Fahimi, *Marketing Systems Group*
Frances Barlas, *Ipsos*
Jennifer Kreslake, *Truth Initiative*
Elizabeth Hair, *Truth Initiative*
Abigail Giles, *Ipsos*
Larry Osborn, *Ipsos*
Randall Thomas, *Ipsos*

Traditional methods of survey sampling, such as those relying on address and telephone frames, continue to struggle with mounting inefficiencies due to coverage problems resulting from growing rates of nonresponse and their associated escalating costs of survey administration. It is from this perspective that practitioners are compelled to look outside of the box for more pragmatic options that do not forego rigor. In this context, it is becoming a common practice to supplement expensive probability samples with those secured from online panels for which cost of data collection is substantially reduced. Carried out effectively, the resulting hybrid samples can address both cost and coverage challenges of traditional samples – particularly when surveying rare or hard-to-get cohorts such as teens or young adults. While the literature on how to improve the external validity of survey estimates from nonprobability samples is evolving, what is currently available has focused solely on surveys of adults. Tasked with the unprecedented mission of developing practical techniques for hybrid sample surveys of teens and young adults that rely on nonprobability samples from online panels, this research has set out to explore weighting and calibration adjustment techniques that can address the unique challenges of surveying this hard-to-get cohort. The authors will present results from two large surveys of individuals 15 to 24 for whom the employed hybrid samples were secured from two sources: a probability-based sample of addresses from the USPS delivery database, and a supplementary sample from various online panels. Specifically, we will illustrate effective calibration procedures that go beyond basic geodemographic weighting adjustments to improve the representation of surveys of teens and young adults from online panels that are often subject to compromised representation.

Advances in MRP & Model-Based Small Area Estimation
Quantifying & Adjusting for Measurement Error

1436 Duke Street Alexandria, VA 22314

info@aapor.org | aapor.org | +1-703-827-1887