

# 2024 FedCASIC Virtual Conference

## Day 1: Tuesday, April 16

9:00 am - 10:00 am, April 16

### **Welcoming Remarks**

Keynote Address: **Data Integration in Survey Research: Possible Approaches to Addressing Future Challenges**

*Joseph Sakshaug, The Institute for Employment Research, Germany*

10:15 am - 11:45 am, April 16

### **Concurrent Sessions**

#### **Session 1A: Advances in Online Surveys**

##### **Revisiting Total Survey Error Framework In a Multimode and MultiData Environment**

*Ting Yan, Westat; Wendy Hicks, Westat*

Declining response rates and increasing cost of data collection continue to challenge survey practitioners, survey researchers, and government agencies. In an attempt to overcome these challenges, surveys are increasingly employing multiple modes to contact sampled persons and to collect data from them. At the same time, organizations and agencies are taking advantages of data from multiple sources (such as administrative data, data from devices and wearables, etc) to supplement and complement data from traditional surveys. The Total Survey Error (TSE) framework has been guiding survey design since developed (Groves, 1989; Biemer and Lyberg, 2003), but has not been revisited within the paradigm of multiple modes and multiple sources increasingly used by surveys. In this talk, we discuss how the TSE framework can be used to evaluate the impact of including multiple modes and multiple sources of data on survey estimates. We will also demonstrate how the TSE framework should be used to guide decisions when including multiple modes and multiple sources of data in a survey effort.

##### **Battling Survey Bots - Experiences and strategies to mitigate fraudulent responses in online surveys**

*Anwar Mohammed, RTI International; Matthew Benson, RTI International; Bob Henne, RTI International; David Schultz, RTI International*

Surveys are critical for federal agencies' to help understand community demographics, needs, preferences, and priorities; Create more relevant policies and services' just to name a few. Unfortunately, online surveys face an increased vulnerability to bots. If the survey has an open link, and involves any form of incentives (Cash, gift card etc.), it could attract the attention of fraudsters. Survey bots are essentially lines of code created to mimic human activity. In this case, they imitate human responses in online surveys, which could result in inaccurate data and making wrong data-driven decisions. These bots can finish surveys quickly and handle lots of them all at once, flooding survey responses, most of which will be fake, likely from users trying to get the incentives in malicious ways. In this presentation we will discuss three such surveys that were impacted by bots and the solutions that were implemented to detect and prevent them. We will also provide additional tips and tricks you can use to reduce the risk of survey bot responses.

### **QR Codes as a Method to Access a Mobile Survey**

*Alda Rivas, U.S. Census Bureau*

Accessing a survey through a smartphone can be a burdensome activity due to the need to type the website in a small keyboard. One potential solution is to include a Quick Response (QR) code in the paper invitations. However, there has been no research exploring the usability of QR codes as a method to access a survey through a smartphone. In this presentation, we share the findings of two rounds of usability evaluations: One round with members of the general population, and one round with older adults. For the general population, we found that all users were able to access the survey using a QR code and did so with high levels of efficiency and satisfaction. Furthermore, no usability issues were observed during this round. For the older adults sample, most users were successful with the task, including some users who have never used a QR code before, and they did so with high levels of efficiency and satisfaction. Furthermore, most users expressed preference for using a QR code in the future instead of typing a website using their smartphone. The usability issues that prevented users in the older adults sample to sign into the survey were due to characteristics of different operating systems (e.g., the design of the notification to open the webpage is mistaken as a text), or due to lack of experience with the QR code process (e.g., taking a picture of the QR code). Regardless of the usability issues encountered, both samples accessed the survey through the QR code with high levels of effectiveness, efficiency, and satisfaction. The high usability of this process observed in these studies indicates that providing a QR code in a paper invite

may be a viable way to reduce respondent burden when accessing a survey through a smartphone.

### **Transitioning In-School Paper Survey to Online Administration**

*Matthew Bensen, RTI International; Bryan Rhodes, RTI International; Sara Carter, RTI International; Elizabeth Parish, RTI International; Jean Robinson, RTI International; Scarlett Pucci, RTI International*

There are reasons to transition from paper and pencil data collection to the online mode, including improved data quality (by using skip patterns and data checks), easier deployment and administration, and the ability to more quickly process/display/analyze survey data. This transition, however, can also present challenges, particularly for school-based surveys that have specific sampling and respondent anonymity requirements. We recently worked with a state government client to transition a biannual school-based student survey from paper to online mode. We will discuss how we wove together three surveys, the student survey, a class sample selection survey, and a teacher survey that gathered data about survey administration, to meet the data collection need. We will discuss details from each of the surveys. We will convey how we programmed the student survey to know the provenance of each student survey without having to use individually based passwords or variables and how data captured in the class selection survey translated into emails that provided class-level information essential for taking the student and teacher surveys. We will show a dashboard providing daily data updates. Finally, we will discuss feedback from teachers and school administrators about the online experience.

### **Using Machine Learning to Identify Careless Web Respondents**

*Ting Yan, Westat; Gizem Korkmaz, Westat; David Cantor, Westat; Kevin Wilson, Westat; Olivia He, Westat; Rashi Saluja, Westat*

Surveys are increasingly employing multimode survey designs to increase response rates and contain cost of data collection. Web surveys are a popular mode of data collection for multimode surveys. However, web surveys are known to have data quality issues. For instance, web surveys may be completed by bots instead of actual human respondents. A larger data quality issue results from actual human respondents who satisfice when completing web surveys. These careless or insincere web respondents do not read survey questions carefully, do not spend time thinking about their answers, and do not follow instructions, and so on. As a result, it is critical to identify careless web respondents and to evaluate the quality of their answers before including them in analysis. Survey literature has documented various

techniques used to identify careless or insincere web respondents such as the use of attention checks, open-ended questions, and so on. This talk describes the use of machine learning to identify careless respondents. In particular, we applied four clustering methods; K-Means Clustering, Hierarchical Clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and Mean Shift Clustering to data from a web survey. We will present results from these four methods and examine convergence of the four methods in identifying careless web respondents. Implications of findings will be discussed.

## **Session 1B: Respondent Centered Design 1**

### **Designing post-submittal follow-ups on establishment surveys**

*Kathryn Harper, ICF; Michael Gibbons, National Center for Science and Engineering Statistics*

Efforts to improve data quality on establishment surveys often include post-survey follow-ups, as well as typical validations within the web questionnaire. These data quality follow-ups can often be the most time-consuming part of a data collection. The methods for these follow-ups are often limited by the survey platform and project timelines. We will present an innovative approach to these follow-ups implemented on the National Center for Science and Engineering Statistics (NCSES) Higher Education Research and Development Survey (HERD).

HERD is an annual census of U.S. colleges and universities that expended at least \$150,000 in separately accounted-for R&D in the fiscal year. The survey collects information on R&D expenditures by field of research, source of funds, types of research and expenses, and headcounts of R&D personnel. Since almost every value collected on the survey is published at the institution-level and a relatively small number of very large research universities can have a notable impact on national data, extensive follow-ups are conducted to ensure data quality. These follow-ups most often focus on drivers of large changes in expenditures from the previous year, item nonresponse, and reporting patterns on questions recently added to the instrument. ICF designed an interface that allows reviewers to post detailed follow-ups within the web questionnaire. Respondents can then respond to those follow-ups by revising submitted data on the web questionnaire or adding a text narrative. In addition to being able to see follow-up questions within the context of the questionnaire and implementing corrections immediately, respondents are interacting with navigation and design elements they have already become familiar with when completing the questionnaire.

## **Leveraging Technology for Survey Governance**

*Katie Petway, US General Services Administration; Camille Tucker, US General Services Administration*

Customer experience (CX) is increasingly a priority for the federal government, as evidenced by the establishment of CX offices in many agencies. However, most CX organizations have limited resources and authority. When CX is a broad mandate, how does a small team drive change management and agency-wide adoption of CX practices? What is the role of survey technology in building customer experience capacity? In this presentation, we will discuss how we have implemented enterprise survey governance as a first step toward more robust customer research practices.

## **Machine Learning for In-Instrument Product Code Search**

*Clayton Knappenberger, U.S. Census Bureau*

The North American Product Classification System (NAPCS) was first introduced in the 2017 Economic Census and provides greater detail on the range of products and services offered by businesses. In the 2022 Economic Census, NAPCS consists of 7,234 codes, and respondents often find that they are unable to identify correct NAPCS codes for their business. These respondents leave written descriptions of their products and services, and over 1 million of these needed to be reviewed by Census Analysts in the 2017 Economic Census. The Smart Instrument NAPCS Classification Tool (SINCT) allows respondents to search for appropriate NAPCS codes based on a written description of their products and services. SINCT is neural network document embedding model that embeds respondent searches in a numerical space and then identifies NAPCS codes that are close to this embedding. Using SINCT to search for product codes in the 2022 Economic Census helped reduce NAPCS write-ins by about 78%, and the majority of remaining NAPCS write-ins lack enough detail to be assigned to a single NAPCS code. Attendees will be interested to learn about how machine learning can improve respondents' experience while also reducing the amount of expensive manual processing that is necessary after collection.

## **QR Codes in the Census Household Panel**

*Casey Eggleston, U.S. Census Bureau; Aleia Clark Fobia, U.S. Census Bureau*

Although survey researchers were initially enthusiastic about the potential for QR codes to improve ease of access to web surveys and increase response rates, early research found that respondent use of QR codes was minimal and that the availability of a QR code had no significant impact on internet response rates (e.g., Marlar 2018;

Lugtig & Lutien 2021). Recently, updates to smartphones that have made QR code use easier and more intuitive as well as societal changes in the use of QR codes in response to the COVID-19 pandemic may have changed the landscape. Updated research is needed to assess whether this promising tool now has more utility as a method of accessing survey questionnaires. In fall 2023, the Census Household Panel, a national online survey panel maintained by the U.S. Census Bureau, recruited an initial set of panel members using a representative sample of 75,000 households in the United States. Invitations to join the panel were sent via U.S. mail, inviting panel members to fill out a recruitment questionnaire online (or via phone if unable to respond online). Invitation letters included both the printed survey URL and a QR code, along with a login ID. In this presentation, I report preliminary results for the Census Household Panel initial recruitment efforts, including what proportion of respondents used the QR code to access the instrument. I also share preliminary analyses of the demographic differences between QR code users and other respondents. I conclude by discussing implications for the use of QR codes in survey research.

### **Testing web surveys for accessibility - Challenges and solutions**

*Gauri Dave, RTI International; Neha Kshatriya, RTI International; Al-Nisa Berry, RTI International; Anwar Mohammed, RTI International*

According to the Census Bureau, 61 million adults in the US live with a disability. 1 in 4 (26%) of adults have visual, auditory, physical, speech, cognitive, or neurological issues. Section 508 Compliance is a federal government requirement to ensure that all information and communications technology developed by federal agencies is accessible for people with disabilities. This applies to all software projects, including surveys that are essential in data collection and critical in formulating government programs, policies, and decision-making.

To be compliant, web surveys should meet Level AA of Web Content Accessibility Guidelines (WCAG) 2.0, consisting of 12 guidelines and 61 success criteria/requirements. Understanding, interpreting, implementing, and testing these requirements could be challenging for developers and testers, resulting in surveys that are not accessible to some end users. This presentation will focus on two major accessibility requirements critical to making surveys compliant - Keyboard accessibility and Form elements. We will demonstrate examples of survey designs that can be made compliant with minor user interface changes or adding additional code (ARIA or HTML tags). We will also discuss common challenges while testing surveys (e.g., impersonating blind/deaf users, testing without a mouse), some best practices and recommendations based on our experience testing surveys for accessibility.

## **Session 1C: Advances in Data Science**

### **Empowering Data Science through Code Modernization: Bridging the Gap between Innovation and Efficiency**

*Roy McKenzie, The Coleridge Initiative; Ekaterina Levitskaya, The Coleridge Initiative; Yunie Le, The Coleridge Initiative; Nathan Barrett, The Coleridge Initiative*

Robust data and data systems serve as the foundation of data science and statistics. As our data systems have matured and become more complex, so too has the need to address legacy coding and best practices used in its development. Code modernization is a critical aspect of ensuring the efficiency, scalability, and sustainability of data infrastructure in today's rapidly evolving technological landscape. Code modernization, in this context, refers to the process of updating and optimizing the underlying codebase that supports data-related operations, processes, and use. Our work addresses the implementation of code modernization through a use case with the Economic Research Service (ERS) within the United States Department of Agriculture (USDA). We partnered with ERS to develop an updated data processing workflow for their Agricultural Resource Management Survey (ARMS). Our approach is centered around the concept of modular programming, and the implementation of modern coding concepts such as unit testing and automated data documentation using open source tools. By adopting this approach, the ERS has significantly improved the maintainability, adaptability, and documentation of its data processing code. These improvements are helping to ensure that the ERS's data infrastructure remains agile and responsive to the evolving needs of agricultural policy and program development.

### **Implementing RDF/OWL to Crosswalk the NAICS, SICS and GICS Industry Classification Systems**

*Christopher N. Carrino, US Census Bureau*

There are various industry classification systems available to classify business activity. This paper will explore OWL graphs of the primary industry classification systems including NAICS, SICS and GICS. We describe the history and relationship across the various systems. We then demonstrate a method to link and combine these graphs using the owl:equivalentClass relationship.

### **Name Entity Recognition for PII redaction on Open Ended Survey Comments for the Defense Organizational Climate Survey**

*Stewart Jollymore, DHRA-DPAC; David Endicott, DHRA-DPAC*

The Defense Organizational Climate Survey (DEOCS) is a continuously fielded survey aimed at measuring aspects of life for members of units across the military. Open ended comments on the DEOCS are a wealth of information that can be leveraged to gauge a wide array of insights, attitudes, and sentiment that traditional survey questions are unable to address. When sharing these comments with unit commanders and researchers there is a desire to redact potential PII what can be used to identify leaders, unit members, and the respondent. Traditional PII removal techniques have relied on hand redaction by a team of trained analysts. Given tight turn arounds for the DEOCS this method is inefficient. This project serves to leverage cutting edge Name Entity Recognition techniques for identifying potential PII in order to redact this information prior to delivery to unit commanders or researchers. The Defense Personnel Analytics Center's Data Science team has used the spaCy library to train models to extract custom entities from our domain specific lexicon which are considered PII (ranks, locations, etc.). When hosted in a production environment these models both speed up the PII redaction process and lead to lower labor costs of hand redaction.

### **Challenges with data linking and attribution in a universe administrative data collection**

*Evan Nielsen, American Institutes for Research; Marisa Pelczar, Institute of Museum and Library Services*

We have undertaken a series of integrated initiatives over the past five years to improve the methodological rigor and utility of an administrative data collection ('survey'). These efforts seek to enable researchers and the public to more effectively use the survey's data to document the value of these important social institutions in their communities. We attempted two research projects that sought to link the survey's data with other federal data sources to enhance the stories about the roles of these social institutions in the U.S. However, both research endeavors met with significant challenges, largely due to the structure of the survey data. Researchers involved in both efforts acknowledged the potential for bias and inaccurate results due to two related factors: (1) Linking a household record from another federal data source to the survey record for the closest institutional location may not accurately reflect the location that the household uses (or would consider using) or even whether the household is located in the service area for that location (which may affect which services household members could use at that location), and (2) attributing an institution's jurisdiction-level data to each institutional location within that jurisdiction masks neighborhood-level variation between those institutional locations. We have

attempted to address these factors by updating the geographic identifiers included in the survey's data files and by exploring the collection of some data at the institutional location level which is currently collected at the jurisdictional level. In this paper, we explain the research that led to these efforts and how we implemented the changes to mitigate the challenges encountered.

### **Automated Data Linking in the Absence of Time-Invariant Identifiers**

*Sun Kyoung Lee, University of Michigan*

This study examines record linking practices for datasets lacking time-invariant individual identifiers. The primary focus is on the application of a random forest classifier (RFC), a supervised machine learning technique for record linkage. The discussion emphasizes the strengths of RFC, including interpretability, scalability, and fine-tuning. Specific attention is given to methods for fine-tuning through k-fold cross-validation. The exploration delves into the selection process of record linking criteria, providing a detailed, step-by-step explanation of the record linking procedure. Additionally, the study addresses match correctness validation by computing false positive metrics based on data characteristics. Finally, a comparative analysis is conducted, evaluating the outputs of various methods against existing record linking techniques such as iterative automated record linkage and the expectation-maximization algorithm, using the same datasets.

1:00 pm - 2:30 pm, April 16

### **Concurrent Sessions**

#### **Session 2A: Advances in Data Collection**

##### **Implementing Custom Integrated Solutions for Efficient and Anonymous Data Collection in HIV Diagnosis Surveillance**

*Dave Roe, ICF; Josh Duell, ICF; Bridget Varian, ICF; Neetika Dembla, ICF; Kim Ethridge, ICF; Ashley Nielsen, ICF; Sam Williams, ICF; Jonathan English, ICF*

The Enhanced Surveillance of Persons with Early and Late HIV Diagnosis (Stage 0/3) project for the Centers for Disease Control and Prevention (CDC) collects vital data from individuals with recent HIV diagnoses at stage 0/3 of infection to help inform prevention, testing and linkage to care strategies. Data collection is a two-stage process that begins with participating health departments recruiting, screening, and registering individuals for the study. Once registered, participants are given the option

to participate in a web or telephone survey conducted and managed by ICF. Telephone surveys can begin immediately via real-time transfer or take place as a scheduled appointment. As part of the process, no personally identifying information (PII) is shared with ICF after participants are recruited by health departments. Data capture is centered around a customized solution that integrates multiple components and facilitates all stages of the project, from recruitment of participants to multi-mode survey data collection, to the distribution and tracking of incentives. Several custom processes have been developed to protect the anonymity of participants and ensure efficient data collection. This presentation will highlight the innovative steps ICF employed to create a cohesive system that keeps data flowing freely and safely. We highlight our approaches to registration and appointments, the assignment of customized IDs for privacy, the instant passing of registration data to the web or phone survey for routing, the masking of inbound caller information, the development of custom reports and the transfer of information back to health departments as a critical component of participant management and nonresponse follow-up.

### **Modernization of an HIV Data Coordinating Center**

*Kristen Flaherty, ICF; Nicole Gonzalez, ICF*

For the past 15 years, ICF has supported and coordinated two large HIV surveillance efforts, establishing a centralized system through which funded grantee sites submit collected data; monitor data collection progress; respond to data quality issues; and ultimately receive processed, cleaned data used to inform local HIV initiatives. These efforts and the centralized system that supports them have contributed greatly to curbing the HIV epidemic by providing data essential to understanding the risk factors for HIV, the populations most at risk for HIV, and the challenges and barriers faced by people living with HIV/AIDS. While the system has operated effectively for 15 years, there has always been a desire to produce high quality data faster so that public health organizations can act quickly. In line with our client's current data modernization efforts, ICF has been given the opportunity to significantly enhance many of the data collection and coordination processes. The Modernization of an HIV Data Coordinating Center presentation will provide an overview of the history of the effort, a summary of the current goals, improvements made to date and planned enhancements. The presentation will also discuss efficiencies, advantages, challenges, and lessons learned when modernizing the data collection and data management strategy for the project.

### **Assessing and redesigning the Identity Theft Supplement**

*Erika Harrell, Bureau of Justice Statistics; Alexandra Thompson, Bureau of Justice Statistics*

The Bureau of Justice Statistics' (BJS) Identity Theft Supplement (ITS) to the National Crime Victimization Survey (NCVS) is a crucial data source on the prevalence and characteristics of identity theft in the United States. The ITS was first administered in 2008 and approximately every two years after. The questionnaire, however, was redesigned prior to the 2021 data collection. BJS contracted RTI International in 2019 to assist in updating the instrument. This presentation will describe the redesign process for the ITS instrument, which included a review of current state and federal laws surrounding identity theft, solutions to address possible measurement issues, cognitive testing of a new instrument based on findings from analysis, and an online pilot test of current and new instruments to examine which survey was most accurate. Findings from the 2021 ITS will also be discussed, such as prevalence of identity theft in 2021, police reporting patterns, and emotional distress to victims.

### **TeleForm Enhancements for Survey Research**

*Daniel Keever, RTI International; Maurice Martin, RTI International; Ansu Koshy, RTI International*

RTI uses TeleForm to create, scan, and extract data from paper surveys. New features were added by RTI to expand the product functionality. These enhancements, based on project requirements and requests, have been created using TeleForm's scripting languages. These enhancements include but are not limited to:

- Too Many Marks. Captures multiple responses to a single response question and thereby differentiates between multiple responses to a single choice question and a question that is intentionally left blank.
- Audit Trail. Any modification to an answer is logged.
- Double Verification, including Data Review. Each form is verified by an operator and then confirmed by a second one.
- There are at least 2 operators assigned to verify a batch. As a data entry check, each operator's verification process is recorded and compared with other operators' records for discrepancies.
- Encrypted data at a dataset or field level.

We plan to briefly describe TeleForm as a survey research tool and review the enhancements. These enhancements have improved data quality and data management.

### **Barriers to the use of AI in Government**

*Jack P. Leeds, Ph.D., U.S. Census Bureau; Troy Brightson, Ph.D., U.S. Census Bureau; Kyana Beckles, M.P.S., U.S. Census Bureau*

This is a discussion about Artificial Intelligence, its usefulness, the communities eagerness to use it, and the barriers that we face in government to using the technology, and related technologies to their full potential. During this discussion, we'll discuss, FedRamp, security levels and other security challenges. We'll talk about definitions for AI, and recent legislation that impacts our use of AI. The technology also gives us opportunities to collect different types of data. Most of the data we collect is static. AI gives us the potential to collect and process auditory and visual data.

## **Session 2B: Improving Response Rates**

### **Efficacy Testing a Game to Educate and Collect Census Data**

*Peter Leveille, MITRE; Zue Lopez Diaz, MITRE*

In this session, we will share the design of an interactive and immersive game created to educate participants about the decennial census while collecting census data. The game is designed to help cue participants about commonly overlooked data, alleviate concerns about privacy, and educate participants about how the data will be used to help their communities. We will describe our process in efficacy testing the game with hard-to-count populations using A/B testing with an online census replica. We will also share the results of the efficacy study.

### **Would you prefer to contact them or shall we? Working with establishment gatekeepers to invite group quarters residents to respond to an online questionnaire**

*Alfred 'Dave' Tuttle, US Census Bureau*

This presentation shares results from a study to evaluate strategies for making a web questionnaire mode available residents of some types of group quarters (GQs). GQs are places where people live or stay in a group living arrangement provided by an organization, e.g., university student housing, military quarters, and worker dormitories. Making a web mode available to GQ residents would require coordination with the administrators who serve as gatekeepers for their organizations and residents.

This study evaluates two options. With the first option, administrators would provide names and email addresses for the residents of their GQs, and the Census Bureau would communicate directly with residents through invitation and reminder emails.

Administrators who prefer to handle communications with residents or are unable to provide names and email addresses would have the option to obtain and distribute login credentials themselves. A secure website will allow GQ administrators to provide residents' names and email addresses or to obtain residents' login information. Administrators would also be asked to send follow-up reminders to nonrespondents. What do GQ administrators think of these options? We conducted usability interviews using a wireframe mockup of the new system to assess the user-friendliness of the interface and look for opportunities for improvement prior to development of a prototype. This presentation will describe the proposed workflow of the coordination website to be used by administrators; administrators' reactions to the workflow, functionality, and tasks they would be expected to perform; and the steps they would have to take to provide residents' contact information and to distribute information to their residents by email.

### **Use of Clustering Methods on Paradata to Inform Surveys**

*Mengshi Zhou, Westat; Gizem Korkmaz, Westat; Ting Yan, Westat; Ryan Hubbard, Westat; Rick Dulaney, Westat; Brad Edwards, Westat*

As survey response rates continue to decline, strategies to increase efficiency of data collection are much needed. Paradata process data collected as part of survey data are often used to understand declining response rates and to inform responsive or adaptive designs. Paradata contain a vast amount of information on when and how sampled persons are contacted and the outcome of each contact attempt.

Researchers have conducted analyses of on call records data to identify unusual behavior and to inform further modeling, to explore stopping rules, to identify productive calling patterns, and to predict response outcome. In this paper, we use paradata associated with the Medical Expenditure Panel Survey (MEPS) and employ clustering techniques to group households with similar characteristics. To enhance the data's coherence and analytical power, we apply rigorous data cleaning procedures. Feature engineering and imputation methods played a pivotal role in shaping the data for machine learning analysis. Notable features include metrics related to contact attempts, such as the number of successful and unsuccessful contacts. Additionally, we extract features capturing household composition, such as the number of members to provide insights into survey dynamics. We use k-means clustering and identify four distinct clusters, each characterized by unique temporal patterns evident in communication models, sequence length, and final outcomes. Additionally, we observe variations among these distinct clusters based on demographic factors such as the race of household heads, language, household size, and the age of household

heads. The results underscore the importance of conducting further exploration of these clusters to inform survey management and interview strategies.

### **Improving Response Rates for the Quarterly Survey of Plant Capacity Utilization (QPC) by Automatically Mining Contact Information from Other Census Surveys**

*Jessica Huang, U.S. Census Bureau; Christian Moscardi, U.S. Census Bureau*

The Quarterly Survey of Plant Capacity Utilization (QPC) is a joint effort from the Federal Reserve Board (FRB), the Defense Logistics Agency (DLA), and the Census Bureau to estimate plant capacity utilization rates. The QPC relies on emails as a second avenue for informing and reminding sampled businesses to respond to the survey. A substantial portion of businesses do not have emails and do not provide quality responses, contributing to lower survey response rates. The goal of this project was to automate finding emails for QPC businesses. Previously, the QPC team has manually searched for missing emails by examining the emails logged by other surveys throughout the Census Bureau. This manual intervention yielded a significant decrease in email missingness and a significant increase in response rate. We present an automated Python process that not only found almost all manually detected emails, but also detected additional emails otherwise undetected by manual intervention. Compared to the manual approach, automated email mining lent itself to a less time intensive, more systematic, and more scalable intervention to decrease email missingness and increase response rates. Adding emails to businesses by mining information from surveys across the Census was one step forwards towards increasing response rates and improving the statistical quality of data products.

### **Session 2C: Applying Natural Language Processing to Survey Text Data - Roundtable Session**

#### **Applying Natural Language Processing to Survey Text Data - Roundtable Session**

*Erin Boon, BLS; Ayme Tomson, BLS; Daniel Todd, BLS; Melissa Pollock, BLS*

There is a lot of interest in expanding our use of machine learning at the Bureau of Labor Statistics (BLS), but Data Scientists face a particular set of challenges when applying natural language processing (NLP) methods to survey text data. For all the hype, NLP methods can yield disappointing results, but there are applications where they have been used effectively. In this roundtable session, panelists will share what they've learned while researching, launching, and maintaining NLP-powered products at BLS. The discussion will include examples of methods and recent projects at

different stages of development, best practices for managing NLP projects, and a review of some of the challenges to maintaining NLP products once built.

2:45 pm - 4:15 pm, April 16

### ***Concurrent Sessions***

#### **Session 3A: Best Practices for Accessible Data Visualization - Instructional Session**

##### **Best Practices for Accessible Data Visualization**

*Karen Moyes, Westat*

Data visualization is an effective way to communicate complex data. Using charts, maps, and infographics can simplify your message and increase comprehension, but unless you build in accessibility, you could lose a large part of your audience.

According to the CDC, 27 percent or up to 1 in 4 people in the U.S. have disabilities. Planning for accessibility early in your design and development is crucial to ensuring your data is clear and easily understood by everyone, no matter how they may access the information. How you use color, how well color elements contrast with each other and the background, and incorporating keyboard accessibility for interactive elements will allow more people to understand your message without the need for a complicated text version. Accessible data visualization benefits everyone. This session will discuss:

- How accessibility and data visualization best practices align and show examples.
- How incorporating accessibility at the beginning can save time, effort, and budget.
- How applying these principles can improve data dissemination and audience understanding across all methodologies.

#### **Session 3B: Innovations in Field Operations**

##### **Approaches to Digital Cleaning**

*Jake Soffronoff, IMLS; Ai Rene Ong, AIR; Matt Sweeney, AIR*

We piloted a first-of-its-kind establishment survey in 2023. Due to the complicated nature of the project's respondent base, previous agency efforts at establishing a suitably robust population frame had proven unsuccessful, and currently-available resources did not cover the field as required by the project. This presentation will discuss the process implemented to refine a commercially-acquired list of potentially-

eligible institutions into the final population frame used for administration of the NMS pilot. This process implemented traditional cleaning methods like data matching with external resources and manual coding to remove ineligible institutions, along with a number of innovative methods, such as using a population AI platform for uncovering business URLs, using web scraping to obtain email addresses, and using a crowdsourcing platform for manual coding and appending the data with contact information. The presentation will detail the challenges faced with each of these platforms, and how those challenges led to each platform earning its niche in IMLS' full process of population frame refinement.

### **Automating Contact Lookup for Telephone Follow-Up**

*John Lombardi, U.S. Census Bureau; Clayton Knappenberger, U.S. Census Bureau; Emily Wiley, U.S. Census Bureau; Jason Bauer, U.S. Census Bureau*

As part of the 2022 Economic Census (EC22), non-respondents are contacted via telephone to remind them to fill out their survey. For a subset of non-response establishments, contact information may be out of date or missing. Historically, clerks have done research to find the correct phone number for an establishment. For EC22, we developed a system to provide phone numbers via the Google API. With an establishment's name and approximate GPS coordinate, we were able to extract phone numbers, QA/QC the phone information, and use record linkage to match results from Google to a sampling frame. This process saved thousands of hours of time and hundreds of thousands of dollars in cost. To validate the accuracy of the phone numbers, we compared results from the Google API to clerk research and achieved similar accuracies between the two. Our work demonstrates that opportunities exist for using data science techniques that can yield significant cost savings for field operations.

### **Exploring New Paradata from Text and Email Contact Attempts in Federal Survey Data Collection**

*Matthew Virgile, U.S. Census Bureau; Laura Hergert, U.S. Census Bureau*

Since 2004, the Census Bureau's Contact History Instrument (CHI) has been used for Computer-Assisted Personal Interview (CAPI) surveys. CHI lets interviewers record information about each contact attempt with potential respondents. Originally, interviewers were asked to record whether a contact attempt was made through a personal visit or a telephone call. In 2022, CHI was redesigned to include options for additional modes of contact, including text messages and e-mails. While survey interviews may only be completed via personal visit or telephone, the addition of contact attempts via text and e-mail may be used in a variety of ways, such as setting

interview appointments or sending information about surveys. Prior to this study, little had been done to explore the reported use of email and text contact attempts in CHI. For our analysis, we use the CHI paradata to calculate the proportion of interviewers that ever attempted contact via these modes, as well as the proportion of sampling units that contacted interviewers via these modes. We also explore the reported reasons for text/email contact, and how text or email usage varies by region within the US. Finally, we compare survey completion rates between sampling units with reported text/email usage and those without text/email usage. This analysis provides a more thorough understanding of how text and email contact attempts are being used and a clearer understanding of their association with survey response.

### **Training our Partners: Lessons Learned from an Integrated Approach to Enhanced Surveillance of Early and Late HIV Diagnosis**

*Bridget Varian, ICF; Dave Roe, ICF*

The Enhanced Surveillance of Persons with Early and Late HIV Diagnosis (Stage 0/3) for the Centers for Disease Control and Prevention (CDC) collects vital data from individuals with recent HIV diagnoses at stage 0 or 3 of infection to help inform prevention, testing and linkage to care strategies. Data collection is a two-stage process that begins with participating health departments recruiting, screening, and registering individuals for the study. To facilitate the first stage, ICF developed a centralized, web-based technical solution that allows health department staff to capture participant information and register participants for participation in the next stage of the study, a multi-mode survey conducted and managed by ICF. The system allows health departments to conduct an immediate transfer to ICF's call center, schedule the participant for a phone interview, or obtain a self-administered web link for the participant. ICF also developed custom reports and the transfer of information back to health departments for participant management and nonresponse follow-up. To maximize the potential of this tool while adequately integrating health departments' activities into the overall data capture process, ICF developed training materials and curricula for health departments to ensure proper operation, detailed information transfer and efficient troubleshooting. ICF also provides technical assistance to health departments and works with them to enhance approaches to recruitment and data collection. This presentation will highlight lessons learned around training, creating practice opportunities for recruiters, and discuss how participant perspectives from the process have helped adapt approaches to interviewer training, data collection, reporting and quality control.

### **Session 3C: Data Science Applications: Text Analysis**

## **Evaluating different Speech to Text Transcription Models**

*Monica Puerto, U.S. Census Bureau; Elizabeth Nichols, U.S. Census Bureau; Curtiss Chapman, U.S. Census Bureau; Micah Harris, U.S. Census Bureau; Brian Sadacca, U.S. Census Bureau; Kevin Zajac, U.S. Census Bureau*

In the last five years, researchers, and practitioners in Natural Language Processing (NLP) have made significant strides to improve the accuracy and abilities of machine learning models. A sector of Natural Language Processing is speech to text which transcribes audio to text through machine learning. There are open-source models available that have shown incredible results. In our presentation, we'll share insights into the use of these models in the context of a survey call-center: the 2020 Census Questionnaire Assistance (CQA) operation. The CQA operation assisted respondents over the telephone with responding to and completing the 2020 Census. CQA team has over 4 million calls involving the caller and the agent. We explore how different transcription models perform (either open sourced or private owned such as Amazon Web Services) and how they compare in non-English languages like Spanish. We will present some pros and cons of each and how to calculate word error rates. If time permits, we can explore how speech to text quality from transcribed text impact downstream machine learning processes.

## **Autocoding federal government survey open responses containing rare words to an ontology of terms**

*Haley Hunter-Zinck, U.S. Census Bureau*

A multiple choice question is a common survey question format and may provide an 'other' option with an open response field asking for more detail in the respondent's own words. Standardized coding of these responses, often to an ontology of terms, is required to incorporate the samples into the larger response dataset. Typical methods for automating the coding procedure (autocoding) take advantage of pretrained models. Since pretrained models are trained on corpora derived, at least in part, from web documents, these models can be inadequate to code responses composed of terms rarely seen on webpages, such as detailed race and ethnicity terms. On a simulated ontology and response dataset, we experiment with how to represent open responses to improve performance of autocoding models for a corpus containing many fabricated words, simulating rare terms in real datasets. We compare different representations and models for predicting associated codes in the ontology. We investigate representations including basic text counts such as bag of words or character n-grams as well as word embeddings, both custom trained and pretrained. Lastly, we compare sentence transformer embedding models with and without fine-

tuning. We conduct fine-tuning using the ontological relationships among assigned codes as a metric of response similarity. We show that sentence transformer representations outperform text count and word embedding representations for predicting codes associated with a response. Additionally, fine-tuning to domain specific data can increase the value of the sentence embedding models for predicting associated codes.

### **Looking for the Needle in the Haystack: Exploring Rare Topics in Large, Unstructured Text Data**

*Curtiss Chapman, U.S. Census Bureau; Elizabeth Nichols, U.S. Census Bureau; Shaun Genter, National Center for Science and Engineering Statistics*

Text mining is a useful process for inspecting patterns in large-scale, unstructured text data. Often, text mining tools are used to segment a dataset into broad categories; however, one may also be interested in whether a text dataset contains niche topics of interest. The current project recommends a method for exploring the presence of known, niche topics within large, unstructured text datasets using data from the Census Questionnaire Assistance (CQA) program. The CQA telephone support line assisted callers by answering questions they had about the census. Of the recorded calls, 91,000 were later transcribed. We created subsets of transcripts potentially related to several niche topics via keyword matching and then created binary labels describing the presence of those topics; questions about young children, Canadian citizens who spend time in the U.S., and people who have died recently. The labels were used to train classifier models to predict the presence of those topics in the larger set of remaining transcripts. Predictions were then validated via clustering models and manual labelling of subsets. Our presentation will explore the success of this method for exploring niche topics in large, unstructured text data and outline recommendations for running such analyses.

### **Opening the Black Box of Self-Employment: Using Machine Learning to Identify Alternative Work Arrangements in the United States**

*Joelle Abramowitz, University of Michigan; Andrew Joung, University of Michigan*

Self-employment consists of an array of work arrangements, yet there exists a dearth of data on such arrangements. To address this gap, this paper uses novel data in the 2003-2019 Panel Study of Income Dynamics (PSID) to examine trends in the prevalence and nature of self-employment work arrangements, transitions across these arrangements, and who works in these arrangements. This work uses machine learning to leverage internal data collected in the PSID on respondent narratives on industry and type of work as well as respondents' employer names. The approach

classifies work arrangements into several categories including informal self-employment, formal self-employment, business ownership, and wage and salaried employment. Findings show disparate trends in the share of workers in different work arrangements and distinct transition patterns across arrangements that would otherwise be masked. Further results suggest that, compared to those in other work arrangements, the informally self-employed are generally worse off. Exploring these questions using these novel data provides unique insights into the changing nature of work relevant to policy considerations across the health, insurance, and poverty dimensions, among others. This study's findings provide greater insight into the nature of alternative work arrangements and permit future work that will more thoroughly consider the causes and implications of differences in work arrangements. This work lays the groundwork for future research examining individuals' work trajectories leading to these roles, movement between different work arrangements, and how these are associated with different levels of economic, physical, and psychological wellbeing over the life course.

### **SOII Autocoder Gender Bias Investigation**

*Gerald Thomas, Bureau of Labor Statistics*

The Bureau of Labor Statistics (BLS) conducts the Survey of Occupational Injuries and Illnesses (SOII) from establishments in private industry and state and local governments, capturing non-fatal workplace injury or illness cases that required at least one day away from work, job transfer, or restriction. Each year, over 200,000 cases are collected through SOII. The SOII Autocoder is a neural network model that assigns codes using the Standard Occupational Classification (SOC) system and the Occupational Injury and Illness Classification System (OIICS) based on the narratives provided in each SOII case. Although no explicit gender indicators are used as inputs to the SOII Autocoder, the SOII case narratives often contain expressions, such as pronouns, that are associated with a specific gender. A pre-trained language model creates a risk that the SOII Autocoder inherits gender bias that potentially exists in the pre-trained language model given that these models are trained using large unfiltered content from the web, which is often far from neutral. Therefore, we evaluate whether SOII Autocoder's predictions vary based on gender expressions that exist in the narrative. To achieve this, we limit the cases to those with narratives containing gender expressions and test for significant differences in the Autocoder's predictions by replacing gender expressions with either all male, all female, or all gender-neutral expressions. This presentation highlights the methodology and results from this gender bias investigation.

4:30 pm - 5:45 pm, April 16

### ***Concurrent Sessions***

#### **Session 4A: Balancing Tradition with Innovation During Unprecedented Times: Lessons from the 2022 Survey of Consumer Finances**

##### **Examining the use of Computer Audio-Recorded Interviewing (CARI) to enhance cultural responsiveness on the Survey of Consumer Finances**

*Heather Sawyer, NORC at the University of Chicago*

Computer Audio-Recorded Interviewing (CARI) can be a powerful tool to monitor data quality in large-scale survey research. CARI has the benefit of capturing interactions between field interviewers and respondents in an unobtrusive way, providing valuable insights about the research process that may otherwise go unnoticed. While CARI has traditionally been used to deter interview falsification, validate interviews, and monitor interviewer adherence to study protocols, this paper examines the potential use of CARI to enhance cultural responsiveness in surveys more broadly, using the 2022 Survey of Consumer Finances (SCF) as an example. Culturally responsive methods address key issues such as how the research is initiated, who is represented in the research, and the relevance of the research to diverse perspectives. I argue the use of CARI presents opportunities to better monitor and assess a study's performance in these key areas. The SCF is a triennial survey sponsored by the Board of Governors of the Federal Reserve and collects detailed financial information from a diverse pool of US households across the full range of socio-economic circumstances. During the 2022 SCF, CARI was used for the first time to augment the study's data quality plan by capturing both interviewer and respondent behavior during key survey questions. In total, over 4,000 completed cases were analyzed by our team of researchers to inform various aspects of data quality. This paper brings together insights from the 2022 CARI analysis with broader discussions regarding diversity, equity, and inclusion in survey research to highlight opportunities to enhance cultural responsiveness on the SCF and beyond.

##### **Field Interviewer Training: Calibrating Delivery Modes**

*Shannon Nelson, NORC at the University of Chicago; Cali Beyer, NORC at the University of Chicago*

The Survey of Consumer Finances (SCF) is a triennial survey sponsored by the Board of Governors of the Federal Reserve (the Board) and collects detailed financial

information from US households across the country. Traditionally, interviewers have all been trained in-person to gain cooperation, identify eligible households, and administer the long and complex instrument. Like so many other surveys, the 2022 SCF had to implement a fully virtual training environment as well as adapt modes during data collection. Even after the pandemic had largely receded people were less likely to engage in-person, both opening their door when initially visited and allowing interviewers in their home to complete interviews. These experiences spurred two main ideas for enhancing future trainings. The first aims to improve trainees learning experiences through use of a tiered, multi-mode training program to better align with data collection realities. The second tailors training according to interviewers' experiences, skillsets, geographic locations, and desires to take on additional specialist roles. In this presentation, we describe what a new model of training could look like for future rounds of the SCF and how other surveys could use these principles to strengthen interviewing training and data collection efforts.

### **Merits of In-Person Field Work: Lessons Learned from the 2022 Survey of Consumer Finances**

*Micah Sjoblom, NORC at the University of Chicago*

The COVID-19 Pandemic has left lasting impacts on large scale in-person data collection across several domains, including shifts in the interviewer workforce, altered mode preferences for conducting outreach, and the continued need for more responsive approaches to engage participants when face-to-face outreach is limited. The 2022 Survey of Consumer Finances (SCF) provides lessons learned in dynamically managing survey operations to introduce phone-first outreach, utilize case prioritization, and promote interviewer skill specialization to meet operational goals. The SCF is a triennial survey sponsored by the Board of Governors of the Federal Reserve (the Board) and collects detailed financial information from a diverse pool of US households across the full range of socio-economic circumstances. The SCF has historically utilized an in-person contacting strategy to guide field outreach. During the 2022 SCF, a series of innovative and responsive processes and procedures were implemented to steer data collection operations in response to more limited in-person fielding. This presentation outlines a set of lessons learned from these efforts designed to both maximize the number of interviews conducted and maintain high levels of data quality.

### **Sharpening our Tools: The Use of R-Indicators in a Challenging Context for the 2022 Survey of Consumer Finances**

*Katherine Archambeau, NORC at the University of Chicago; Kate Bachtell, NORC at the University of Chicago; Steven Pedlow, NORC at the University of Chicago; Cathy Haggerty, NORC at the University of Chicago*

The years following the 2020 COVID-19 pandemic were marked by concerns about the safety of in-person contact and shortages in several labor markets. This unique context heightened the need for adaptive approaches to survey data collection. Survey administrators were especially challenged to navigate stay-at-home orders in many locations and offset shortages in part-time workers available for interviewing activities. In this paper we present the 2022 Survey of Consumer Finances (SCF) as a case study and describe how we used responsiveness indicators ('R-indicators') to monitor and improve the representation of key populations in the survey data. Our 2022 work was built on R-indicator models developed for the 2019 SCF (and presented at FedCASIC in 2021). The 2022 SCF focused on a collaborative effort between the central office staff monitoring data collection and the remote field operations team. When potential demographic shortfalls were flagged by R-indicators, a team was in place to immediately brainstorm possible field interventions that would remediate the deficit and reduce bias, such as special mailings and/or higher incentive offers. The 2022 SCF achieved success in using information obtained from R-indicators to alter data collection strategies and boost response from targeted populations. This paper will discuss the collaboration efforts between office and field staff, as well as elaborate on the interventions used and the successful results obtained from a real-life application of the indicators. In addition to discussion of the 2022 SCF, we will share plans for building on our findings for the 2025 SCF, as we look to apply R-indicators findings to make further advances in the survey's representation.

## **Session 4B: Data Maturity Assessment - Instructional Session**

### **Data Maturity Assessment**

*Troy Brightson, Ph.D., U.S Census Bureau; Crecilla Scott, M.A., U.S Census Bureau; Arnold DelaRosa, U.S Census Bureau; Kyana Beckles M.P.S., U.S Census Bureau; Jack P Leeds, Ph.D., U.S Census Bureau*

The Census Bureau conducted the Data Maturity Assessments to analyze policies, procedures, and operations related to data and data infrastructure. The purpose of the assessment is for an agency to evaluate itself against:

- Documented best practices
- Determine gaps
- And identify areas for improvement

The topics that will be covered in the assessment include:

- Data governance
- Data management
- Data culture
- Data systems
- Data analytics staff skills and capacity, And
- Resource capacity and compliance with law and policy

## **Session 4C: Alternative Data Sources**

### **An Alternative Data Approach to Index Creation**

*Ayme Tomson, BLS*

Our Data Science team assists different program offices in integrating alternative data into existing indexes. Alternative data can be difficult to validate and incorporate into existing program production methodologies. One program's Airline Team has been researching index creation from a third-party data source. This presentation will discuss alternative data source challenges such as data size, data cleaning, and data linkage as they relate to the program index methodologies. The work presented will cover multiple research projects spanning several years of effort to validate the third-party alternative data source for index creation.

### **Can It Work for Employers? Evaluating the Expansion of Administrative Records Use beyond Nonemployer Demographic Statistics (NES-D)**

*Adela Luque, U.S. Census Bureau*

Our ability to identify and research business demographic trends and performance disparities across demographic groups hinges upon the availability of reliable, frequent, and timely business demographics data. In response to declining response rates, and increasing imputation rates and costs, starting with reference year 2017, the Census Bureau began providing nonemployer demographics not through a survey, but a program that leverages administrative and census records to assign demographics to the universe of nonemployer firms: the annual Nonemployer Statistics by Demographics series (NES-D). Given NES-D's success, Census is now evaluating the feasibility of assigning demographic characteristics to U.S. employer businesses using administrative records. In the presentation we will describe the background, methodology, ongoing challenges, current results and next steps of this ongoing effort. The use of administrative records in business demographics statistics should be viewed as a complement to surveys, and a vehicle to unburden

respondents and allow the survey to measure issues that cannot appropriately be captured with administrative records or third-party data; hence strengthening and expanding the capacity of the Federal Statistical Ecosystem.

### **Pulse Surveys as a Tool for Improving Survey Organizations**

*Herman A. Alvarado, SAMHSA; Kathryn Piscopo, SAMHSA; Michael Inguillo, SAMHSA*

In recent years, there has been an increased interest to develop continuous opportunities to learn from employees' experiences in organizations. This approach requires diverse ways to gather information in a short amount of time accomplished by leveraging advancements in technology. Online pulse surveys are one of the preferred tools due to their flexible design, easy implementation, and faster results. For this presentation, we will describe how pulse surveys, or surveys that take quick but broad ongoing assessments, have been effectively used at the Center for Behavioral Health Statistics and Quality (CBHSQ) in the Substance Abuse and Mental Health Services Administration (SAMHSA). As an OMB federally-recognized statistical unit, CBHSQ is a small center responsible for many different data collections and dissemination. These operations require a staff with highly specialized skills, and as such, staff retention is a priority. In the last two years, CBHSQ has used pulse surveys as one of the critical tools to analyze and understand issues associated to staff engagement. The presentation will include examples on the different topics addressed in the pulse surveys, the type of questions, the available tools used for quick analysis, and how results could be translated into best engagement practices for the organization.

### **Using Public Sources to Validate the Record Linkage of the Survey of Doctorate Recipients to Public Inventor Data**

*Ekaterina Levitskaya, Coleridge Initiative; Corey Sparks, Coleridge Initiative; Wan-Ying Chang, National Center for Science and Engineering Statistics within NSF*

Currently, there is no single large-scale dataset with individual links between doctoral recipients and inventors in the United States. A dataset linking these sources could allow the study of the innovation process and output of doctoral recipients. The data sources to conduct such record linkage exist: the Survey of Doctorate Recipients (SDR), conducted by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation, contains restricted use microdata on doctoral recipients and United States Patent and Trademark Office (USPTO) publishes public patent data with inventor names. However, in order to validate and assess the quality of a resulting linked dataset, a gold standard dataset is needed, where the true match status of records is known. In this work, we propose a method of

gathering public resume information from online sources, which contain data on both education and patenting activity. The public resume data are then used to validate inventor status for a representative random sample drawn from the SDR respondents. A wide array of public resume data is explored. We report the feasibility and efficiency of scraping public resume data for constructing a gold standard dataset and discuss challenges related to potential coverage bias and other data quality concerns. This methodology will ultimately help validate linkage results and inform a data quality assessment of linked data.

## **Day 2: Wednesday, April 17**

9:00 am - 10:00 am, April 17

### ***Plenary Session***

#### **Transforming the Labour Force Survey in the UK**

*David Freeman, The Office for National Statistics, UK*

10:15 am - 11:45 am, April 17

### ***Concurrent Sessions***

#### **Session 5A: Data Dissemination**

##### **Modernization Challenges of the Digest of Education Statistics**

*Kristi Donaldson, National Center for Education Statistics*

The Digest of Education Statistics is one of the core statistical products of the National Center for Education Statistics (NCES) and draws on nearly 100 data sources, from NCES and the Department of Education, and other Federal agencies. In addition to serving as a standalone comprehensive statistical reference covering the broad field of education, it underlies many NCES reports and publications, including the Congressionally mandated Condition of Education and the Equity in Education Dashboard. The Digest updates more than 400 statistical tables annually, though production has not kept pace with technological advances. In this presentation, we will discuss our goals to improve the timeliness, utility, and accessibility of Digest through production modernization, including:

1. pain points in existing production procedures, which include inefficiency in table updates and reviews, long lag times between data availability and table publication, and difficulty finding information across the large number tables;

2. the current transitional period as we plan for the future of the Digest, including producing machine-readable tables and beginning to build out table metadata; and
3. our goals with modernization, including standardizing table schemas, redesigning products to be flexible for various downstream uses and to streamline production of multiple workflows, and building toward an enterprise-level analytical data system.

## **High Frequency Economic Data Collection**

*Aaron Keith Savage, U.S. Census Bureau*

The COVID-19 pandemic presented many challenges to businesses in the U.S., with small businesses facing particularly large burdens. It became clear early in the pandemic that there was a need for near real time data that governments, businesses, and researchers could rely on to understand the changing business climate. The U.S. Census Bureau recognized this need and launched the Small Business Pulse Survey (SBPS) from April 2020 - April 2022. This weekly survey provided a snapshot of the obstacles facing small businesses in America during the pandemic. Capitalizing on the success of SBPS, the U.S. Census Bureau launched the Business Trends and Outlook Survey (BTOS) in July 2022 that employs a novel, ongoing data collection strategy to produce near real time data on current economic conditions and trends. BTOS is the only biweekly business tendency survey produced by the federal statistical system, providing unique and detailed data during times of economic or other emergencies. BTOS is representative of all non-farm employer businesses and produces data at state, sector, subsector and employment class levels. Additionally, the recently approved Emergency Economic Information Collection (EEIC) allows access to a vast question bank that can be drawn on quickly and added to the BTOS or another Census economic survey. This positions BTOS uniquely to capture economic data in times of crisis. BTOS also asks supplemental questions on relevant topics such as AI adoption among US businesses. This innovative data source is a critical piece in the U.S. Census Bureau's portfolio, fulfilling its mission to serve as the leading source of valuable data about the nation's people and economy.

## **The Supply Chain Insights Platform: Synthesizing economic data for supply chain analysis**

*Christian Moscardi, U.S. Census Bureau; Kevin Li, U.S. Census Bureau; Krista Chan, U.S. Census Bureau*

Stakeholders across the private and public sectors demand information about the health of the nation's supply chain, and federal agencies are working rapidly to address those demands. Among other efforts to improve the quality of supply chain

data, the U.S. Census Bureau has begun work to consolidate existing data assets into views that are of particular use to stakeholders with an interest in the supply chain. Dubbed the Supply Chain Insights Platform (SCIP), we have developed a tool that lets a user search in plain language for a particular product (e.g. 'baby formula') and uses a combination of machine learning and metadata to surface as much relevant data as possible on that product and its supply chain. During this presentation, we will describe and demo the tool, share data sources we are using, and discuss some of the most salient challenges to combining the Census Bureau's economic data into a holistic picture of the supply chain. We will discuss data gaps we have discovered through feedback from supply chain stakeholders, and underscore the importance of easily accessible, interoperable, scalable, and high-quality supply chain data for those stakeholders' needs.

### **Unlocking Census Data in data.census.gov**

*Maria Valdisera, U.S. Census Bureau*

As the Census Bureau's primary data dissemination tool, data.census.gov provides free access to the nation's most reliable source of demographic, social, and economic data for dozens of geographic areas across the country. These data provide a benchmark for community diversity and can help quantify initiatives to reach nearly any population based on collection methods. This tool allows users to access Census data in a number of unique formats, including tables, maps, Geography Profiles, and Population Pyramids. In this session, attendees will understand the ease of use with data.census.gov, the depth of data available within the site, and observe demonstrations utilizing it. Demonstrations will involve reviewing search methods, download and exporting capabilities, and key mapping features. Due to user feedback, many updates and enhancements have been made to data.census.gov since its implementation in 2020. We will also be highlighting the latest changes and improvements made to the site to help keep contractors, data users, and other federal employees up to date on the many ways data can be accessed through data.census.gov.

### **Using Synthetic Data to Reduce Disclosure Risks in Municipal Health Surveys**

*Wen Qin Deng, New York City Department of Health and Mental Hygiene; Tashema Bholanath, NYC DOHMH; Stephen Immerwahr, NYC DOHMH; Jingchen Hu, Vassar College*

Releasing public-use micro-level data files from health surveys holds immense value for science, health policy, and government accountability. However, even after removing personally identifying information, the confidentiality of survey respondents

may still be compromised through combinations of variables in these datasets. Given the increasing availability of other public data sources and technological advancements, rigorous and systematic methods are needed to assess and mitigate disclosure risk before the public release of micro-level data. Using population estimates through a combination of key variables, we identified high-risk observations within a large New York City population-representative health surveillance survey. We compared three different solutions to mitigate the risk of re-identification; suppression, synthesis using Classification and Regression Trees, and synthesis via Bayesian models and assess their impact on both risk and loss of utility of the resulting protected data. We measured the utility using the overlap of the 95% confidence intervals of health estimates and their mean squared errors from the confidential and protected datasets. While both synthesis methods resulted in slightly higher disclosure risks compared to the suppression method, the synthetic datasets preserved a higher level of utility. Synthesis via Bayesian models was chosen due to its satisfactory balance between risk reduction and data utility preservation. We provide guidance in determining key variables and risky observations, propose solutions to avoid over-protecting and potentially obscuring estimates for underserved and vulnerable groups, and seek to share our experiences with data curators in advancing disclosure risk controls and data sharing in public health.

## **Session 5B: Respondent Centered Design 2**

### **Use of Qualitative Message Boards with Sensitive Questions**

*Rachel Walker-Kulzick*

Our presentation will provide an overview of the uses of qualitative message boards to collect data on sensitive topics. As part of a project looking at sexual harassment and discrimination in the science and engineering (S&E) enterprise, we had planned to conduct virtual focus groups. However due to limitations that emerged throughout the research project, we transitioned our data collection approach to an online message board platform. We used qualitative message boards to collect data on workplace culture and climate, including perceptions of sexual harassment and other forms of workplace harassment. The period of study was October-November 2022. Data were collected from a total of 177 students and professionals in the S&E enterprise. We also specifically focused on recruiting sexual and gender minority (SGM) students and professionals in order to understand their unique workplace experiences. Due to the sensitive nature of the subject matter, our presentation focuses on the methods we employed to engage participants in the message board environment. This includes the development of our message board protocol and roles of message board moderators.

While we leveraged some of the techniques, we would typically use in focus groups, there are limitations to these methods in the message board environment. As such, we describe how we adapted our data collection approach to gather crucial information, particularly as the data collected inform the development of strategies to measure sexual harassment and similar constructs among S&E students and professionals.

### **Improving Efficiency and Data Quality of an In-mover Interview**

*Kimberly Collins, U.S. Census Bureau; Amy Fischer, U.S. Census Bureau*

During a decennial census, it is critical to count the population where they were living or staying on the reference date, April 1 (aka Census Day). When someone moves to a different housing unit after Census Day, it can be difficult to understand which address to use when responding to a census questionnaire, especially if an enumerator asks about multiple addresses. The Field Data Capture (FDC) application, the instrument used in Decennial field operations, had a unique interview path to enumerate in-movers, who, by definition, resided at the case address on the day of the interview but resided elsewhere on Census Day. When an in-mover was identified during a Nonresponse Followup (NRFU) visit in 2020, the enumerator attempted to collect demographic data for the Census Day inhabitants of the case address. Additionally, if the in-mover did not already complete a questionnaire for their Census Day address, the census taker added the address (aka in-field add) to the system and collected the demographic information for said address as of Census Day. This procedure differs from a survey interview, where the census taker would update the Household Roster on file and attempt to interview the in-mover(s), but they would not attempt to collect any data on the person(s) who lived at the address on the reference date. The Decennial in-mover interview is more complex because it entailed asking the respondent about residents at two different locations: the case address and their Census Day address. It also includes an in-field add submission. Although the amount of in-mover adds comprised less than two percent of the entire NRFU workload, enumerators felt this interview path was more complicated than most NRFU interviews because it did not flow smoothly (Gibb, S. et al. 2023). In-movers may be highly mobile or more likely to rent their homes, which can result in being undercounted in the census (Stempowski, D. 2023). The complexity of the in-mover interview, with its multiple components, means in-movers can also be hard to persuade (Stempowski, D. 2023). As the Census Bureau considers improvements for the 2030 Census, it is important to ensure that the in-mover interview process is clear. To reach this goal, we are streamlining the automated in-mover interview path and the training design to make the enumerator's job easier in navigating the interview. A more streamlined path

of questions and training improvements designed to increase knowledge of in-movers will ensure a positive, less burdensome interaction for the respondent, and thereby improve coverage and data quality.

## **Applying a Panel Member-First Approach on KnowledgePanel**

*Nick Bertoni, Ipsos*

This year marks the 25th anniversary of KnowledgePanel, a platform born in 1999 that has undergone numerous transformative changes since its inception. The platform has seen significant advancements in areas like recruitment, sampling, weighting, and surveying, creating a highly accurate and trusted online, probability-based data collection vehicle over the last two decades. Federal agencies like the CDC, NIH, FDA and others have relied upon KnowledgePanel for their data collection needs.

Benchmarking against federal point estimates is one common way that researchers can evaluate and quantify the quality of online probability panels. This is an area where KnowledgePanel has excelled. Beyond this tangible measure is an intriguing question: Are there other management strategies that can directly enhance the quality of data collection, even if they are harder to define or measure? In 2023, the Ipsos KnowledgePanel team set out to rethink panel management by adopting a 'panel member-first' design. Being panel-member first means being intentional about managing the panel member experience in a way that promotes and improves member engagement. The tone, cadence, and overall content of messaging was overhauled in an attempt to increase engagement. Some extra rewards along the way can also be a pleasant surprise. The impact of this approach was clearly seen within the first year of implementation, as this presentation will demonstrate.

KnowledgePanel saw substantial improvements in panel retention and completion rates in 2023. Changes in methodology, messaging, and incentivization will be presented to demonstrate how treating panel members better can directly lead to better panel performance. These lessons learned can be used to inform federal survey efforts.

## **Updating the MCBS Questionnaire to Collect Sexual Orientation and Gender Identity**

*Melissa Heim Viox, NORC; Marisa Wishart, NORC; Hannah Murrow, NORC; Andrea Mayfield, NORC*

In January 2021, the White House released Executive Order 13988 on Preventing and Combating Discrimination on the Basis of Gender Identity or Sexual Orientation, which called upon agencies to identify existing and new policies to promote equal treatment under the law and ensure that all persons can access healthcare and other essential

services without being subjected to sex discrimination. In response, the Medicare Current Beneficiary Survey (MCBS), a continuous survey of a nationally representative sample of the Medicare population, introduced socio-demographic items sourced from the National Health Interview Survey and the U.S. Census Bureau's Household Pulse Survey about sexual orientation and gender identity (SOGI) to expand measures of health equity in 2023. Certain aspects of the MCBS protocol, such as the use of proxy respondents for over 10% of interviews, the longitudinal design, and the target population of primarily older adults, introduce manageable challenges for adding SOGI measures into the existing questionnaire. This presentation will describe methods used to incorporate SOGI into the questionnaire, including item selection, placement within the questionnaire, impact on item universe for sex-specific health status questions, and consideration for proxy interviews. We will also present lessons learned from the first complete round of SOGI data collection. With thoughtful planning, SOGI items can be incorporated into a complex survey protocol like the MCBS and administered successfully with the Medicare population.

## **Session 5C: Projects in progress: Machine learning for interview and survey data analysis**

### **A Semi-Automated Nonresponse Detector (SANDS) model for open-response data**

*Kristen Cibelli Hibben, National Center for Health Statistics; Zachary Smith, National Center for Health Statistics; Valerie Ryan, National Center for Health Statistics; Ben Rogers, National Center for Health Statistics; Paul Scanlon, National Center for Health Statistics; Travis Hoppe, National Center for Health Statistics*

Data quality presents a key challenge for open-text responses which are more prone to item nonresponse, inadequate, or irrelevant responses, particularly in online self-administered surveys. Fortunately, the rapidly changing technology landscape has given rise to more approaches better suited to managing and processing open-text data. Harnessing recent technological advancements in combination with targeted-human coding, the National Center for Health Statistics (NCHS) developed the Semi-Automated Nonresponse Detector for Surveys (SANDS) model to help researchers detect nonresponse in open-ended survey text. This model uses state-of-the-art natural language processing with pre-trained large language models (LLMs). SANDS has previously demonstrated high sensitivity and specificity in distinguishing between valid responses and non-codable responses in longer open-ended responses, particularly those related to health topics. In this presentation, we extend prior

evaluation efforts presented previously, including new topics such as gender, sexual education, and discrimination. We also evaluate new types of open-ended text, including other (specify) responses and demonstrate that SANDS continues to perform well in these contexts. We compare model output against human-coded sources of truth or hand reviewed random samples and use sensitivity and specificity calculations to quantify model performance. Finally, we present updated information about model access and guidelines and tips for best practice. Data are from the NCHS Research and Development Survey (RANDS) Rounds 4, 6, 7 & 8 and RANDS During COVID-19 Rounds 1-3. NORC collected the data between 2020 and 2023 using a probability-based panel representative of the US adult English-speaking non-institutionalized population.

### **Applying Machine-Learning to Interviewer Monitoring and Question Assessment**

*Hanyu Sun, Westat; Ting Yan, Westat; Anil Battalahalli, Westat*

Survey organizations has long used Computer Assisted Recorded Interviewing (CARI) as a quality control tool. 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 process tends to be labor intensive and time consuming. To improve the effectiveness and efficiency of using CARI to monitor interviewer performance and assess survey questions, we developed a pipeline that heavily draws on the use of machine learning to process audio recorded interviews. In the presentation, we will describe the various components and outcome metrics of the pipeline, how the pipeline can be used to detect potential interviewer falsification, identify undesirable interviewer behaviors, and detect survey questions at a higher risk of poor performance, and how to incorporate the pipeline to the conventional quality control process.

### **Automated Identification of Unit-Level Interviewer Falsification using Machine Learning**

*Jerry Timbrook, RTI International; Kirsty Weitzel, RTI International; Peter Baumgartner, RTI International*

Unit-level falsification of survey interviews (curbstoning) may introduce bias into studies. Verifying that interviewers complete genuine interviews is thus crucial for quality assurance of interviewer-administered surveys, and is often achieved by listening to interview recordings for two voices: interviewer and respondent. Yet manual review of all study interviews can be cost and time prohibitive, so surveyors often review a random subsample (e.g., 10%) instead. Advances in machine learning

(ML) offer new possibilities for automatic detection of interview falsification, potentially saving time and money. Previous ML approaches to this problem such as speaker diarization (e.g., using spectral clustering to identify the number of voices in a recording) show high error in a lab setting and are agnostic to the expected number of speakers; yet interviews are expected to have at least two voices: interviewer and respondent. Other clustering algorithms (e.g., k-means, agglomerative) might yield higher accuracy when combined with a test of the assumption that each recording has two voices. Yet this approach is uninvestigated in a survey context. We test this new approach using 711 recordings from an actual telephone survey (i.e., not lab recordings). For each recording, we apply voice activity detection (determining whether parts of the recording contain a human voice), calculate embeddings, use multiple algorithms to cluster these embeddings assuming that there are two unique voices present, and then calculate the cosine similarity between the clusters to determine whether a two-voice assumption is supported. We will review results of our ongoing work, identify methodological issues for audience discussion, and explore implications for interviewer-administered surveys.

### **Measuring Response Latency Using Text Analysis of Interview Transcripts from the Consumer Expenditure Survey**

*Victoria R. Narine, Bureau of Labor Statistics; Erica Yu, Bureau of Labor Statistics; Brett McBride, Bureau of Labor Statistics*

Traditional analyses of survey data can provide valuable information about respondent behavior but does not reveal information about the factors that affect the response process, including the role of the interviewer or respondent requests for clarification. Computer Audio-Recorded Interviewing (CARI) allows researchers to gain insights into interviewer-respondent interactions and listen to the response process. Using transcripts from CARI, this project will explore how major changes to scripted question reading in the Consumer Expenditure Survey, which provides data on expenditures, income, and demographics of consumers residing in the U.S., affect response latency, or the time it takes a respondent to answer a question. In this research, response latency is a proxy for respondent burden. Major changes are defined as changes to question wording that are likely to affect how the respondent would answer the question. Pauses or hesitations in speech (i.e., verbal disfluencies) and respondent remarks such as requests for clarification and think-aloud reflections will be assessed using text analysis. Because this research is ongoing, the presentation will share selected preliminary findings and discuss the challenges of working with unstructured interview data, including preparing it for text analysis, transcription errors, speaker interruptions, and other intricacies of human dialogue. We will also share lessons

learned for accessing and working with data stored in another agency's cloud and using data science techniques in survey methodology research. Lastly, presenters will identify methodological issues for discussion with the audience.

### **Research program investigating use of machine learning for Census Bureau contact center operations**

*Kevin Zajac, U.S. Census Bureau; Elizabeth Nichols, U.S. Census Bureau*

The Census Bureau is researching ways that machine learning methods could enhance the efficiency and effectiveness of our contact centers, while making the experience more satisfactory for both our callers and live agents. Our current contact center operations typically employ an Interactive Voice Response (IVR) system and/or live agents who use scripts to assist them in answering caller questions. For our IVR, the caller listens to predefined options and selects the appropriate menu item from their phone keypad to hear the answer to their question. During the 2020 Census, the contact centers were handled by a contractor due to the large nature of the operation. The contact center solution used both an IVR system and live agents, and calls were taken in English and 12 other non-English languages. No machine learning techniques were used in that operation, but millions of calls were recorded with caller permission. Using these recorded calls, the Census Bureau is currently investigating whether machine learning can improve the efficiency and effectiveness of contact center operations in several areas, such as improving the IVR so that it can answer more caller questions automatically, reducing call handle time for live agent calls, automating some manual contact center processes such as quality monitoring, and identifying new topics or improving existing topics in the operational scripts read to callers. This presentation will focus on ongoing research and some of the challenges we have faced in answering the research questions with machine learning, including error rates in transcription models, and creating or using labeled datasets for English and non-English languages. We will also discuss the reliance on contractor knowledge and expertise for this project.

1:00 pm - 2:30 pm, April 17

### ***Concurrent Sessions***

#### **Session 6A: Tools for Data Processing**

## **Heartbeatpy: Enabling respondent-level survey analysis in python**

*Kirsty Weitzel, RTI International; Peter Baumgartner, RTI International*

Heartbeatpy is a python package developed at RTI International based on the heartbeat analysis methodology that was awarded the 2021 Wiley Award for Excellence in Survey Research developed by Christopher Patton and Justin Purl. Heartbeat analysis enables survey researchers to focus on variation within an individual's responses by standardizing individual responses based on their own mean and standard deviation creating 'upvotes' and 'downvotes'. This enables researchers to understand which questions illicit the strongest opinions across a survey. Heartbeatpy enables the use of python for survey analysis, where previous implementations were written in R. This presentation will present an overview of heartbeat analysis and enable users to get started using the heartbeatpy python package.

## **Maintaining Data Integrity: The Evolution of the Census Bureau's Field Quality Monitoring Program**

*Elizabeth M. Mahoney, U.S. Census Bureau; Mary C. Davis, U.S. Census Bureau; Richard A. Denby, U.S. Census Bureau; Scott W. Glendye, U.S. Census Bureau; Laura B. Hergert, U.S. Census Bureau; Rachel Huang, U.S. Census Bureau; Sadaf Rohani, U.S. Census Bureau; Rafael E. Puello, U.S. Census Bureau*

As data collectors, we bear the crucial responsibility of ensuring the accuracy and quality of our data. This intricate task entails monitoring the work of interviewers across numerous surveys, meticulously verifying adherence to established procedures. To address this challenge, the Field Quality Monitoring (FQM) Program was launched in 2021, implementing a near real-time monitoring system utilizing paradata to detect, mitigate, and rectify potential data quality issues.

Since its inception, the FQM Program has undergone a remarkable transformation. Initially designed to monitor a single survey, it has evolved into a dynamic cross-survey monitoring system, equipped with a diverse array of innovative tools to enhance its effectiveness. This presentation delves into the progress of the FQM program, introducing novel metrics developed for outlier detection and exploring the accompanying investigative tools. Furthermore, it evaluates the effectiveness of these metrics in swiftly identifying threats to data quality.

Moving forward, the FQM program envisions an ambitious future, encompassing the development of advanced predictive models to proactively anticipate and avert data quality issues. Additionally, it aims to integrate machine learning algorithms to automate outlier detection and streamline the investigative process. By embracing

these advancements, the FQM program will continue to safeguard the integrity of the data collected, ensuring its reliability and trustworthiness for generations to come.

## **Geographic Update Partnership Software (GUPS) Web - Streamlining Geographic Data Collection at the U.S. Census Bureau**

*Maria Panaccione, US Census Bureau*

The Census Bureau is modernizing the collection of data in geographic partnership programs with the latest version of the Geographic Update Partnership Software (GUPS) Web, which continues to utilize a flexible and cost-effective open source architecture but now takes full advantage of the scalability, flexibility, and security of a containerized cloud solution. GUPS Web evolves earlier versions of GUPS into a web application that will further streamline data collection for geographic partnership programs and enable new functionality not possible in a standalone application such as real time sharing of data with partners. This session will discuss how GUPS Web uses open source architecture in the first cloud native web application developed and deployed by the Census Bureau to provide license-free GIS tools that incorporate quality checks, hundreds of business rules, and custom topology, which standardizes the data update process for 40,000 tribal, state, and local governments across the nation.

## **Leveraging Large Language Models (LLMs) in Census Data Processing and Analysis**

*Yezzi Angi Lee, Reveal Global Consulting; Taylor Wilson, Reveal Global Consulting; Hector Ferronato, Reveal Global Consulting; Cameron Milne, Reveal Global Consulting*

This presentation will outline a collection of proof-of-concepts for implementing Large Language Models (LLMs) and other natural language techniques to enhance the efficiency and accuracy of Census data processing. These tools, some of which are already in development at Census, include Dr. NAICS, a classification and analyst support tool, language translation support for historically undercounted populations, and a code conversion tool for SAS to Python. Our primary objective is to streamline manual processes and to assist analysts and developers to reallocate their effort away from time-intensive tasks. This initiative aligns with the Census Bureau's movement towards statistical modernization within the Economic Statistical Methods Division, influencing various aspects across the organization. We will discuss the development process, including key challenges, strategies, and solutions. We will also highlight AI governance practices in federal program offices and how they can be adapted to work more closely with these new methods and technologies. Lastly, we will discuss what

kind of technology stack was necessary to implement to effectively integrate these tools into a data mature organization and how this might differ in organizations at different parts of their journey. Our findings demonstrate the immense potential of LLMs and other natural language tools to modernize processes in federal programs, achieving enhanced operational efficiency.

## **Session 6B: Reducing Non-Response**

### **Equipping State Agency Staff to Analyze Nonresponse Bias in Federal Survey Programs**

*Benjamin Schneider, Westat; Tamara Nimkoff, Westat; Anthony Fucci, Westat; Andy Cruse, Westat*

As response rates continue to decline in most large-scale federal surveys, nonresponse bias analysis has become an increasingly important part of the planning, weighting, and analysis components of federal statistical programs. While federal policy increasingly mandates the analysis of potential nonresponse bias in data collection programs, state-level staff responsible for collecting and analyzing the data often lack training and tools in how to conduct such analyses. To address these challenges in a large, distributed data collection program for the Department of Education, we developed a free, open-source interactive application and accompanying statistical software package. We also developed a set of educational resources and training materials supplemented by direct technical assistance, to empower state agency staff with varied backgrounds and survey contexts to adopt current best methods for nonresponse bias analysis. In this presentation, we provide an overview of the application and the underlying analysis methods it implements. We also provide reflections and recommendations on training users of the application in how to conduct nonresponse bias analyses appropriate to their particular survey context.

### **Expanding interview window times and effects on trends for survey outcomes**

*Kristen Brown, RTI; Lauren K. Warren, RTI; Jennifer Hoenig, SAMHSA; Tenecia Smith SAMHSA*

The National Survey on Drug Use and Health (NSDUH) was conducted only in person until March 2020 when data collection was halted due to the COVID-19 pandemic. Data collection resumed in October 2020 with a web mode option added to in-person data collection in areas in which COVID-19 infection rates were low. Since then, NSDUH has been conducted via multimode data collection, utilizing web and in-

person modes. In 2023, the timeframe for completing the web interview, if the interview was started but not completed on the same occasion, was expanded to increase respondent participation. From 2021 to 2022, respondents who began the interview on the web had to complete the interview within 24 hours of initiating the interview or start the interview all over again. Beginning in 2023, respondents were given 4 weeks to complete the interview before having to start the interview all over again.

SAMHSA is interested in how this change in data collection protocol impacted the resulting data. Analyses suggest that the respondents who completed the interview in the later portion of the window (i.e., between 24 hours and 4 weeks, labeled as 'late responders' for this analysis) report higher rates of substance use compared with those respondents who completed the interview during the first 24 hours of initiating the interview (labeled as 'early responders'). This presentation will cover demographic and behavioral differences between the samples of web respondents who completed the interview in different portions of the web interview window, as well as adjusted comparisons of substance use and mental health outcomes obtained by controlling for the demographic characteristics. We will also discuss the extent to which this change is expected to impact annual survey results.

### **Sometimes (or Late) is Not the Same as Never: Health Survey Panelist Response**

*Emily Wichmann, The University of North Carolina at Chapel Hill; Tashma Bholonath, New York City Department of Health and Mental Hygiene; Stephen Immerwahr, New York City Department of Health and Mental Hygiene*

As panel surveys become more widely used for measuring health and related, more information on health-related nonresponse bias is needed. We analyzed 24 surveys conducted over a 34-month period via the NYC Health Panel. We sought to identify health attributes of post-enrollment participation; detect possible differences between early and late respondents; and investigate the potential of late respondents as proxies for non-responding panelists. Panelists' propensity to respond to survey invitations was characterized as 'Never', 'Sometimes' (completed fewer than half of the invited surveys), and 'Often' (completed at least half of the invitations). For response timing, respondents were divided into 'Early' (by the midpoint of the field period,) and 'Late'. Available health measures, recorded during the panel enrollment survey, were self-rated general health, days of poor physical health, days of poor mental health, and days of impaired activities in the past 30 days. We used Chi-Square tests to examine associations between each type of response and demographic and health measures and tested statistical significance at two-sided p-value < 0.05. The 'Never'

and 'Sometimes' respondents shared similar demographic attributes, but the 'Sometimes' respondents reported worse health outcomes than either 'Never' or 'Often' respondents. We also found that 'Early' respondents generally had better self-reported health than 'Late' respondents, irrespective of whether the total field period was two weeks or four weeks. Finally, 'Late' respondents typically differed from 'Never' respondents, particularly in terms of household size and education level. These findings hold potential for better understanding non-response bias in health-related panel surveys.

### **Survey of Unpaid Time: A first non-response bias study since the pandemic**

*Cindy Ubartas, Statistics Canada; Geneviève Vézina, Statistics Canada; Marie-Hélène Miville; Statistics Canada*

In the context of declining response rates for social surveys, our organization is concerned by the presence of potential bias. Therefore, non-response bias studies are now considered. An innovative and agile framework with accelerated timelines for development has been put in place in 2023. It assumes the conception of a shorter questionnaire to compare respondents' characteristics and responses to a few key questions from the main survey. It also uses a different collection strategy with a mixed mode approach. A pilot project was done in the fall 2023 with the Time Use Survey (General Social Survey). Its main objective was to evaluate the feasibility in terms of operations of such a study and assess what could be gained from it. This presentation will describe the strategy used and provide a summary of the results and lessons learned.

## **Session 6C: Data Science Applications: Large Language Models**

### **Down the LLM Rabbit Hole: Evaluating the Performance of Different Large Language Models in Coding Open-Ended Questions**

*Michael Link, Ipsos; Nick Bertoni, Ipsos*

Large language models (LLMs), advanced AI systems, use deep learning to process and analyze text data for tasks like text classification and sentiment analysis. However, LLMs can be a 'black box' due to their varying computational programming and training data. They may struggle with inputs different from training data, leading to inaccurate results. Their coherent outputs may lack factual accuracy. We tested a tool; TACATIC -- powered by OpenAI LLMs at Ipsos that auto codes with or without training, categorizing, and coding sentiment from open-ended responses. We compared: (1) human coders to the LLM method; (2) model usage changes over time;

(3) input language changes. Results showed LLMs can extract useful information from numerous survey responses, potentially saving time and costs. The TACTIC tool can automate this process, reducing time and simplifying data flows. However, variability in more specific quantitative assessments was clear, as conclusions could vary based on coded response counts.

### **Enhancing Big Data Classification with Zero-Shot Learning Models**

*Francis Smart, Censeo Consulting Group; Issa Abboud, General Services Administration*

This study investigates the application of Large Language Model (LLM) Zero-Shot classifiers for categorizing vast quantities of federal transactional sales data. We demonstrate how complexity can be significantly reduced by strategically organizing text-based transaction descriptions according to their frequency and monetary value. This enables us to transform a daunting task into a more manageable one for 99% of the data's total value. Our examination of various LLMs revealed that the Bart-Large model, pre-trained on the MultiNLI dataset, outperformed others. Though results were satisfactory, the model's effectiveness aligned with human judgment at 69% with a confidence threshold of 0.4, which improved to 82% at a 0.6 threshold. Certain categories presented ambiguity, affecting both human and LLM accuracy. Notably, the model excelled in identifying 'professional labor hours', a category of key interest, achieving 83% accuracy at the lower weight (.4) threshold and 90% at the higher one (.6). With the 0.4 threshold, the leading LLM correctly identified 83% of the True Positives, keeping the false negative rate acceptably low, at most 10% of the true values. This false negative rate decreased to merely 1.46% at the 0.6 threshold. These findings underscore the potential of LLM Zero-Shot classifiers to efficiently handle and analyze enormous datasets with high precision, particularly in categories where the definitions of the category are well understood.

### **Using Large Language Models for Other-Specify Coding**

*Wei Huang Wong, NORC*

Some questionnaire items allow respondents to select an "Other" option and specify more information about what "Other" means through an open text response. In 'other-specify'(OS) coding, also known as backcoding, these open text responses are coded back into the existing codeframe or as new options. OS-coding is time- and labor-intensive. Survey assistants must code responses, adjudicate disagreements, and quality check the work. Large language models (LLMs) excel at natural language processing tasks and could help reduce the time and labor burden of OS-coding. We present an experiment using LLMs to backcode open text responses from the 2019

National Survey of Early Care and Education (NSECE). The NSECE is a nationally representative survey of ECE supply and demand, sponsored by the Office of Planning, Research, and Evaluation in the Administration for Children and Families. Our experiment evaluated model outputs against human-generated codes. We utilized open-source LLMs deployed locally on consumer-grade hardware. We show results from various models and prompt formulations. We discuss implications of using LLMs to reduce the burden of OS-coding in survey research.

## **Enhancing Data Analysis with Large Language Models: Revolutionizing Data Discovery Through Semantic Search, Summarization, and Captioning**

*Irina Belyaeva, U.S. Census Bureau, Center Enterprise Dissemination and Consumer Innovation*

The emergence of large language models (LLMs) has introduced a transformative paradigm in information discovery within structured data, significantly impacting fields such as federal statistics and the broader scope of national federal surveys. This paradigm leverages the capabilities of LLMs in semantic search, content summarization, and semantic captioning, directly addressing the unique challenges faced in the analysis and interpretation of survey data and federal statistics. By transcending traditional keyword-based approaches, LLMs offer a nuanced understanding of data, crucial for effective semantic search and interpretation in federal statistical analysis. This research underscores the relevance of LLMs to the analysis of the survey data by highlighting their intrinsic strengths in grasping the essence and context of structured data—a cornerstone in advancing federal statistical methods. Through unraveling the semantic relationships inherent in survey data, LLMs facilitate a shift from conventional searching to a more insightful interpretation of structured information, promising to revolutionize data exploration and enhance decision-making in federal statistics. Moreover, LLMs' prowess in content summarization and semantic captioning enables them to distill complex datasets into concise summaries and generate meaningful descriptions for individual data points, illuminating their underlying context and connections. This not only empowers data analysts to quickly grasp the core essence of extensive datasets but also aids in communicating statistical findings more effectively. While delving into the theoretical foundations of employing LLMs for these transformative tasks, this research also acknowledges the challenges such as bias mitigation and interpretability that need addressing. In conclusion, this study lays the groundwork for future endeavors aimed at harnessing the full potential of LLMs in the domain of data-driven information discovery, particularly within the context of federal statistics, thereby contributing to more informed policy-making and enhanced public understanding.

## **Using Large Language Models to Support Survey Coding**

*Rob Chew, RTI International; John Bollenbacher, RTI International; Michael Wenger, RTI International; Jessica Speer, RTI International; Annice Kim, RTI International*

Survey coding is a widely used method for analyzing open-ended responses. While useful, coding is often burdensome and time consuming since it requires researchers to read, interpret, and reliably categorize a large body of unstructured text. Large language models (LLMs), like ChatGPT, are a class of quickly evolving AI tools that can perform a range of natural language processing and reasoning tasks. In this study, we explore the use of LLMs to reduce the time it takes for coding while retaining the flexibility of a traditional qualitative approaches, such as content analysis. We outline the proposed approach, called LLM-assisted content analysis (LACA) on four publicly available data sets to assess the broader question of how well GPT-3.5 performs across a range of deductive coding tasks. Overall, we find that GPT-3.5 can often perform deductive coding at levels of agreement comparable to human coders. Additionally, we demonstrate that LACA can help refine prompts for deductive coding, identify codes for which an LLM is randomly guessing, and help assess when to use LLMs vs. human coders for coding. We conclude with several implications for future practice of survey coding and related research methods.

2:45 pm - 4:15 pm, April 17

### ***Concurrent Sessions***

## **Session 7A: Video Interviewing in Quantitative Surveys, Today and Tomorrow - Roundtable Session**

### **Video Interviewing in Quantitative Surveys, Today and Tomorrow**

*Hanyu Sun, Westat (U.S.); Rick Dulaney, Westat (U.S.); Heidi Guyer, RTI International (U.S.); Andrew Hupp, University of Michigan (U.S.); Dina Neiger, Social Research Centre (Australia)*

In the past 4 years many research organizations around the world have added video as a data collection mode, prompted by the social distancing invoked by the COVID-19 pandemic and bolstered by experimental work that began in the early 2000s. The pandemic also motivated many more potential survey respondents to become comfortable with video for synchronous communication. This kind of convergence between a survey mode and a respondent's daily life is rare, and welcome. How is the

mode faring: This roundtable brings together methodologists from four survey organizations to share their experiences. Some researchers in this panel have extended our understanding of the mode through new experiments and lab work; others have taken early steps into survey production environments. All will discuss the future of this new data collection tool for quantitative surveys. Each participant will provide a 15-minute overview of their experience and some thoughts about the mode's future in official government surveys. Then in the next 10-15 minutes the moderator will pose some questions to the participants, such as:

- What has surprised you most about video interviewing?
- What are the main advantages and disadvantages?
- Some have speculated that video as a data collection mode falls somewhere between face to face and telephone. Can you shed any light on this?
- What practical advice can you give to a survey organization that is poised to begin video interviewing?

The final 15-20 minutes of the will feature an open discussion with the audience, giving everyone a chance to ask questions and offer remarks.

## **Session 7B: Record Linkage**

### **Some Novel Record Linkage Algorithms**

*Sanguthevar Rajasekaran, University of Connecticut*

In this presentation we plan to summarize some of the novel algorithms that we have recently proposed in the context of record linkage. Blocking is a technique that is typically used to speed up record linkage algorithms. Recently, we have introduced a novel algorithm for blocking called SuperBlocking. We have created novel record linkage algorithms that employ SuperBlocking. Experimental comparisons reveal that our algorithms outperform state-of-the-art algorithms for record linkage. We have also developed parallel versions of our record linkage algorithms and they obtain close to linear speedups. We will provide details on these algorithms in this presentation. We can think of each record as a string of characters. Numerous distance metrics can be found in the literature for strings. The performance of a record linkage algorithm might depend on the distance metric used. Some popular ones are: edit distance (also known as the Levenshtein distance), q-gram distance, Hausdorff distance, etc. Jaro is one such popular distance metric that is being widely used for applications such as record linkage. The best-known prior algorithms for computing the Jaro distance between two strings took quadratic time. Recently, we have presented a linear time algorithm for Jaro distance computation. We will summarize this algorithm also in this presentation.

## **OmniTrustAI: Revolutionizing AI with the Train Once, Apply Anywhere Framework**

*Xiaowei Xu, University of Arkansas, Little Rock; Xingqiao Wang, Vivekanandan Gunasekaran*

The development of Artificial Intelligence has led to sophisticated language models that rival human writing. However, their use in specialized areas can yield unsafe, biased, or factually incorrect outputs. Our innovative AI framework adopts a 'Train Once, Apply Anywhere' (TOAA) approach, modifying these Foundation models for safer, more robust use across different domains. Our method involves transferring knowledge from a foundational Large Language Model into a Customized Language Model (CLLM). This process significantly reduces the model's size while maintaining its performance, enabling efficient operation on consumer-grade computers. The CLLM offers fast processing, cost-effectiveness, and enhanced accuracy. A key feature of our CLLM is its single-training, multi-domain application capability, contrasting with traditional AI models limited to their training domain. This flexibility marks a significant shift in AI methodologies. We tested our TOAA framework using various foundation models, including GPT-3.5, Dolly, and LLAMA, focusing on entity matching; an essential task for data integrity. Our CLLM, trained on one dataset, excelled across multiple domains, showcasing superior accuracy, a 50-fold increase in speed, and linear cost savings compared to using foundation models. This study confirms the TOAA framework's effectiveness for domain-specific tasks, advancing AI deployment's practicality, safety, and efficiency; an Omni Trust AI.

## **Mitigating LLM Vulnerability**

*Vivek Gunasekaran, UALR; Xiaowei Xu, UALR*

The proliferation of Large Language Models (LLMs) has significantly transformed the landscape of natural language processing, content generation, and information retrieval. However, their widespread adoption raises concerns regarding potential vulnerabilities that can be exploited for malicious purposes. This study provides an in-depth exploration of LLM vulnerability implementation, encompassing a thorough analysis of theoretical foundations, practical implications, and proactive mitigation strategies. The research identifies key factors, such as model architectures, training data, and deployment scenarios, that can introduce inherent weaknesses in LLMs. In addition, these vulnerabilities can be introduced during various phases, such as the design, development, deployment, maintenance, and operations of LLM-based applications. Ongoing monitoring and iterative model updates were also discussed as essential components of a dynamic and adaptive security strategy. In conclusion, this

study provides a comprehensive examination of LLM vulnerability implementation, offering a nuanced understanding of its theoretical foundations, practical implications, and proactive mitigation strategies. By addressing these vulnerabilities, this study contributes to the development of more secure, responsible, and trustworthy LLMs that foster confidence in their applications across various domains.

### **Demo of OMNIMatch and Guidance**

*Xingqiao Wang, UALR; Xiaowei Xu, UALR; Vivek Gunasekaran, UALR*

In the evolving landscape of data management, entity matching stands as a critical yet challenging task. OMNIMatch emerges as a revolutionary solution, harnessing the power of Large Language Models (LLM) to redefine entity matching. This demo introduces OMNIMatch, highlighting its role in simplifying and enhancing the accuracy of entity matching processes. The demonstration will expertly guide viewers through a variety of real-world scenarios, effectively showcasing OMNIMatch's extensive applicability in multiple tasks. It will highlight the tool's proficiency in processing two specific types of datasets: one that is similar to US Census data, and another comprising simulated household data. Each scenario is carefully selected to demonstrate OMNIMatch's exceptional capability in managing these intricate data structures, showcasing its powerful and versatile functionality. Designed for data professionals and business analysts alike, OMNIMatch's applications span across sectors, offering transformative benefits in data quality and insights. This demo invites you to witness firsthand the future of entity matching, showcasing how OMNIMatch stands at the forefront of data management innovation.

### **Large-scale simulated population data for an authentic entity resolution challenge**

*Beatrix Haddock, Institute for Health Metrics and Evaluation, University of Washington; Alix Pletcher, Institute for Health Metrics and Evaluation, University of Washington; Nathaniel Blair-Stahn, Institute for Health Metrics and Evaluation, University of Washington; Os Keyes, Institute for Health Metrics and Evaluation, University of Washington; Matt Kappel, Institute for Health Metrics and Evaluation, University of Washington; Steve Bachmeier, Institute for Health Metrics and Evaluation, University of Washington; Syl Lutze, Institute for Health Metrics and Evaluation, University of Washington; James Albright, Institute for Health Metrics and Evaluation, University of Washington; Alison Bowman, Institute for Health Metrics and Evaluation, University of Washington; Caroline Kinuthia, Institute for Health Metrics and Evaluation, University of Washington; Rajan Mudambi, Institute for Health Metrics and Evaluation, University of Washington; Abraham D. Flaxman, Institute for Health*

*Metrics and Evaluation, University of Washington; Zeb Burke-Conte (pronouns: he/him), Institute for Health Metrics and Evaluation, University of Washington*

Entity resolution (also known as record linkage) is the data science challenge of determining which records correspond to the same real-life entity, such as a person, business, or establishment. The United States Census Bureau regularly performs entity resolution on administrative lists containing hundreds of millions to billions of records. However, these administrative lists contain PII and are highly confidential, preventing those outside the Bureau from understanding the Bureau's entity resolution challenges in detail. In this session, we present pseudopeople, an open-source Python package that generates simulated datasets with hundreds of millions of records, which resemble the administrative lists linked by the Census Bureau. pseudopeople is based on an individual-based microsimulation of the United States population, including dynamics such as migration, mortality, and fertility. pseudopeople users can customize the noise present in the datasets generated. pseudopeople data can be used to create authentic entity resolution tasks for testing new methods or software. We present an example of an entity resolution pipeline emulating the methods used by the Census Bureau, using only freely available open-source software and simulated data.

### **Address Parsing with Active Learning**

*Onais Khan Mohammed, University of Arkansas at Little Rock; John R. Talburt, University of Arkansas at Little Rock; Adeeba Tarannum, University of Arkansas at Little Rock; Abdul Kareem Khan Kashif, University of Arkansas at Little Rock; Salman Khan, University of Arkansas at Little Rock; Khizer Syed, University of Arkansas at Little Rock*

This work describes the research and development of a tool to parse demographic items into a standard set of fields to achieve metadata alignment using an active learning technique based on token pattern mappings augmented by active learning. Input strings are tokenized and then a token mask is created by replacing each token with a single-character code indicating the tokens potential function in the input string. A user-created mapping then directs each token represented in the mask to its correct functional category. Testing has shown the system to be as accurate, and in some cases, more accurate than comparable parsing systems. The primary advantage of this approach over other systems is that it allows a user to easily add a new mapping when an input does not conform to any previously encoded mappings instead of having to reprogram system parsing rules or retrain a supervised parsing machine learning model. These address components are essential for the use of HiPER indices, Boolean rules, and scoring rules, and these rules play a crucial role in the

implementation of various data preparation functions. This includes identifying and separating the street address, city, state, and postal code. These components are then stored in a structured format, allowing them to be easily retrieved and used in various applications.