

# 2026 FedCASIC Virtual Conference

## Day 1: Tuesday, April 21

11:00 am – 12:00 pm, April 21

### **Welcoming Remarks and Keynote Address**

#### **Responsible AI to Advance Survey Design and Data Collection: A Framework-Driven Approach**

*Ting Yan, NORC at the University of Chicago*

Artificial intelligence is reshaping industries across the globe, and survey research is no exception. While recent scholarship—including Westwood’s (2025) PNAS piece—raises concerns about AI as a potential threat to traditional survey methodology, AI also offers meaningful opportunities to enhance survey processes, increase efficiency, and reduce costs. Realizing these benefits requires grounding any adoption of AI in rigorous methodological frameworks rather than pursuing technology for its own sake.

This keynote illustrates how the Survey Process framework and the Total Survey Error (TSE) framework together provide the structure necessary to identify where—and how—AI can responsibly improve survey design and data collection while managing risks and maintaining quality. The Survey Process framework enables a systematic assessment of pain points across the survey lifecycle, from initial study design through data collection, processing, and dissemination. Using this lens, we can pinpoint where AI interventions are most likely to alleviate bottlenecks, address resource constraints, or strengthen data quality. The TSE framework complements this approach by mapping each AI intervention to the error sources it aims to reduce, making visible trade-offs between error and cost. Just as importantly, the TSE framework provides a principled way to assess new error sources introduced by AI itself, including algorithmic bias, model drift, and other unintended consequences. By anchoring AI innovation in established survey methodology, organizations can responsibly harness these technologies to generate higher-quality, more trustworthy data.

1:00 pm – 2:30 pm, April 21

### **Concurrent Sessions**

#### **Session 1A: Integrating Human Expertise and AI to Advance Research with Large-Scale Data**

##### **Overview and Perspectives on Data Science and Artificial Intelligence in Research**

*Emanuel Robinson, National Academies*

This presentation provides an overview of recent Artificial Intelligence (AI)-related work by a national non-profit organization that advances evidence-based research. It explores the transformative role of AI in impacting scientific discovery, technology use, and the workforce. Recent activities and reports to be highlighted include syntheses of human-AI teaming research, discussions of AI's impacts on the technical workforce, an overview of AI's uses in statistical research and foundational models for scientific discovery, and opportunities in machine learning for safety-critical applications. Key considerations will be discussed, including methodological and research advancements, potential policy impacts, future directions, and applications in a range of fields. Attendees will gain insights into emerging AI-driven approaches and practical considerations for human-AI integration into scientific endeavors in a responsible and effective manner. Information about publicly available reports, presentations, and related materials and resources will be shared.

### **Predicting Missing Responses with Process Data: A Multiclass Machine Learning Approach**

*Qiwei He, Georgetown University*

Missing values pose a significant challenge, undermining the integrity and reliability of subsequent studies. Traditional methodologies have predominantly focused on statistical or Machine Learning techniques that rely on feature extraction to mitigate these issues. However, the effectiveness of these approaches is heavily contingent on the careful selection and engineering of features, thereby limiting their scalability and adaptability. This presentation describes the objective of this study in two-fold: (1) to investigate whether process data is sufficiently informative to predict missing values, and (2) to identify the most robust behavioral factors in missing response prediction. This study proposes multiclass machine learning methods to identify missing response patterns and pinpoint potential causes. A total of 11,468 respondents from 9 countries in the International Computer and Information Literacy Study 2018 cycle were used in the present study. In multiclass machine learning prediction, the macro-level accuracy ranges from 0.6 to 0.9 per item. The missing values in the high-performance group were better predicted than those in the low-performance group. Three predictors (i.e., response time, coding length, and remaining time) were found to be most robust in predicting missing responses, especially in adjacent items.

### **Using Machine Learning to Improve Qualitative Coding of High School Courses to Enhance Data Quality**

*Judy H. Tang, Westat*

The National Assessment of Educational Progress (NAEP) High School Transcript Study (HSTS) analyzes high school graduates' transcripts to report national trends in coursework and academic performance. A key component of HSTS is the qualitative coding of course catalogs, which requires assigning standardized School Courses for the Exchange of Data (SCED) codes to locally defined

courses to ensure comparability across schools and states. This process is labor-intensive and central to data quality. This presentation examines how machine learning (ML), specifically natural language processing (NLP), was integrated into the course catalog coding process to support human coders and improve efficiency, accuracy, and consistency. The ML approach uses a Python-based framework to generate semantic embeddings for course titles. These embeddings capture contextual meaning and are compared with a database of previously coded HSTS courses to generate similarity scores and suggest likely SCED codes, providing a more robust approach than lexical matching alone. Implementation followed a multi-phase experimental design, including system integration testing, output validation, and user testing with experienced coders. In output testing, 72.3 percent of the first SCED code suggestions matched the final assigned codes, and 91.2 percent of the top five suggestions included the correct code. User testing showed higher accuracy with ML assistance (74.2 percent versus 65.8 percent without ML) and faster coding times, particularly for routine courses (e.g., English 9, Algebra I). The presentation concludes with lessons learned and discusses how ML-supported, human-guided qualitative coding can improve data quality and operational efficiency in large-scale education and federal survey datasets.

## **Session 1B: Survey Respondents and Data Quality**

### **Practice Effects on Knowledge Items and Personality Measurement**

*Jesus Arrue, Westat*

*Hanyu Sun, Westat*

*David Cantor, Westat*

While practice effects are well documented in psychological and cognitive testing, less is known about whether they occur in address-based sampling (ABS) longitudinal web surveys. We examine practice effects as a mechanism of panel conditioning in a randomized four-wave longitudinal web survey. In web surveys, repeated exposure to identical items may inflate correct responses or internal consistency due to familiarity, memory of prior answers, or answer lookup rather than true change, potentially biasing estimates of change and distorting measurement error. As federal agencies increasingly rely on ABS push-to-web designs for official statistics and program evaluation, identifying and quantifying such effects is critical for maintaining data quality. The experiment was embedded in an ABS push-to-web panel of registered voters in two U.S. states.

Respondents were randomly assigned to one of three conditions: Group 1 answered the same knowledge and personality items in all four waves; Group 2 answered them in Waves 2 and 4; and Group 3 only in Wave 4, allowing us to isolate cumulative exposure effects across up to four administrations. We hypothesize that repeated measurement will increase correct responses to knowledge items and internal consistency of the personality scale in Group 1 relative to the other groups. We first conduct cross-sectional comparisons at Wave 4 using respondents observed in all waves. We then estimate structural equation models to assess model fit and test whether cumulative exposure shifts item thresholds, loadings, or residual variances across groups and waves. Findings inform questionnaire design, item

rotation strategies, and the assessment of measurement errors in federal longitudinal web surveys.

### **What do parents know? A National Health Interview Survey-Teen Investigation**

*Benjamin Zablotsky, National Center for Health Statistics*

*Amanda E. Ng, National Center for Health Statistics*

*Lindsey I. Black, National Center for Health Statistics*

*Jonaki Bose, National Center for Health Statistics*

*Jessica R. Jones, National Center for Health Statistics*

*Stephen J. Blumberg, National Center for Health Statistics*

Most nationally representative survey data focused on teen health use parent-report rather than directly asking teenagers. Parents may not be the ideal reporter for many experiences that shape teen health, like bullying by peers or internalizing emotions. Limited studies have examined concordance between parents and their teenagers, and few have done so using nationally representative data. The current study leveraged self-reported online data from the National Health Interview Survey-Teen (NHIS-Teen), in comparison with parent-reported data on the NHIS, an interviewer administered survey. NHIS-Teen was conducted between July 2021-December 2023. NHIS and NHIS-Teen are nationally representative household surveys conducted by the National Center for Health Statistics. An index of objective (e.g. doctor visit in past year) and subjective questions (e.g. life satisfaction) was created using questions asked of both parents and their teenagers (n=1,035 pairs). Differences in concordance between indices were explored by household-level characteristics (family income, highest parental education, urbanicity, number of parents). Overall concordance was found to be higher for objective questions when compared to subjective questions. For both objective and subjective questions, concordance was higher in families with higher income and higher parental education. For subjective questions only, concordance was higher in two-parent households. Findings from this study suggest that the likelihood of parent-teen concordance significantly varies by sociodemographic characteristics. Surveys may consider using online modes to obtain data directly from teenagers to supplement traditional parent-report surveys, especially when low concordance between parents and teens is expected.

### **Optimizing Call Attempts Using Response Propensity and Marginal Yield Analysis**

*Rafael Sobrino, Statistics Canada*

*Rachid Boudiaf, Statistics Canada*

Statistical agencies are increasingly facing the dual challenge of rising collection costs and declining response rates, necessitating a shift towards more dynamic, data-driven collection strategies. This presentation introduces a project undertaken at Statistics Canada to optimize telephone follow-up operations for a major household survey. The core objective is to move beyond one-size-fits-all rules by leveraging response propensity (the likelihood that sampled individuals will choose to participate in the survey) models to tailor

collection efforts. Our approach involves segmenting the survey sample into quintiles based on a predicted response propensity score, derived from an Elastic Net model using rich socio-demographic data. This segmentation provides a framework for analyzing the effectiveness of Computer-Assisted Telephone Interviewing (CATI) call attempts across different respondent profiles. We will present the methodology used to model response propensity and the distinct socio-demographic and behavioral profiles of the resulting segments. The presentation will then focus on a cost-benefit analysis that quantifies the marginal yield of successive call attempts for each segment. We will demonstrate how this analysis identifies a clear point of diminishing returns, typically around 12 attempts, beyond which continued calling yields minimal gains in completed interviews. We will discuss the operational implications of implementing a data-driven stopping rule, including significant potential cost savings and the opportunity to reallocate resources to more effective contact strategies. The findings provide actionable evidence for developing more efficient, targeted, and sustainable data collection frameworks.

### **Examining the Link Between Survey Cooperation and Response Quality in a Diary Study**

*Christopher Antoun, University of Maryland*

*Joseph B. Rodhouse, U.S. Department of Agriculture, Economic Research Service*

*Elina T. Page, U.S. Department of Agriculture, Economic Research Service*

This study explores the potential link between nonresponse error and measurement error, two issues often treated separately in survey research. Using paradata and self-reports from a U.S. national food diary survey, we examine whether an individual's likelihood to cooperate with a survey request influences their response behavior upon participating. We also examine factors such as respondent motivation, cognitive ability, and task difficulty that may simultaneously affect both response propensity and response behavior. Our results indicate that individuals with a lower propensity to participate in the diary survey were more likely to delay logging food events and backfill—adding entries for previous days that were missed—compared to those with a higher response propensity. However, no differences were found between these groups for other indicators of response quality. Additionally, while the factors we examined predicted cooperation, they did not account for variations in response quality. These findings suggest that recruiting reluctant respondents may inadvertently lead to increased problematic behaviors, such as backfilling. Nonetheless, the reasons behind the connection between response propensity and response behavior remain unclear, warranting further investigation.

2:45 pm – 4:15 pm, April 21

### **Concurrent Sessions**

#### **Session 2A: Data Dissemination and Communication**

## **Alignment of Transparent Reporting with Priorities for Enhancement of Scientific Integrity**

*John Eltinge, U.S. Census Bureau (Retired)*

This paper explores the alignment of transparent reporting with stakeholder priorities for enhancement of scientific integrity in systems for production, dissemination and use of statistical information. Five concepts receive principal emphasis: (1) Complementary uses of transparent information for evaluating and reporting on the current production in our statistical information systems; and for future improvement of those systems. (2) Transparent reporting on intended zones of statistical work, as defined by (a) high-priority products; (b) intended uses of the resulting statistical information; (c) preferred features of the resulting profiles quality, risk and cost; and (d) assumptions regarding the operating environment. (3) Transparent reporting on multiple layers of outcomes, including (a) information on the impact of specific design and environmental factors on performance profiles from (2); (b) the distribution of important environmental factors, and of slippage in design factors; and (c) related empirical results for products from (2.a), and for related diagnostics.

## **Use of Linked Micromaps for Public Exploration of Official Statistics**

*Randall Powers, U.S. Bureau of Labor Statistics*

*Wendy Martinez, U.S. Census Bureau*

Linked micromaps were developed to display geographically indexed statistics in an intuitive way by linking them to a sequence of small maps. The approach integrates several visualization design principles, such as small multiples, discrete color indexing, and ordering. Linked micromaps allow for other types of data displays that are connected to geography, including scatterplots, boxplots, time series plots, confidence intervals, and more. In this presentation, we will show how linked micromaps can be used to better understand and explore relationships for populations and subpopulations, explore multivariate relationships, discover patterns of heterogeneity across time and space, and evaluate data quality. We will illustrate how linked micromaps can be used with examples using data from the Bureau of Labor Statistics Quarterly Census of Employment and Wages program and the Occupational Employment and Wage Statistics program.

## **Using the Census Data API to Access Census Data**

*Faith Whittington, U.S. Census Bureau*

*Tyson Weister, U.S. Census Bureau*

The Census Data API offers another way to access the same data that users see on the Census Bureau's main data platform, [data.census.gov](https://data.census.gov). Join this session for a brief overview of the Census Data API and how it can be used alongside [data.census.gov](https://data.census.gov). This session will focus on highlighting the advantages of disseminating data via an API for public data users. Specifically, we will discuss use cases where the API benefits different types of users – from routine Census data users to technical web developers.

## **Session 2B: Designing the Survey Collection Instrument**

### **The Effect of Display Design and Number of Follow-up Questions on the Accuracy of Survey Responses**

*Alda G Rivas, U.S. Census Bureau*

A branching question may lead respondents to additional follow-up questions, which results in higher burden. This may motivate respondents to change their response to the branching question to avoid the follow-up questions, or to provide inaccurate responses to the follow-up questions in an effort to reduce their burden. It is possible that the design used to display the branching question may interact with the number of follow-up questions to produce different levels of accuracy. We implemented a 3 x 3 between-subjects design to explore the effect of display design (next page, unfolding, grayed-out) and number of follow-up questions (3, 5, 7) on the accuracy of responses to the branching question and the follow-up questions. Our results indicated that most participants answered the branching question accurately, with most of the inaccurate responses being observed in the "unfolding" display design. We also found that the "grayed-out" and "unfolding" display designs were more likely to contain accurate responses among the follow-up questions compared to the "next-page" design, regardless of the number of follow-up questions, and that presenting participants with either 5 or 7 follow-up questions decreased accuracy, compared to presenting 3 follow-up questions. Based on our findings, our general recommendations include to avoid a "next-page" design and aim to present a smaller number (e.g., 3) of follow-up questions. However, if the survey design includes 5 or 7 follow-up questions, a "grayed-out" or "unfolding" display design is preferable. The findings from our study provide a clear guide for survey practitioners to choose the optimal combination of display design and number of follow-up questions to minimize respondent burden and maximize accuracy of responses.

### **Can We Predict Dropout? The Predictive Power of In-Survey Burden Evaluations**

*Erica C. Yu, U.S. Bureau of Labor Statistics*

*Robin L. Kaplan, U.S. Bureau of Labor Statistics*

*Douglas Williams, U.S. Bureau of Labor Statistics*

The Bureau of Labor Statistics and Census Bureau are conducting research on the modernization of the Current Population Survey (CPS). The CPS is a monthly survey with a longitudinal design collected using a combination of telephone and in-person interviews. Ongoing research is being conducted on ways to improve response rates, the respondent experience, and manage survey costs. One way to address these issues is to add a web mode of collection. Currently, the CPS is interviewer-administered only and adding the web mode would add a new self-administered option. A pilot test of mixed-mode administration

of the CPS was conducted in 2025. All respondents, immediately after completing a wave of the CPS either by self-response on the web or by personal interview, had the option to provide feedback about their survey experience by answering closed-ended questions, including about how burdensome the survey was, how relevant the questions felt, and how easy or difficult it was to answer the questions. The debriefing questions were repeated at the end of each monthly wave of interviews; respondents could participate in up to three waves for this pilot test. This design resulted in a dataset of more than 2,500 cases with data on response mode, demographics, debriefing responses, and whether respondents dropped out in later waves. Analysis focused on evaluating the survey experience between web and personal interview modes, whether labor force status is associated with ratings of difficulty and burden, and debriefing responses associated with survey non-response in subsequent waves. Discussion will include debriefing questionnaire design, factors related to the respondent experience, and the relationship between response mode and burden.

### **Web-Probing: Motivating Detailed Responses in the Absence of an Interviewer**

*Robin Kaplan, U.S. Bureau of Labor Statistics*

*Tywanquila Walker, U.S. Bureau of Labor Statistics*

Web-probing, or web-based unmoderated pretesting, is an increasingly common survey pretesting methodology that complements traditional cognitive interviewing in an online self-administered format. Participants in web-probing studies are often asked a mix of closed- and open-ended probes to provide insight into their response processes, including comprehension, recall, judgment, and response formation. The lack of an interviewer in unmoderated testing may result in participants satisficing, insufficient responses, or skipping questions entirely, resulting in poor pretesting results or no data. Research shows that motivational statements (prompts) asking survey respondents to provide answers to items they skipped, or to important questions, can be effective at reducing item nonresponse and increasing response quality in web surveys. However, use of such prompts has not been assessed in web-probing studies. To understand how effective prompts are in the context of web-probing, an experiment was embedded into a web-based study pretesting survey questions about labor force participation and disability. The instrument included seven open-ended probes. Online participants (N=380) from a nonprobability panel were randomly assigned to either receive a prompt encouraging them to provide a detailed response to each probe or were presented with the open-ended probe and no prompt. Item nonresponse and character count were assessed to determine whether prompts can increase response and length of open-ended responses. We found that prompts did not affect item nonresponse but increased the length of open-ended responses. Prompts worked best when the probes asked about topics that were personally relevant to participants. Implications for the design of web-probing studies are discussed.

### **Using Remote CAPI in Controlled Residential Settings: The Implementation of the Survey of Prison Inmates R&D Field and Cognitive Test**

*Scarlett Pucci, RTI International*

*Ashley Murray, RTI International*

*Eliza Snee, RTI International*

*Tim Smith, RTI International*

The Survey of Prison Inmates (SPI) is a cross-sectional survey of state and federal prison populations conducted by the Bureau of Justice Statistics (BJS) since 1974, with seven collections, most recently in 2016. Traditionally, the SPI has been administered in-person by field interviewers. In 2024, BJS initiated the SPI Research and Design (SPI R&D) program to investigate ways to conduct the interviews remotely, potentially reducing the responsibility placed on facility staff and travel costs. We worked with BJS to design and conduct a field test and cognitive interviews. This research utilized both in-person and remote computer assisted personal interviews (CAPI) to test for mode effects when interviewing incarcerated populations. CAPI interviews relied on study- and facility-provided devices (e.g., laptops, WiFi, hot spots) and video teleconferencing platforms. For the field test, sample members were randomly assigned to the control (in-person) or treatment (remote) group within each facility. Results were analyzed to compare mode differences in terms of response, respondent burden, and data quality. Following the field test, protocols were updated to incorporate lessons learned and cognitive interviews were conducted using the same modes and devices. The cognitive interviews tested for differences in response quality between modes. This research is groundbreaking for the field of survey research by addressing whether data collection efforts within controlled settings can utilize remote data collection to collect reliable and comparable data across modes. Takeaways and lessons learned will be presented.

4:30 pm – 5:40 pm, April 21

### **Session 3B: Building the Survey Collection Instrument**

#### **A Scalable Framework for Standardizing Survey Questionnaire Versions**

*Shane Trahan, RTI International*

*Uma Maryada, RTI International*

*John Colin Matthews, RTI International*

*Madeline Cannon, FDNY Fire Department of New York*

Large survey programs such as those supporting FDNY's World Trade Center Health Program often accumulate multiple survey versions with overlapping analytical intent but inconsistent question wording, identifiers, and data structures. In many legacy data models, each survey version is stored in a separate table using version-specific question and answer identifiers. As a result, analytically equivalent questions may appear under different column names and answer codes across survey waves, requiring analysts to manually identify comparable questions, query multiple tables, and reconcile

inconsistencies. This approach is time-consuming, error-prone, and difficult to scale. This presentation describes a scalable methodology for survey question standardization that combines automated semantic matching with structured validation to support a unified survey data model. A standardized baseline survey definition is established using the most recent and contextually aligned survey versions. Questions from earlier waves are compared against this baseline using text embeddings and cosine similarity to identify exact and near-exact semantic matches. High-scoring candidate matches are further evaluated using a large language model to assess equivalence in meaning and analytical intent, producing interpretable similarity scores. Closely matched items undergo targeted manual review. Using these mappings, responses from all survey versions are consolidated into a standardized response structure with consistent question and answer identifiers, abstracting survey-specific differences into a reusable mapping layer. Applied to a multi-year FDNY health program, this approach reduced analytical complexity, improved cross-wave comparability, and established a maintainable framework

### **Leveraging Generative AI for Quality Assurance in Survey Questionnaire Development and Instrument Programming**

*Weihuang Wong, NORC at the University of Chicago*

*Ting Yan, NORC at the University of Chicago*

*Lilian Huang, NORC at the University of Chicago*

We describe a quality assurance (QA) pipeline for survey questionnaire development and instrument programming that applies generative AI to assist with key stages of this process, from questionnaire standardization through instrument validation. Our pipeline addresses four stages in the QA pipeline: (1) converting human-friendly questionnaires from varied Word document formats into standardized metadata structures, (2) performing quality checks on questionnaire logic including skip pattern validation and typo detection, (3) generating comprehensive test cases for instrument validation, and (4) verifying that the programmed instrument matches questionnaire specifications. At each stage, we suggest how generative AI can augment human review by systematically analyzing structured metadata, identifying potential issues, and generating test artifacts. We present detailed evaluations of two specific applications. First, we use generative AI to parse questionnaires and identify eligibility criteria for each item, a task traditionally requiring line-by-line manual review. Second, we apply generative AI to validate instrument programming code against intended survey flow specifications, checking that skip logic and fills match questionnaire requirements. We discuss challenges including handling complex skip logic and managing questionnaire formatting variability. We identify opportunities where generative AI shows promise for reducing QA burden while highlighting scenarios where human expertise remains essential. Our findings provide practical guidance for survey organizations considering AI-assisted quality assurance workflows.

### **Clear and Consistent: Building Internet Paradata Standards for Census Surveys**

*Renee Ellis, U.S. Census Bureau*

Internet paradata collects information about user and instrument actions over the course of filling out a survey. This provides important information for understanding user behavior that can be useful for improving surveys, targeting problems, understanding survey taking behavior, identifying fraud and many other survey uses. For over a decade, the Census Bureau has applied paradata in these ways and found it valuable. However, because paradata is non-standard, complex data, it presents some challenges. Although there is a great deal of collaboration across surveys, there are no universal paradata analysis standards like exist for sampling or survey methods, and much of the analysis of internet paradata has been done independently by individual surveys. While we have created some generic programs, definitions and formulas in the past, lack of analysis standards meant they were inconsistently used. As we move towards increased consistency in systems for all Census surveys, we have recognized a need for the creation of standard internet paradata formulas, definitions and methods for use across all surveys. Implementation of standards can decrease duplication of effort and increase consistency in the use and presentation of paradata analysis. This presentation details our current standardization efforts.

## **Day 2: Wednesday, April 22**

11:00 am – 12:00 pm, April 22

### ***Plenary Session***

#### **Leveraging Recent Technology for Data Collection**

*Christopher Antoun, University of Maryland*

Survey approaches that take advantage of recent technology, particularly smartphones, hold great promise for improving survey measurement. However, they also introduce complexities in design and implementation. Methodological research is essential for navigating these issues and enabling innovations in survey design. In this talk, I will describe my work examining three newer survey approaches: questionnaires optimized for mobile devices, app-based surveys modularized into shorter parts, and in-the-moment surveys triggered by geolocation data. Through these examples, I aim to highlight both the opportunities and challenges that mobile data collection methods present for survey researchers.

1:00 pm – 2:30 pm, April 22

### ***Concurrent Sessions***

#### **Session 4A: Sampling and Response**

## **The Latest on Limitations of Non-probability Online Panels and New Evidence on the Wisdom of Probability Sample Panels**

*Jon A. Krosnick, Stanford University*

Quickly after the Internet arrived, InterSurvey created the KnowledgePanel, a probability sample Internet panel that lives on today. But nearly as quickly, numerous firms created non-probability opt-in panels of people who volunteered to do surveys for money. Those latter companies quickly eclipsed the probability sample Internet panels inspired by the KnowledgePanel (AmeriSpeak, the Understanding America Panel, etc.) in terms of profit and sheer volume of business, due to rock bottom prices and claims of superior data quality. Slowly but surely, those claims have been proven wrong, over and over. One particularly visible incident involved highly publicized polls that were inaccurate in 2016, due entirely to the flooding of the marketplace with opt-in and river sample data. At the same time, polls using probability samples measuring the same phenomena were extremely accurate. And since then, an accumulation of evidence has made it clearer and clearer that fraud is an insurmountable challenge for opt-in polls. As long as sampling is not done by reaching real people through random selection methods, the “efficient” contact methods are at grave risk for attracting bots driven by wise AI that can pass all checks and bias survey results in any direction of interest to the bot creator. Equally challenging are “click farms” of employees in developing countries who pretend to be survey respondents without ever reading questions and each completing huge numbers of questionnaires. This talk will review the latest evidence showing the limitations of opt-in and river sampling and reviewing evidence that probability samples continue to be superbly accurate. All this attests to the wisdom of creating the Ask U.S. Panel.

## **A Second Pre-Incentive and Real-Time Monitoring: Impact on Clinician Survey Response Rates**

*Fenose Osedeme, RTI International*

*Austin DeSpirito, RTI International*

Evidence shows that pre-incentives can increase survey response rates, but effectiveness depend on factors such as timing and incentive structure. The National Dementia Workforce Study (NDWS) Community Clinician Survey in the Wave 1 phase, randomized 25,000 clinicians to three experiment incentive groups: 1) a \$100 post-paid; 2) a \$125 "early bird" post-paid for surveys completed within 10 days, then \$100 post-paid thereafter; or 3) a split-incentive with \$10 cash pre-incentive and a \$90 post-paid. After the initial invitation and one reminder, all sampled clinicians received a second \$10 cash pre-incentive with a paper (PAPI) follow-up mailing. We present AAPOR-Standard response rate comparisons across experiment arms following the second pre-incentive and identify which incentive structures yielded the highest response rate. The presentation highlights the technological systems and workflows that supported the experiment. We describe our integrated monitoring dashboard used for real-time response tracking by treatment arm and clinician type, automated system triggers for web, email and PAPI follow-ups, and field staff workflow

coordination. We present web-survey outreach innovations used in fielding including personalized links, targeted early bird messaging, paradata capture (timing and device), and adaptive reminders. Attendees will gain evidence-driven findings on pre-incentive effectiveness plus practical guidance on dashboard features, paradata to monitor, and workflow practices needed to run similar, technology-enabled experiments with hard-to-reach professional populations.

### **Using REDCap to Manage Respondent Driven Sampling (RDS) and Longitudinal Data in a Community Wound Study of People Who Use Illicit Opioids in Rural NC**

*Deirdre Middleton, RTI*

*Erin Erickson, RTI*

*Arnie Aldridge, RTI*

*David Leblond, RTI*

*Jon E Zibbell, RTI*

This study investigates a disturbing trend in rural North Carolina of deep and necrotic, soft tissue wounds appearing on people who use illicit fentanyl. We employed respondent-driven sampling (RDS) to recruit a targeted sample of 500 fentanyl consumers who will recruit their peers with numbered coupons to participate. Study modules include laboratory testing of drug samples, diagnostic picture-taking, a medical questionnaire, and a behavioral health survey. Photos and medical information are examined by a notable physician and dermatologist to identify and classify wound types. Participants return post-baseline for four follow-up visits incentivized by cash stipends that increase with each subsequent visit. This presentation demonstrates a custom REDCap implementation that integrates RDS management, longitudinal scheduling, incentive tracking, survey data collection, medical pictures, and laboratory-tested, drug results. REDCap workflows track coupon status, link recruiters to recruits via unique IDs, track appointments and dual incentives, and store diverse data sources over time. We describe RTI-developed REDCap modules and configuration choices that support data quality, reporting, participant management, and data security requirements. Our experience shows that REDCap can be adapted to manage complex RDS recruitment in longitudinal community-based studies that involve sensitive data from a vulnerable and hard-to-reach population.

### **Use of Linked Micromaps for Understanding Empirical Effects of Unit Nonresponse**

*Darcy Morris, U.S. Census Bureau*

*John Eltinge, U.S. Census Bureau (Retired)*

Linked micromaps display geographically indexed statistics in an intuitive way by linking exploratory and analytic data plots with a sequence of small maps. In this presentation, we show how linked micromaps can be used to explore and understand potential unit nonresponse bias. Geographic visualizations allow survey analysts to assess disparate empirical effects of unit nonresponse weight adjustments. We use Household Pulse Survey public microdata to visualize geographic variation in outcome estimate sensitivity

considering different survey weighting scenarios. Patterns uncovered with linked micromaps can be used to inform improvements in nonresponse weight adjustment methods as well as improvements in sampling design for subsequent survey implementations. Keywords: nonresponse bias; sensitivity analysis; survey weight adjustments; calibration; Household Pulse Survey.

#### **Session 4B: From System-Driven to Self-Service Models of Survey Development: Challenges and Opportunities – Roundtable Session**

##### **From System-Driven to Self-Service Models of Survey Development: Challenges and Opportunities**

*Roxanne Moadel-Attie, U.S. Census Bureau*

*Emily Reece, U.S. Census Bureau*

*Dameka Reese, U.S. Census Bureau*

*Mark Govoni, U.S. Census Bureau*

The Data Ingest and Collection for the Enterprise (DICE) Program at the U.S. Census Bureau aims to modernize computer-assisted data collection for more than 130 surveys. By advancing the technology that supports internet self-response (ISR), computer-assisted personal interviewing (CAPI), and computer-assisted telephone interviewing (CATI), the DICE Program has developed new platforms that enable a self-service approach to survey instrument creation and adaptive data collection. This roundtable discussion brings together both technological and survey owner perspectives on the shift from a system-driven approach to a self-service model of survey development. In particular, this roundtable will explore the challenges and opportunities of self-service survey development, including impacts on testing, workforce training, survey instrument and respondent material creation, business rule creation, cross-survey standardization, data extraction, operational use of paradata amongst other related topics.

2:45 pm – 4:15 pm, April 22

#### **Concurrent Sessions**

#### **Session 5A: Creating a National Survey on Special Education Spending (NSSSES)**

##### **Building the Student Resources Survey for the National Study of Special Education Spending**

*Jeremy Redford, American Institutes for Research*

*Alli Gilmour, American Institutes for Research*

*Kelly Linker, NORC at the University of Chicago*

The Institute for Educational Sciences, within the U.S. Department of Education, partnered with the American Institutes for Research (AIR), NORC at the University of Chicago (NORC), and Allovue, a PowerSchool company, to conduct the study to design the National Study of

Special Education Spending (NSSES). The NSSES will develop national estimates of what is spent to educate students who receive special education, on average and for students with specific types of disabilities. The design work for this study began in 2022, and the NSSES is currently being pilot tested with a national sample of districts, schools, and school staff. In this presentation, we will present on the development and testing of the Student Resources Survey (SRS), the core instrument for capturing the special and general education services students receive. Building on the U.S. Department of Education's Special Education Expenditure Project, originally administered via paper questionnaires, the team created a modernized, modular survey aligned with current data needs. We developed a measurement framework identifying all data sources required to construct spending estimates. Cognitive interviews with special education teachers and other school staff examined item clarity, response burden, ideal respondents, and factors motivating survey completion. The presentation will highlight how this iterative process produced a modular, highly adaptive SRS that enables respondents to designate other school staff to complete relevant sections and includes a student schedule component capturing daily resources and service providers. We will share findings from usability testing of the modular SRS, including respondents' experiences with the screener questions and the student schedule.

### **Sampling**

*Cong Ye, American Institutes for Research*

We will present a practical, two-stage sampling approach designed to produce balanced sample sizes across school district characteristics (such as district type and poverty level) while maintaining comparable precision for estimates of multiple student subpopulations. The design addresses common operational challenges in education surveys, including the uneven presence and size of subpopulations across districts and substantial variation in within-district intraclass correlations for key outcome measures. In the first stage, districts are selected using probability proportional to size sampling with both explicit and implicit stratification. In the second stage, student rosters are collected from sampled districts and used to guide subpopulation-specific student selection. Rather than relying on fixed sampling rates, student selection probabilities are dynamically adjusted through simulation to achieve target margins of error for each subpopulation. These adjustments are made in real time as sample sizes accrue, enabling target allocation of sampling effort across districts of varying size and composition. To reduce variability in student weights, conditional student selection probabilities are calibrated to district selection probabilities, accounting for the non-proportional nature of the first-stage sample. Clustering effects are controlled by setting maximum student selections per subpopulation within districts, with larger districts contributing more students to meet overall precision targets. This approach illustrates how adaptive, simulation-based sampling can be operationalized in complex, multi-stage surveys to improve precision, manage field constraints, and support robust subgroup estimation.

## **Estimation**

*Drew Atchison, American Institutes for Research*

*Arun Kolar, American Institutes for Research*

This study presents a methodological framework for estimating expenditures on education services using recent advances in AI-assisted data processing to reduce respondent burden while maintaining analytic precision. The approach synthesizes bottom-up estimates resources allocated to individual students collected through the Student Resources Survey (SRS) with top-down estimates derived from average per-pupil expenditures reported by local education agencies (LEAs). To limit reliance on survey-based methods for top-down estimates, we generate LEA-level spending measures from administrative data processed through machine learning models developed by Allovue/PowerSchool. Survey methods have been used to obtain top-down estimates despite substantial respondent burden due to challenges in generating consistent spending estimates using LEA administrative data. LEAs report fiscal data using different organizational structures and ways of categorizing expenditures, making it difficult to consistently and accurately report spending in different categories across LEAs. To lower researcher burden, these challenges are addressed using machine learning models that more consistently and accurately categorize expenditures. Using the AI-assisted categorized expenditures, we estimate average per-student overhead expenditures for LEA administrative functions separately for general and special education. Final student-level spending estimates are then constructed by integrating the LEA per-student overhead spending measures with the SRS-generated individual student spending estimates. This combined approach yields a methodologically robust estimate of education spending that reduces respondent burden while enhancing the accuracy and comparability of expenditure measures across LEAs.

## **Session 5B: Multi-Channel Outreach: A Survey Strategy**

### **Meeting the citizen customer where they are**

*Alisha Kim, Accenture Federal Services*

*Margot Moody, Accenture Federal Services*

Reliant on citizen trust and satisfaction, government agencies face the ongoing challenge of delivering exceptional, equitable, and responsive customer service. Multi-channel survey data collection offers a transformative approach by offering feedback mechanisms that meet the federal customer where they are. This presentation examines the implementation and impact of multi-channel surveys within a public sector context, offering actionable insights for government practitioners. By leveraging multiple feedback platforms---including phone-based surveys, web-based surveys, and mobile application or chatbot-based surveys---agencies can capture comprehensive customer experiences while overcoming geographic constraints and reaching heterogenous respondent populations, whether individuals or other entities such as households or businesses. This use case---a large

federal agency with customers dispersed around CONUS and OCONUS---demonstrates how multi-channel surveys have successfully increased response rates, improved processes, and enhanced transparency. Attendees will learn about multi-channel strategies for designing accessible surveys across platforms that meet government standards while maintaining data quality, as well as techniques for integrating analytics and visualization tools to harmonize data from multiple sources for real-time decision-making. Special emphasis will be placed on overcoming the challenges of connecting with a large public sector audience, linking survey data collection to service delivery channel, and the critical need for secure, compliant data handling.

### **A multi-faceted view of the federal customer**

*Alisha Kim, Accenture Federal Services*

*Margot Moody, Accenture Federal Services*

To facilitate data-driven decision-making, a multi-faceted understanding of survey respondents is critical to achieving meaningful insights. This presentation explores the adoption of a multi-channel survey strategy to enhance the representativeness and depth of survey data. By leveraging multiple survey channels---such as online platforms, mobile applications, telephone interviews, and in-person interactions---researchers can capture varied perspectives and mitigate the inherent biases associated with single-channel limitations. The presentation emphasizes the importance of channel selection in designing a robust survey strategy. Key factors influencing this choice include the target population, survey objectives, resource constraints, and the nature of the information being collected. For instance, younger, tech-savvy respondents may prefer mobile or online surveys, while older populations might respond better to telephone or in-person methods. A multi-channel approach not only broadens the respondent base but also accommodates communication preferences, ultimately leading to richer datasets. However, this strategy requires careful planning to ensure consistency across channels and maintain data integrity. The discussion includes practical considerations, such as aligning question formats, managing logistical challenges, and addressing privacy concerns. This presentation will provide context for selecting and integrating multiple survey channels effectively. Participants will gain actionable insights on tailoring survey strategies to specific research contexts. By adopting a multi-faceted view of respondents, researchers can unlock a deeper understanding of their audience and drive more impactful outcomes in federal data collection.

### **The power of more data**

*Alisha Kim, Accenture Federal Services*

*Margot Moody, Accenture Federal Services*

Data quality is imperative for effective decision-making, and its importance is magnified in the context of multi-channel survey strategies. By leveraging varied channels---such as online platforms, mobile applications, social media, and traditional methods like phone---

the volume and variety of data generated present opportunities to derive actionable insights, faster. We will explore the power of more data in enhancing data quality and accuracy, emphasizing the role of enriched datasets in improving representativeness of the target population and uncovering nuanced trends. The integration of data from multiple channels allows for cross-validation, ensuring consistency and reliability across responses. For example, discrepancies identified between online and offline surveys can signal potential biases, allowing for refined methodologies and adjustments for underrepresented demographics. Furthermore, the breadth of data collected through multi-channel approaches contributes to the robustness of post-hoc analysis, increasing confidence in results. However, the abundance of data also necessitates careful attention to data harmonization, cleaning, and ethical considerations. Techniques such as deduplication, normalization, and machine learning-based anomaly detection play pivotal roles in ensuring the integrity of multi-channel datasets. Additionally, the ethical use of survey participant data requires transparency, informed consent, and adherence to privacy standards to maintain public trust. We will show that agencies can harness deeper insights with more representation of the citizen customer by embracing the potential of multiple data collection sources. When managed effectively, multi-channel survey strategies are a dynamic tool for achieving superior data quality.

4:30 pm – 5:20 pm, April 22

## **Session 6A: Alternatives to Analysis of Free Response Data**

### **The People Behind the Enterprises: The Use of Administrative Records in Business Owners Statistics**

*Adela Luque, U.S. Census Bureau*

*Vitaliy Novik, U.S. Census Bureau*

*V. Araujo, U.S. Census Bureau*

*J. Earle, George Mason University, IZA, and U.S. Census Bureau*

*L. Ekerdt, U.S. Census Bureau*

*N. LaBerge, U.S. Census Bureau*

*J. Wold, George Mason University and U.S. Census Bureau*

*S. Young, Arizona State University and U.S. Census Bureau*

*J. Noon, U.S. Census Bureau*

In response to declining response rates, starting in 2020 the Census Bureau began providing nonemployer demographics not through a survey, but a program that leverages existing administrative and census records to identify the business owner universe and their characteristics: the annual Nonemployer Statistics by Demographics series (NES-D). Following the successful transition from surveying nonemployers to the adrec-based NES-D, a related initiative for employers, known as the Employer Characteristics Project (EC),

has been underway. While the methodology underlying EC is related to that of NES-D, employer businesses introduce novel challenges due to their more complex organizational structures. In this presentation we will cover our methodology, challenges, and comparisons of administrative record based demographics to those from surveys. Beyond unburdening surveys and lowering costs, we expect the Employer Characteristics data to enable substantial advances in business research. The linkage between businesses and business owners will allow researchers to examine the outcomes of owners of exiting businesses, the role of owner characteristic in firm ex ante heterogeneity, and more.

### **Using Microsoft CoPilot to assist with Data Analysis of Open-Ended Unstructured Data**

*Karen Castellanos-Brown, U.S.D.A. Food and Nutrition Service*

*Amanda Reat, U.S.D.A. Food and Nutrition Service*

*Olivia Deavers, U.S.D.A. Food and Nutrition Service*

*Savina Sparker, U.S.D.A. Food and Nutrition Service*

Qualitative data analysis can be very time consuming, in particular, when you have a large amount of unstructured data. Two research analysts at USDA Food and Nutrition Service were asked to analyze multiple years worth of WIC Breastfeeding Award application data. This task was taking a great deal of time given the large amounts of data there was in narrative form and in an unstructured format. The surveys used to collect the data consisted of some close-ended questions and many open-ended narrative questions and supporting documentation as attachments. This presentation will share how we used Microsoft CoPilot to facilitate and expedite our data analysis, the issues we encountered in using Microsoft CoPilot, and the validity checks we implemented to ensure our findings were accurate. We will also share our suggestions for ways Microsoft Copilot can be used to assist with data analysis, including what validity checks are important to prevent hallucinations and inaccurate results.