Federal Committee on Statistical Methodology

2024 Research and Policy Conference

The Relevance, Timeliness, and Integrity of Federal Statistics

October 22-24, 2024

College Park Marriott Hotel and Conference Center
Hyattsville, MD

Preliminary Program

v1.1

Hosted By: COPAFS
Council of Professional Associations on Federal Statistics

Table of Contents

2024 FCSM Conference **1**
Sponsors & Exhibitors **2**
Federal Sponsors 2
Organizational Sponsors 3
Exhibitors 4
2024 FCSM Planning Committee **5**
Federal Committee on Statistical Methodology **6**
COPAFS Mission & Objectives **7**
Keynote Speaker: Jeri Mulrow **8**
Schedule Overview **9**

Abstract Listings for October 22 ..... **12**
Concurrent Sessions A 13
Plenary Session 25
Concurrent Sessions B 26
Concurrent Sessions C 35

Abstract Listings for October 23 ..... **47**
Concurrent Sessions D 48
Concurrent Sessions E 57
Concurrent Sessions F 68
Concurrent Sessions G 78

Abstract Listings for October 24 ..... **90**
Concurrent Sessions H 91
Concurrent Sessions I 102
Concurrent Sessions J 113
Concurrent Sessions K 124

2024 FCSM Research and Policy Conference

The Federal Committee on Statistical Methodology (FCSM) was founded in 1975 by the Statistical and Science Policy Branch (formerly, the Office of Statistical Policy) in the Office of Information and Regulatory Affairs in the Office of Management and Budget (OMB) to assist in carrying out SSP/OMB's role in setting and coordinating statistical policy.

The FCSM serves as a resource for OMB and the federal statistical system to inform decision making on matters of statistical policy and to provide technical assistance and guidance on statistical and methodological issues. The FCSM sponsors regular conferences (in partnership with the Council on Professional Associations on Federal Statistics), hosts seminars and workshops (often in collaboration with partner organizations such as the Washington Statistical Society), writes reports aimed at the federal statistical system and the community, and creates subcommittees and interest groups focused on topics relevant to the federal statistical community. Recent news and resources from FCSM are available at www.fcsm.gov.

The FCSM Research and Policy Conference helps the committee achieve their major goals, which are to:

- Communicate and disseminate information on statistical practice among all federal statistical agencies;
- Recommend the introduction of new methodologies in federal statistical programs to improve data quality; and
- Provide a mechanism for statisticians in different federal agencies to meet and exchange ideas.

The theme for the 2024 FCSM Research and Policy Conference is: *The Relevance, Timeliness, and Integrity of Federal Statistics*. The conference provides a forum for experts and practitioners from around the world to discuss and exchange current methodological knowledge and policy insights about topics of current and critical importance to federal agencies as well as the Federal Statistical System as a whole.

Each day of the conference will offer papers on a wide range of topics relevant to the production, quality and use of federal statistics. Attendees from a range of backgrounds will find sessions of interest, including statistical methods, administrative data, questionnaire design, program evaluation, policy making, and more.

Sessions feature presentations by government, private sector, and academic researchers from multiple countries. All sessions will include an open discussion and some sessions will include a formal discussion. Presentations will be made available on the conference website following the conference.

Federal Sponsors

The 2024 FCSM Research and Policy Conference received generous support from the following federal agencies.

Organizational Sponsors

The 2024 FCSM Research and Policy Conference received generous support from the following organizations.

Exhibitors

Thank you to all our exhibitors! Learn more about these organizations by visiting the exhibit tables in the Main Hall.

NORC at the University of Chicago

AMERICA'S DATAHUB CONSORTIUM

NCES National Center for Education Statistics

MITRE

bea Bureau of Economic Analysis U.S. DEPARTMENT OF COMMERCE

2024 FCSM Research and Policy Conference
Program Committee

Jeffrey Gonzalez, Co-Chair, *Bureau of Labor Statistics*
Jessica Holzberg, Co-Chair, *U.S. Census Bureau*
Jessica Graber, Co-Chair, *National Center for Health Statistics*
Darius Singpurwalla, Co-Chair, *National Center for Science and Engineering Statistics*
David Kashihara, Co-Chair, *Agency for Healthcare Research and Quality*
Paul Schroeder, Co-Chair, *Council of Professional Associations on Federal Statistics*

Melissa Cidade, *Bureau of Justice Statistics*
Gavin Corral, *National Agricultural Statistics Service*
Derrick Dennis, *Internal Revenue Service*
Aneta Erdie, *U.S. Census Bureau*
Shelley Feuer, *National Center for Science and Engineering Statistics*
Ellen Galantucci, *Federal Maritime Commission*
Shelley Hyland, *Bureau of Justice Statistics*
Alyssa Holdren, *Bureau of Economic Analysis*
Yan Liu, *Internal Revenue Service*
Jonathan Mendelson, *Bureau of Labor Statistics*
Rachel Morgan, *Bureau of Justice Statistics*
Jennifer Rhorer, *National Agricultural Statistics Service*
Jordan Riddle, *Bureau of Transportation Statistics*
Joseph Rodhouse, *National Agricultural Statistics Service*
Aneer Rukh-Kamaa, *Office of Homeland Security Statistics*
Sharon M. Stern, *U.S. Census Bureau*
Ian Thomas, *U.S. Census Bureau*
Tywanquila Walker, *Bureau of Labor Statistics*
Julie Weeks, *National Center for Health Statistics*
Doug Williams, *Bureau of Labor Statistics*
Jeramiah Yeksavich, *Energy Information Administration*

Federal Committee on Statistical Methodology

Ann E. Carson, *Bureau of Justice Statistics*
Chris Chapman, *National Center for Education Statistics*
Jennifer Hunter Childs, *U.S. Census Bureau*
Tala Fakhouri, *U.S. Food and Drug Administration*
John Finamore, *National Center for Science and Engineering Statistics*
Ellen Galantucci, *Federal Maritime Commission*
Jeffrey Gonzalez, *Bureau of Labor Statistics*
Michael Hawes, *U.S. Census Bureau*
Travis Hoppe, *National Center for Health Statistics*
Brandon Kopp, *Bureau of Labor Statistics*
Carla Medalia, *U.S. Census Bureau*
Lisa Mirel, *National Center for Science and Engineering Statistics*
Jennifer Nielsen, *National Center for Education Statistics*
Jennifer M. Ortman, *U.S. Census Bureau*
Anne Parker, *Internal Revenue Service*
Jennifer D. Parker, *National Center for Health Statistics*
Mark Prell, *Economic Research Service*
Michael Ratcliffe, *U.S. Census Bureau*
Rolf R. Schmitt, *Bureau of Transportation Statistics*
Robert Siviniski, *Office of Management and Budget*
Matt Spence, *U.S. Census Bureau*
Ed Strocko, *Bureau of Transportation Statistics*
G. David Williamson, *Centers for Disease Control and Prevention*
Linda J. Young, *National Agricultural Statistics Service*

COPAFS

Council of Professional Associations on Federal Statistics

The Council of Professional Associations on Federal Statistics (COPAFS) is devoted to educational activities and preserving the public good represented by federal statistical collections. Since 1980, COPAFS has provided an open dialog between those who use federal statistics in professional contexts and the Federal statistical agencies that produce those statistics for the public good. Supporting organizations include professional associations, businesses, research institutes, and others that help to produce and/or use federal statistics. Our Goal: Advancing Excellence in Federal Statistics.

COPAFS' objectives are to:

- Increase the level and scope of knowledge about developments affecting Federal statistics;
- Encourage discussion within and among professional organizations to respond to important issues in Federal statistics and bring the views of professional associations to bear on decisions affecting Federal statistical programs.

In support of these objectives, COPAFS:

- Obtains information on developments in statistics through discussions with officials, attendance at congressional hearings and meetings of statistical advisory committees, engaging with the broader statistical community, and reviewing recent reports or directives affecting the Federal statistical system;
- Disseminates information and encourages discussion and action on developments in federal statistics through correspondence and presentations at COPAFS and professional association meetings, direct calls for action via email, and announcements on social media; and
- Plans and presents educational programs on uses of statistics in policy formulation, public and private decision- making, research, the distribution of products, and the allocation of resources.

COPAFS helps:

- Professional associations and other organizations obtain and share information about developments affecting federal statistical programs;
- Federal agencies to disseminate information on developments of interest to the professional community and to obtain advice about professional societies' concerns and priorities;
- Congressional offices to clarify issues and questions about the federal statistical system, to plan hearings related to federal statistical programs, and to identify experts to testify; and
- The public to learn more about the federal statistical agencies, to communicate views of data users concerning Federal statistical activities, and to obtain a better understanding of how policy and budget are likely to affect the availability of federal statistics.

Keynote Speaker

Jeri Mulrow, Westat

Tuesday, October 22 11:00AM

Jeri Mulrow is Vice President and Sector Lead, Data Solutions, leading Westat's units of Statistics and Data Science, Large Surveys, Research Operations, and Field Services staff, overseeing the collaboration of these teams' shared processes to promote efficiency and facilitate innovative research solutions. From 2019 to 2023, she served as Director of Statistics and Data Science. Her areas of statistical expertise include sampling and estimation methodologies, survey methodology and design, data quality and process improvement, and data visualization and information dissemination.

Mulrow also served as the Principal Deputy Director for the Bureau of Justice Statistics in the U.S. Department of Justice, providing leadership and oversight on the design, collection, analysis, and dissemination of data on the U.S. justice system. Previously, she was the Deputy Division Director for the National Center for Science and Engineering Statistics at the National Science Foundation where she provided leadership and direction on the design, collection, analysis, and dissemination of data on the U.S. science and engineering enterprise.

Mulrow was recently elected President by the American Statistical Association membership for the year 2026.

2024 FCSM Schedule Overview

Tuesday, October 22

8:00 am – 5:00 pm **Registration Open**

8:00 – 9:00 am **Continental Breakfast,** *Sponsored by Westat*

9:00 – 10:30 am **CONCURRENT SESSIONS A**

A-1. Innovations in Federal Data Management and Evaluation (*Vessey 1*)
A-2. When the Rubber Hits the Road: Integrating Alternative Data Sources (*Vessey 2*)
A-3. Text Tactics: Natural Language Processing in Action (*Patuxent*)
A-4. Disclosure Avoidance and Product Structure in the 2020 Census S-DHC File (*Room 2110*)
A-5. Advancing Survey Methodologies: Innovations in Survey Modernization (*Room 2100*)
A-6. Make Everybody Count: Including Hard to Reach People in Federal Statistics (*Room 0105*)

10:30 - 11:00 am **Coffee Break**

11:00 - 12:15 pm **PLENARY SESSION** *(Chesapeake Ballroom)*

- Welcoming Remarks
- Jeri Mulrow, Westat

12:15 - 2:00 pm **Lunch on Your Own**

2:00 – 3:30 pm **CONCURRENT SESSIONS B**

B-1. PANEL: Engaging Stakeholders to Develop Innovative Public-facing Data Products (*Vessey 1*)
B-2. From Data to Decisions: Leveraging Administrative Records (*Vessey 2*)
B-3. Open Source Software in the Federal Statistical System (*Patuxent*)
B-4. Delivering Solutions for Evidence Building in a NSDS Demonstration (*Room 2110*)
B-5. Improvements to Model-Based Small Area Estimations (*Room 2100*)
B-6. Maximizing Response Rates: Practical Solutions for Higher Engagement (*Room 0105*)

3:30 – 3:45 pm **Break**

3:45 – 5:15 pm **CONCURRENT SESSIONS C**

C-1. Innovative Methods for Assessing Nonresponse Bias and Refusals (*Vessey 1*)
C-2. Designing an Economy-Wide Survey: The AIES (*Vessey 2*)
C-3. Data Science Methodologies and Applications (*Patuxent*)
C-4. Visualizing Insights: Unveiling the Power of Data (*Room 2110*)
C-5. Data Insights Across Federal Statistics (*Room 2100*)
C-6. Transforming Data Collection: Advancements in Field Operations (*Room 0105*)

Wednesday, October 23

7:30 am – 5:00 pm **Registration Open**

7:30 – 8:30 am **Continental Breakfast**

8:30 - 10:00 am **CONCURRENT SESSIONS D**

D-1. AI-Ready Data: Promise, Practices, Policies from the Dept of Commerce (*Chesapeake A*)
D-2. What's Going on in There? Applications of Formal Privacy (*Chesapeake B*)
D-3. Expanding Gender and Sexuality Measurement (*Chesapeake C*)
D-4. How Should Federal Statistical Agencies Measure 'Relevancy' of Their Data? (*Vessey 1*)
D-5. Using Record Linkage of Administrative Records to Improve Fed Stats (*Vessey 2*)
D-6. Toward a More Perfect Sample: Case Studies in Sample Redesign (*Room 0105*)

10:00 - 10:30 am **Coffee Break,** *Sponsored by the American Statistical Association*

10:30 - 12:00 pm **CONCURRENT SESSIONS E**

E-1. Cracking the Code to Higher Response Rates: Evidence and Innovations (*Chesapeake A*)
E-2. Accessible Data for Young Learners: Making Federal Data Relevant to Students (*Chesapeake B*)
E-3. Establishing the Establishments: Innovations in Establishment Data (*Chesapeake C*)
E-4. PANEL: Roles and Responsibilities of Statistical Officials (*Vessey 1*)
E-5. Creating a System of Data on HH Income, Consumption, and Wealth (*Vessey 2*)
E-6. Sexual Orientation and Gender Identity (SOGI) Data Action Plans (*Room 0105*)

12:00 – 1:45 pm **Lunch on Your Own**

1:45 – 3:15 pm **CONCURRENT SESSIONS F**

F-1. Using AI and ML to Improve Data Analytics 1 (*Chesapeake A*)
F-2. New Bayesian Methods for Statistical Data Privacy (*Chesapeake B*)
F-3. Understanding the Landscape – the Annual Business Survey (*Chesapeake C*)
F-4. Improving Government Efficiency with Official Statistics (*Vessey 1*)
F-5. Appraise, Assess, & Apply: Advancing Alternative Data Use (*Vessey 2*)
F-6. Assessing Quality in an Online Probability Panel Survey: NCHS RSS (*Room 0105*)

3:15 - 3:30 pm **Break**

3:30 – 5:00 pm **CONCURRENT SESSIONS G**

G-1. Best Practices and Challenges of Disclosure Reviews (*Chesapeake A*)
G-2. Effective Strategies in the Era of Evidence Building; The Future is Now! (*Chesapeake B*)
G-3. Making the Invisible Visible: Illuminating Race, Gender, Class Inequalities (*Chesapeake C*)
G-4. The Roles of Data Quality and Transparency in Supporting Scientific Integrity (*Vessey 1*)
G-5. Evidence-based Bayesian Methods (*Vessey 2*)
G-6. Striking the Right Balance of Old and New: Combining Paper and Web (*Room 0105*)

Thursday, October 24

7:30 am – 5:00 pm **Registration Open**

7:30 – 8:30 am **Continental Breakfast**

8:30 - 10:00 am **CONCURRENT SESSIONS H**

H-1. Innovations in Sample Design: From Theory to Practice (*Chesapeake A*)
H-2. Balancing Data Privacy and Utility (*Chesapeake B*)
H-3. Federal Advances in Sexual Orientation and Gender Identity Measurement (*Chesapeake C*)
H-4. PANEL: Measuring the Health of the Federal Statistical System (*Vessey 1*)
H-5. Bringing it All Together: Exploring Ways to Mesh Data from Multiple Sources (*Vessey 2*)
H-6. The National Household Travel Survey: Applications and Opportunities (*Room 0105*)

10:00 - 10:30 am **Break**

10:30 am - 12:00 pm **CONCURRENT SESSIONS I**

I-1. LLMs Unleashed: Revolutionizing Surveys with AI (Chesapeake A)
I-2. New Explorations of Disclosure Avoidance Techniques (Chesapeake B)
I-3. Vital Signs and Data: A Healthy Dose of Analytics (Chesapeake C)
I-4. Leading the Federal Statistical System: Today's Lessons for Tomorrow's Success (Vessey 1)
I-5. Data Linkages Support Expanded Analyses in the U.S. Health System (Vessey 2)
I-6. Piecing Together the Puzzle: Novel Methods for Missing Data (*Room 0105*)

12:00 - 1:45 pm **Lunch on Your Own**

1:45 – 3:15 pm **CONCURRENT SESSIONS J**

J-1. Keeping AI Out of Trouble: Guardrails and Applications for Federal Data (*Chesapeake A*)
J-2. Communicating Quality in an Evolving Federal Data Ecosystem (*Chesapeake B*)
J-3. Advancing Equity in Survey Research: Innovations in Assessment (*Chesapeake C*)
J-4. The U.S. Federal Statistical System: Why, What, Who, and Career Opportunities (*Vessey 1*)
J-5. Let's Link Up: Data Integration Insights, Innovations, and Impacts (*Vessey 2*)
J-6. Asking Good Questions: Survey Instrument Design Considerations (*Room 0105*)

3:15 – 3:30 pm **Break**

3:30 – 5:00 pm **CONCURRENT SESSIONS K**

K-1. Using AI and ML to Improve Data Analytics 2 (*Chesapeake A*)
K-2. Safe Data Technologies: Safely Expanding Access to Administrative Tax Data (*Chesapeake B*)
K-3. Recent Findings and Advances in Probability-based Panel Research (*Chesapeake C*)
K-4. The Future is Now: How the U.S. Statistical System is Enabling Evidence Building (*Vessey 1*)
K-5. Optimizing Survey Methodologies for Enhanced Data Quality (*Vessey 2*)
K-6. Shifting into Safer Gear: Data-Driven Strategies in Transportation (*Room 0105*)

Abstract Listings for Tuesday, October 22

- **Concurrent Sessions A** **9:00 am – 10:30 am**
- **Plenary Session** **11:00 am – 12:15 pm**
- **Concurrent Sessions B** **2:00 pm – 3:30 pm**
- **Concurrent Sessions C** **3:45 pm – 5:15 pm**

Session A-1: Innovations in Federal Data Management and Evaluation: Enhancing Standards, Collaboration, and Accessibility

Organizer: Tywanquila Walker, *Bureau of Labor Statistics*
Chair: Tywanquila Walker, *Bureau of Labor Statistics*
Vessey 1

Building Capacity: The benefits of an Interdisciplinary Team in Federal Program Evaluation and Research

Anna Maria Calcagno, *Small Business Administration*
Shabori Sen, *Small Business Administration*
Shay Meinzer, *Small Business Administration*
Sana Ahmed Wilder, *Small Business Administration*
Travis Dorsey, *Small Business Administration*
Olivia Gonzalez, *Small Business Administration*
Placeholder, *General Services Administration*

Federal program evaluations often struggle to provide practical and actionable recommendations for program staff to implement. The U.S. Small Business Administration's Office of Strategic Management and Enterprise Integrity (OSMEI) addresses this by fostering interdisciplinary teams with economists, program evaluators, and data scientists. This approach leverages mixed methods to gain a more comprehensive understanding of programs, and our Federal Equity in Procurement Evaluation Portfolio will serve as the use case for this presentation. This presentation highlights the benefits of interdisciplinary teams and mixed methods for professionals seeking to improve the impact of their work. Taking this approach to evidence-building allows us to comprehensively bring together a variety of tools in our toolkit especially when data limitations and ambiguity in definitions exist. These tools include, but are not limited to data collection, process improvement, automation, machine learning, evaluation, and policy design. Our team initiated a mixed-methods portfolio approach to examine equity in federal contracting that has employed literature reviews, quantitative analysis, qualitative evaluations, and experiments. This resulted in practical guidance for agencies, such as dashboards and implementation guides. In an ongoing effort, we are partnering with the GSA Office of Evaluation Sciences for the behavioral experiments. In this session, we highlight this portfolio as an illustrative example of OSMEI's vision to better harness the power of interdisciplinary teams for developing richer program insights. Beyond methodological expertise, we will also discuss the importance of communications and strategic thinking in building actionable evidence for programs to be more effective.

Operationalizing a Congressional Mandate: Examples from the Cybersecurity Workforce Data Initiative

Amber Levanon Seligson, *National Center for Science and Engineering Statistics*
Kelly Phou, *National Center for Science and Engineering Statistics*
Gigi Jones, *National Center for Science and Engineering Statistics*
Caren A. Arbeit, *RTI International*
Michael Hogan, *RTI International*
Erin Dunlop Velez, *RTI International*

Federal Statistical Agencies have long been required by Congress to collect or report data. For example, the Science and Engineering Equal Opportunities Act of 1980 required that the National Science Foundation (NSF) report biennially on the sex and race and ethnicity composition of the scientific workforce. Operationalizing Congressional data collection or reporting mandates can be challenging due to the typically abstract nature of the requirements. This presentation describes how the National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation has designed and implemented methodologies for meeting the initial stages of a new mandate. The CHIPS and Science Act of 2022, section 10317, requires that NCSES and other entities assess whether it is possible to generate "nationally representative estimates and

statistical information on the cybersecurity workforce." In October 2023, NCSES launched the Cybersecurity Workforce Data Initiative (CWDI). This presentation informs audiences tasked with collecting federal statistical data on how the CWDI has addressed three hurdles in operationalizing the mandate: (1) determining the optimal approach to collaboration with other federal entities in responding to the mandate; (2) choosing a methodology for defining the population that is the focus of the mandate; (3) identifying pragmatic means to finding out whether data already exist that can meet the mandate, or whether new data need to be collected. By describing practical approaches to meeting a congressional data collection or reporting mandate, this presentation can help inform other federal statistical agencies facing similar requirements.

Re-envisioning and Revitalizing the Health and Human Services (HHS) Data Council: The Process of Implementing Changes to the Goals, Work, and Charter of the Data Council

Sarah Lessem, *HHS/ASPE*

The HHS Data Council is the principal internal advisory body to the Secretary on health and human services data policy. It was established in 1995 to coordinate data policy issues and conduct research to improve long-term collection and use of HHS data. The charter stated that the Data Council would:

- Develop the Department-wide data collection strategies,
- Work with HSS and external stakeholders to ensure data collected by HSS is useful and accessible to all users,
- Work towards consensus in data standards and privacy,
- Serve as liaison to the HHS federal advisory committee and respond to issues it presents related to data standards and privacy,
- Oversee survey and general statistical analysis, and
- Implement interagency efforts on information systems policy.

Over the past 30 years, other data governing, evaluation, privacy, and security boards, councils, committees, and groups have arisen to address many of these issues. This resulted in a shift in the need for and focus of Data Council work and uncertainty to its new purpose. Therefore, the Data Council is going through a strategic overhaul and rechartering process. The new HHS Data Council will address existing HHS data challenges without duplicating the work of other, more recently formed, councils. The presentation will discuss the process of re-envisioning the 30-year-old HHS Data Council, determining its new goals, obtaining stakeholder feedback, examining membership, updating its charter, and creating working groups that are effective in a virtual world. While this presentation will focus on the HHS Data Council, the strategies we discuss can be adapted to be used in revitalizing other government councils and processes.

Statistical Classifications: A FAIRy Tale

Peter Meyer, *Bureau of Labor Statistics*
Daniel Gillman, *Bureau of Labor Statistics*
Faouzi Aloulou, *Energy Information Administration*
William Savino, *U.S. Census Bureau*

The FAIR principles (Findable, Accessible, Interoperable, Reusable) were developed with scientific data in mind. Soon after they were published, it became clear the principles were applicable to data from all domains. A large number of organizations and research projects have adopted FAIR. In addition, a set of guidelines for implementing FAIR were developed. Conferences devoted to FAIR are now regular events.

At the same time as the FAIR principles have gained such popularity, the standard classification systems used in the federal statistical community are maintained in outdated ways, mostly in human readable formats, and buried within the web sites where they are maintained, there are Excel spreadsheet files. These at least are machine actionable, but they are not immediately findable for humans and difficult to find automatically.

In addition to the advent of the FAIR principles, knowledge representation languages and techniques were developed and in increasing use. Specific frameworks written in these languages, called vocabularies, have been written with statistical classifications in mind. One of the these is the eXtended Knowledge Organization System (XKOS) maintained under the DDI Alliance. The use of XKOS or similar technology to maintain and

disseminate statistical classifications will make them FAIR.

This talk addresses these concerns in more detail. Examples from around the world will illustrate the ideas. A clear rationale for adopting the approach will be provided.

Return to Schedule Overview

Tuesday, October 22, 2024 9:00 AM

Session A-2: When the Rubber Hits the Road: Practicalities of Integrating Alternative Data Sources

Organizer: Nikolas Pharris-Ciurej, *U.S. Census Bureau*
Chair: Linden Mcbride*, U.S. Census Bureau*
Vessey 2

Partnering with State agencies to Acquire and Use Administrative Records

Renuka Bhaskar, *U.S. Census Bureau*
Maria Perez-Patron, *U.S. Census Bureau*

The U.S. Census Bureau has a long history of entering into agreements to receive administrative data from federal, state, and local agencies in an effort to provide quality statistics and improve census and survey operations. One example of this is a longstanding effort to enter into agreements with state agencies that administer nutrition and cash assistance programs – the Supplemental Nutrition Assistance Program (SNAP), the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), and Temporary Assistance for Needy Families (TANF). States that send administrative data to the Census Bureau have the option to receive estimates of program eligibility and participation derived from their administrative data linked to survey data. In this presentation, we will describe the process of working with state agencies to acquire, clean, and utilize state administrative data to develop a product for participating states. We will also highlight the SNAP Longitudinal Data Project, a joint effort of the USDA-Food and Nutrition Service, the U.S. Census Bureau, and state SNAP agencies, in which we are developing a standardized research database using state SNAP administrative records, and present some preliminary results of the type of estimates that researchers will be able to obtain using this database.

Maximizing Linkage in Address Data: Spatial, Exact, and Fuzzy Matching

Timothy Champney, *MITRE*
Hongxun Qin, *MITRE*
Stephanie Coffey, *U.S. Census Bureau*

High quality record linkages are critical for enriching survey data with alternative data sources. To enhance the American Community Survey (ACS), we evaluated several options to match commercially available property data to housing unit records in the ACS and the Census Master Address File (MAF) data, with the ultimate goal of supplementing data collected in the survey with the commercial property data. Techniques to match address data come in two flavors: spatial matching and address matching.

For our application, spatial matching is done by overlaying commercial boundary shape files on the lat-long coordinates on the MAF to associate Census housing unit records to commercial property parcel records. This method is useful because it does not require matching of text fields, but performs poorly when parcels include many housing units (e.g., large apartment buildings). Address matching, or entity resolution at the address level, links records across the two sources based on the content of various address fields, and offers the possibility of disambiguating multiple matches and matching objects that cannot be successfully assigned a unique match through spatial matching. Several approaches are available for address matching, including rule-based, deterministic, fuzzy, and probabilistic methods.

In our paper/presentation we compare various combinations of spatial and address matching techniques to illustrate the tradeoff between linkage rates and linkage quality. We also present a hybrid solution that leverages spatial matching as well as several types of address matching to maximize high quality linkages for our research.

Enriching Survey Data with Road Network Analysis for Emergency Response

Joey Marshall, *U.S. Census Bureau*
Abraham Cheung, *U.S. Census Bureau*
Bethany DeSalvo, *U.S. Census Bureau*
Chase Sawyer, *U.S. Census Bureau*
Heather King, *U.S. Census Bureau*
Isabelle Notter, *U.S. Census Bureau*

In the context of emergency management, empirical questions often emerge that must be answered quickly to guide disaster response. Under these conditions, time constraints prohibit the collection of new survey data on the populations affected by a disaster. However, the utility of existing survey data can be enhanced to answer new and complex research questions that emerge in the rapidly-evolving aftermath of a disaster.

Following the March 26, 2024 collapse of the Francis Scott Key Bridge in Baltimore, the U.S. Census Bureau rapidly responded by enriching American Community Survey (ACS) data with road network data and routing analysis to identify commuters affected by the Bridge collapse. This was accomplished by plotting the likely driving route to work for every car-commuting ACS respondent who lived in Maryland, worked in Baltimore city or Baltimore County, and drove or carpooled to work. Driving routes were calculated using OpenStreetMap data and open-source routing software in conjunction with information on ACS respondents' residence, place of work, and means of transportation to work. After identifying a subset of commuters who likely drove across the bridge, Census Bureau analysts were able to identify distinct sociodemographic differences between bridge and non-bridge commuters.

This analysis demonstrates how survey data can be enriched and used in new ways to answer emerging research questions in the context of emergency management, and how information can be rapidly disseminated among decision makers in the days and weeks following a disaster.

Public Opinion on Specific Potential Uses of Administrative Data

Aleia Fobia, *U.S. Census Bureau*

In the face of declining survey participation rates and increasing demand for fast and efficient data products, governments are increasing their reliance on data from alternative sources including records and commercially available data. Public perception is one of the many challenges faced by this shift in federal data infrastructure. As the push to increase data sharing and data linkage has strengthened, it has brought increased scrutiny to how changes might affect the trust relationship between respondent and government data collection efforts.

Privacy and confidentiality concerns have been the focus of previous research on public opinion around the use of administrative data. In earlier studies, we found little association between general attitudes about privacy and opinions on the use of administrative data. Contextual theories of privacy provide some explanation for the dissonance by highlighting the critical importance of context for how individuals make decisions around appropriate information flows and distributional norms.

In this presentation, we present results from a mixed-methods study about public perception of specific uses of administrative data in Census Bureau survey and census operations. We provide in-depth context to respondents about specific plans to use administrative and auxiliary data to modify and target contact methods, adjust responses, and determine vacancy status. Preliminary results support the importance of respondents' expectations around the use and accuracy of government data and the influence of context on privacy concerns.

Return to Schedule Overview

Session A-3: Text tactics: Natural Language Processing in Action

Organizer: Ellen Galantucci, *Federal Maritime Commission*
Chair: Ellen Galantucci, *Federal Maritime Commission*
Patuxent

Exploring Data Science Methods in Analyzing Text Information for Datasets

WenWei Zeng, *Bureau of Transportation Statistics*
Young-Jun Kweon, *Bureau of Transportation Statistics*

Research papers play a crucial role in literature review and topic identification within the academic community. The process of reviewing papers and articles requires a substantial amount of effort to extract relevant information for incorporation into an individual's research. Policy makers and transportation professionals at various levels of government rely on data from the Commodity Flow Survey (CFS), collected by the U.S. Department of Transportation Bureau of Transportation Statistics and the U.S. Census Bureau, to evaluate transportation demand, energy consumption, safety risks, and environmental impacts.

Traditionally, researchers have relied on manual methods, such as reading and taking notes, to identify relevant topics within datasets like the CFS. However, this process can be time-consuming and inefficient when dealing with many papers. Text analysis techniques, such as text frequency analysis and topic models, offer an additional efficient way to quantify the use of datasets. These methods help researchers identify key words and themes within large document collections, providing valuable insights into the subject matter.

Findings from text analysis reveal common topics covered in research articles that utilize the CFS, such as economics, environmental subjects, supply chain, and transportation. These methods are not limited to the CFS dataset and can be applied to various other datasets at the federal or state level. Additionally, text analysis techniques have broader applications, including identifying research trends in science fields and fulfilling information technology requests. By leveraging these tools, researchers can streamline the process of extracting valuable insights from large document collections.

Leveraging Natural Language Processing for Legislative Research

Pavani Samala, *Reveal Global Consulting*
Mercera Silva, *U.S. Census Bureau*
Amber Hennessey, *U.S. Census Bureau*
Taylor Wilson, *Reveal Global Consulting*
Jayram Athimoolam, *Reveal Global Consulting*

The U.S. Census Bureau's Public Sector Frame and Classification Branch (PSFCB) reviews all legislation the 50 state governments and the District of Columbia enact, with the goal of identifying and classifying relevant legislative changes for local and state government births, deaths, and mergers. This research supports the frame maintenance for the Census of Governments (CoG) and is the basis for the Individual State Descriptions report, which provides information about how state and local governments are organized. Analyst review of this legislation is labor-intensive, encompassing hundreds of bills for each state annually, and very few bills provide relevant changes needed for classification.

This research explores methods to reduce analyst burden and to expedite the research process with a combination of web scraping, natural language processing (NLP), and machine learning to develop an automation tool. The purpose of this tool is to recommend specific bills per legislative session for analyst review, based on language combinations and placement within the text. The automation process is designed to recommend rather than automatically classify bills to ensure the most relevant changes are captured and to allow analysts to make final decisions on whether a change is necessary.

This presentation will discuss the web scraping process, the methodology associated with classification, and how NLP improves analysts' workflow. We will explore methods of measuring burden reduction and present

preliminary findings from user testing.

Machine Learning for In-Instrument Product Code Search

Emily Wiley, *U.S. Census Bureau*
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. NAPCS consists of over 7,000 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 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. In this talk, we will discuss the development and results of the SINCT model and will highlight how machine learning can improve respondents' experience while also reducing the amount of expensive manual processing that is necessary after collection.

NLP Methods for Detecting Non-Therapeutic Drug Use in Clinical Notes

Nikki Adams, *National Center for Health Statistics*
Rihem Badwe, *National Center for Health Statistics*

The National Center for Health Statistics (NCHS) received a Patient-Centered Outcomes Research Trust Fund investment to develop an algorithm to detect the non-therapeutic co-use of opioids and stimulants in the National Hospital Care Survey, a nationally representative survey of hospital encounters. In addition to medical codes, clinical notes were analyzed using natural language processing (NLP) methods for hospitals providing electronic health records (EHR) data.

FIHR-standard data comes with useful metadata such as what type of EHR section the text concerns, and this information is helpful when analyzing clinical notes. This metadata can indicate, for example, whether the text is a list of medications or patient's social history. However, EHR data may not always come with this metadata. When this happens, NLP methods can be used to make up for this lack of information.

Initial findings indicated that a combination of note type metadata (for example, "medications" versus "social history") and drug term regular expressions was highly accurate in detecting opioid and stimulant non-therapeutic use. Where encounters did not have that metadata, initial algorithm performance was poor. A logistic regression machine learning model, trained through a fairly simple and fast process, was used to detect non-therapeutic contexts. This acted as a proxy of the metadata that was missing, increasing overall algorithm performance from 0.65 Matthew's Correlation Coefficient to 0.85 for any non-therapeutic stimulant use. This presentation will discuss the use of NLP to improve algorithm performance for identifying non-therapeutic stimulant and opioid use.

Streamlining ACS Autocoding: Leveraging Natural Language Processing and LLMs

Alexander Zakrzeski, *Reveal Global Consulting, LLC*
Julia Beckhusen, *U.S. Census Bureau*
Ana J Montalvo, *U.S. Census Bureau*
Jackson Chen, *Reveal Global Consulting, LLC*
Lynda Laughlin, *U.S. Census Bureau*
Yezzi Angi Lee, *Reveal Global Consulting, LLC*

The American Community Survey (ACS) is conducted by the U.S. Census Bureau, aimed at gathering comprehensive demographic, social, economic, and housing data from a representative sample of households across the United States. Within the ACS, both industry and occupation write-in descriptions are collected from respondents and converted into specific Census Bureau Industry/Occupation Codes through an autocoding

process and clerical coding.

In this research, we analyzed multiple years of ACS industry and occupation write-in data to identify common titles and descriptions to update the Alphabetical Indexes of Industry and Occupation. Next, we investigated preprocessing methods such as spell check and fuzzy matching to enhance data quality before coding. This process includes identifying the most common item-non responses and refusals, and standardizing abbreviations and acronyms.

Finally, we explored new approaches to developing a new autocoder process, leveraging natural language modeling techniques and large language models (LLMs). We utilized an improved search engine based on semantic search and powered by LLMs. We will delve into experiments in exploring different embedding models, comparing context retrieval window sizes and sequential workflows, and reranking strategies.
This presentation will discuss the results of our analysis, the challenges encountered during the process, and propose potential solutions. Our framework aims to streamline coding procedures, ensuring accuracy and efficiency in the ACS data processing system.

Return to Schedule Overview

Tuesday, October 22, 2024 9:00 AM

Session A-4: Disclosure Avoidance and Product Structure in the 2020 Decennial Census Supplemental Demographic and Housing Characteristics File

Organizer: Philip Leclerc, *U.S. Census Bureau*
Chair: Philip Leclerc*, U.S. Census Bureau*
Discussant: Sallie Keller*, U.S. Census Bureau*
Room 2110

The Supplemental Demographic and Housing Characteristics File: Structure and Tabulations

Michael Hawes, *U.S. Census Bureau*
Alexandra Krause, *U.S. Census Bureau*

We introduce the structure and tabulations in the Supplemental Demographic and Housing Characteristics File (S-DHC). The S-DHC product includes counts and averages for the population living in households and families of different types. For example, this includes average household size by age and tenure, average family size, and total population in households by tenure, published at Nation and State levels. In this talk, we discuss the overall structure of the S-DHC product, the tabulations in it, and changes in the set of tabulations published from the corresponding tabulations in 2010. By introducing the S-DHC product and explaining its structure, this talk will sets the stage for and helps to frame the other three talks in this session.

PHSafe: The Disclosure Avoidance Algorithm for the Supplemental Demographic and Housing Characteristics File

William Sexton, *Tumult Labs*
Ashwin Machanavajjhala, *Tumult Labs*
David Pujol, *Tumult Labs*

We present the disclosure avoidance algorithm, PHSafe, that the U.S. Census Bureau is using to protect the Supplemental Demographic and Housing Characteristics File (S-DHC) of the 2020 Census. The tabulations contain statistics of demographic and housing characteristics of U.S. persons living in households. We describe the PHSafe algorithm, which is based on adding noise drawn from a discrete Gaussian distribution to the statistics of interest. We discuss why the S-DHC tabulations are challenging to protect while providing adequate

accuracy. We highlight several policy decision points that arose during the development of PHSafe and discuss the tooling Tumult developed to help evaluate trade-offs in these decision points.

Per-Attribute Privacy Semantics for the Supplemental Demographic and Housing Characteristics File

David Pujol, *Tumult Labs*
Ashwin Machanavajjhala, *Tumult Labs*
William Sexton, *Tumult Labs*

Differentially private releases such as the U.S. Census Bureau's Supplemental Demographic and Housing Characteristics File (S-DHC) ensure quantifiable privacy protections for all individuals in the dataset. The strength of the privacy guarantee is measured by the differentially private privacy loss, which is a measure of the amount of information any adversary can gain about individuals from the data. This value is instrumental in understanding the privacy risks associated with any release. However, since this privacy loss applies to all records and to all information about that record, it is often an overestimate of the privacy risk associated with qualities of an individual such as their age and location.

To understand the privacy risk of an individual's qualities, we propose the use of per-attribute privacy loss, an alternative measure of privacy risk for existing differentially private algorithms. The per-attribute privacy loss measures the privacy risk associated with attributes of an individual, and therefore measures an adversary's ability to infer these attributes from a differentially private data release. This measure allows a data curator to get a more fine-grained analysis of the specific privacy risks associated with a data release and how those risks are distributed across a population.

We conduct per-attribute privacy loss analysis on the S-DHC data product to demonstrate how, in practice, the attribute privacy loss is significantly less than the total privacy loss for many protected secrets and how the attribute privacy loss can differ depending on the individual characteristics of the individual.

Bayesian Methods to Improve the Accuracy of Differentially Private Measurements of Constrained Parameters

Scott H. Holan, *University of Missouri and U.S. Census Bureau*
Ryan Janicki, *U.S. Census Bureau*
Kyle Irimata, *U.S. Census Bureau*
James Livsey, *U.S. Census Bureau*
Andrew Raim, *U.S. Census Bureau*

Formal disclosure avoidance techniques are necessary to ensure that published data cannot be used to identify information about individuals. The addition of statistical noise to unpublished data can be implemented to achieve differential privacy, which provides a formal mathematical privacy guarantee. However, the infusion of noise results in data releases which are less precise than if no noise had been added and can lead to some of the individual data points being nonsensical. Examples of this are estimates of population counts which are negative or estimates of the ratio of counts which violate known constraints. A straightforward way to guarantee that published estimates satisfy these known constraints is to specify a model and incorporate a prior which properly constrains the parameter space. We utilize rejection sampling methods for drawing samples from the posterior distribution and we show that this implementation produces estimates which maintain formal privacy, are more precise than the original unconstrained noisy measurements, and are guaranteed to satisfy prior constraints. Applications to state/nation level modeling of U.S. Census Bureau's 2020 S-DHC will be highlighted and additional modeling opportunities for decennial census data will also be discussed.

Return to Schedule Overview

Session A-5: Advancing Survey Methodologies: Innovations in Survey Modernization

Organizer: Erica Olmsted, *U.S. Census Bureau*
Chair: Erica Olmsted, *U.S. Census Bureau*
Room 2100

Data Modernization Efforts for the National Ambulatory Medical Care Survey

Nicole Cummings, *National Center for Health Statistics*
Carol DeFrances, *National Center for Health Statistics*
Prachi Mehta, *National Center for Health Statistics*

The National Ambulatory Medical Care Survey (NAMCS), conducted by the National Center for Health Statistics (NCHS), is a nationally representative survey designed to meet the need for objective, reliable information about the provision and use of ambulatory medical care services in the United States. As ambulatory health care delivery continues to change, NAMCS has needed to adapt to a different mode of data collection and include additional types of ambulatory care providers. Recently, both the NAMCS Health Center (HC) Component and the Provider Component were redesigned and launched. In 2021, the NAMCS HC Component began collecting patient visit data from HCs using electronic health records (EHR). The goal of integrating EHRs into data collection was to increase timeliness, efficiency, volume of records, and clinical data quality. In 2023, the NAMCS Provider Component began to survey a sample of physician associates/physician assistants (PAs) along with sampled physicians. Inclusion of PAs starts to address the changing scope of ambulatory care practices. This presentation provides an update on the NAMCS redesign and progress made as a result of its modernization efforts. Also highlighted are several additional efforts to enhance electronic data collection at NCHS. These include the development of NCHS' first Confidential Information Protection Statistical Efficiency Act compliant cloud for the collection, processing, and storing of EHR data, the piloting of an automated EHR data transmission mechanism using the Fast Healthcare Inoperability Resources (FHIR) standard, and ensuring program sustainability by building internal technical capacity to manage FHIR-based applications.

Maximizing Overlap of NAEP School Samples to Optimize both Trend and Cross-sectional Estimates

Lloyd Hicks, *Westat*
Amy Lin, *Westat*
Yiting Long, *Westat*
Keith Rust, *Westat*

This presentation examines the school sampling methods used for the Long-Term Trend (LTT) assessments for 2022. Historically, the National Center for Education Statistics (NCES) has used these assessments since the early 1970s to monitor student performance in mathematics and reading across three age groups: 9, 13, and 17. Typically conducted every four years, the 2022 LTT assessments were advanced by two years to assess the impact of the pandemic on student performance and to enable comparisons with the pre-pandemic 2020 data. The preferred approach for the 2022 LTT sample design was to replicate the 2020 school sample to the greatest extent possible, since reassessing the same students from the 2020 assessments is not an option. Overlap control, specifically maximizing overlap, is infrequently used for NAEP assessments; but when the need arises, Keyfitz procedures (based on the application of Bayes Theorem for conditional probabilities) are used. Keyfitz procedures by themselves, do not ensure exact school sample replication, particularly in cases of significant reductions in student enrollment in individual schools. However, the 2022 LTT sample design adopted an innovative approach to ensure the maximum school sample overlap. The methodologies discussed have implications for educational assessments and extend to other establishment-type survey samples requiring high degrees of overlap.

Paradata Analysis of Participant Recruitment and Retention for COVID Household Transmission Study

Xiaoshu Zhu, *Westat*
Steph Battan-Wraith, *Westat*
Vanessa Olivo, *Westat*
Kerry Grace Morrissey, *Westat*

As part of the CDC-sponsored Respiratory Virus Transmission Network National (RVTN-N), a COVID-19 household transmission study with virtual enrollment was conducted with the aim of achieving greater geographic diversity and inclusion in the study population. Using a case-ascertained design, this study remotely enrolled index cases who tested positive for COVID-19 within the past 5 days and requested the participation of the majority of household members. Both index cases and household members went through multiple surveys to assess their eligibility before they were enrolled. The purpose of this analysis is to examine how study features - especially those derived from paradata - influenced participants' response status. Various statistical models were conducted, including multilevel logistic regression and conditional inference tree analysis to assess feature impact on response status changes between stages. Discrete time survival analysis and survival tree models evaluated the association between study features and the time the participants withdrew from the study. This analysis found strong retention effects from both participant characteristics (e.g., participation type, parental permission, and household type and size) and recruiting features (e.g., reminder contact mode, number of session logins, and response device).

Using the Annual Business Survey to Improve Measurement of the U.S. Cybersecurity Workforce

Shelley Feuer, *National Center for Science and Engineering Statistics*
Gigi Jones, *National Center for Science and Engineering Statistics*
Michael Prebil, *National Institute of Standards and Technology*
Karen Wetzel, *National Institute of Standards and Technology*

Cyber technology brings innovation and automation; it also brings cyberattacks, threatening U.S. citizens, businesses, and national security. However, data on the cybersecurity workforce are limited. The National Center for Science and Engineering Statistics (NCSES) is mandated by the CHIPS and Science Act of 2022 (CHIPS) to assess the feasibility of producing national estimates on this workforce. Relatedly, NCSES is developing a cybersecurity module on its mandatory Annual Business Survey (ABS). An ABS module can provide essential industry data about employed cybersecurity workers.

NCSES draws on the Workforce Framework for Cybersecurity (commonly referred to as the NICE framework) from the National Institute of Standards and Technology. The NICE framework is applied in CyberSeek, which measures the cybersecurity supply and demand gaps through web scraping job postings. While this method provides some estimates on cybersecurity professionals, it's limited in capturing industry-market data and meeting the federal statistical standards for producing official statistics. The ABS module will fill critical information gaps, augmenting NICE data assets and meeting the CHIPS mandate.

This presentation will provide an overview of the NICE framework, highlighting aspects applicable to the collection of these data in federal surveys. Results of cognitive testing of ABS cybersecurity survey questions will be shared with an assessment of methods for including questions on: industry definition(s) of cybersecurity, taxonomy, roles, educational profiles, and cost breakdown of services. We will discuss the utility and limitations of the NICE framework and the ABS when developing potential survey questions to improve measurement of the cybersecurity workforce.

Return to Schedule Overview

Session A-6: Make Everybody Count: Including Under-represented and Hard to Reach People in Federal Statistics

Organizer: Bob Sivinski, *Office of Management and Budget*
Chair: Rob Santos*, U.S. Census Bureau*
Discussant: Rob Santos*, U.S. Census Bureau*
Room 0105

Public Participation in the Review of OMB's Race and Ethnicity Statistical Standards

Jennifer Saindon, *U.S. Census Bureau*

In 2022, the Chief Statistician of the United States convened an Interagency Technical Working Group (Working Group) to update the Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity (SPD 15) since the demographics of the nation have changed since the last revision in 1997. There were numerous public campaigns and requests for updates to the Race and Ethnicity directive. On March 28, 2024, the Chief Statistician published the updates to SPD 15. In addition to research on the topics conducted by federal agencies, a key input to the Working Group's recommendations was Public Participation.

The Working Group used a multi-faceted approach to engage with the public: initial recommendations were published in the Federal Register to solicit feedback from the public; public town halls and listening sessions were conducted as well as a tribal consultation; and, the Working Group presented to key stakeholder groups, including at professional conferences. The themes of the topics from these engagements were varied – some were expected, and others were new ideas. All were considered and even if they were not adopted in this round of revisions, recurring themes were specifically named for future research.

This presentation will discuss the themes included in the revisions as well as topics for future research.

Engaging with External Advocates: The Power of Coalition and Community Organizing

Meeta Anand, *The Leadership Conference on Civil and Human Rights*

This talk will focus on the importance of community engagement in achieving relevant, timely and meaningful data equity. Through examples of work in race & ethnicity, SOGI and disability data, the talk will illuminate how engaging with external partners can help to collect statistical data that are meaningful to the communities from whom the data are captured. The discussion will include both strategies and different pathways for engagement as well as recommendations for uplifting and highlighting the importance of equitable data within communities and government.

Improving Data Collection for People with Disabilities: Challenges, Opportunities, and Recommended Actions

Adam Politis, *OSTP*

People with disabilities are a complex, heterogenous population that historically have not been accurately represented in many Federal statistics. Though the reasons for this are multifactorial, one increasingly recognized factor is that people with disabilities are a hard-to-count population. Traditional survey methods are not sufficient to fully include them in the data collection process, due to many subpopulations of people with disabilities being hard to locate, hard to contact, hard to persuade, and/or hard to interview. This talk will provide an overview of these challenges and their implications, discuss strategies for inclusion that are used with hard-to-count populations and examine their applicability to people with disabilities, and provide recommendations for specific actions to ensure that people with disabilities are more accurately enumerated.

Providing Value in Return for Data Collected from Tribal Nations
TBD, *TBD*

TBD

Return to Schedule Overview

Tuesday, October 22, 2024 11:00 AM

Plenary Session

Organizer: Linda J. Young, *National Agricultural Statistics Service*
Chair: Karin Orvis, *Office of Management and Budget*
Chesapeake Ballroom

Opening Remarks
Karin Orvis, *Office of Management and Budget*

TBD
Jeri Mulrow, *Westat*

TBD

Return to Schedule Overview

Tuesday, October 22, 2024 2:00 PM

Session B-1: PANEL: Engaging Stakeholders to Develop Innovative Public-facing Data Products

Organizers: Carla Medalia, Jennifer Ortman, Jennifer Hunter Childs, *U.S. Census Bureau*
Chair: Sallie Keller*, U.S. Census Bureau*
Vessey 1

In order to address evolving stakeholder requirements, federal agencies are constantly innovating their public-facing statistical products. These changes range from improving the way agencies disseminate existing data products to combining an agency's data assets to develop new insights. This panel will highlight the approaches taken across five agencies to accomplish this goal: the Census Bureau, the Department of Agriculture, the National Center for Science and Engineering Statistics, the National Center for Education Statistics, and the Bureau of Labor Statistics. The panelists will describe various initiatives, the challenges they encountered, and solutions they identified to ensure that agencies can meet the needs of their stakeholders. This interactive session will include discussion amongst the panel and engage the audience's questions.

Panelists

- Erika Becker Medina, *U.S. Census Bureau*
- Linda Young, *National Agricultural Statistics Service*
- John Finnamore, *National Center for Science and Engineering Statistics*
- Adam Safir, *Bureau of Labor Statistics*

Return to Schedule Overview

Tuesday, October 22, 2024 2:00 PM

Session B-2: From Data to Decisions: Leveraging Administrative Records in Statistics

Organizer: Gavin Corral, *National Agricultural Statistics Service*
Chair: Gavin Corral*, National Agricultural Statistics Service*
Vessey 2

Analyzing the Residential Status of Demographic Frame Addresses

Brandon Pipher, *U.S. Census Bureau*
Tom Mule, *U.S. Census Bureau*

The Census Bureau's Demographic Frame is a comprehensive, person-level frame consisting of geographic, demographic, social, and economic characteristics. Person-address records on the Demographic Frame are derived from administrative, third-party, census and survey data records.

A person found on the Demographic Frame may have multiple address records, as they are often associated with several addresses across the various data sources. These multiple address-records create difficulty in placing a person at their correct residential address according to a given reference day.

This analysis will evaluate addresses for people from these models on the Demographic Frame based on a reference date of July 1, 2021. As part of this analysis, we will compare these Demographic Frame addresses to addresses in the 2020 Census and the 2021 American Community Survey frames. This comparison will allow us to examine the Demographic Frame addresses that are not found within these other Census products, which may provide information to help identify whether subsets of these addresses may be more likely to be residential or non-residential addresses. This analysis can potentially improve the quality of the Demographic

Frame and its person-place records by increasing the chance of placing a person at their correct residential address.

Identifying Underserved Areas using Administrative Child Passenger Safety Data

Elizabeth Petraglia, *Westat*

Car seat checks are one of the most useful ways to ensure that children are traveling safely in vehicles (using the correct car seat for their weight and age, and properly installed and secured in the seat). However, not all caregivers take advantage of this resource. The National Digital Car Seat Check Form (NDCF) provides a national database of car seat checks performed by certified child passenger safety technicians, including detailed information on car seat use and misuse, but without any personally identifying information other than zip code. By attaching 2022 ACS 5-year estimates via zip code, we can estimate the proportion of families with children under age 8 in a given area that participate in and benefit from car seat checks. We also assess zip code-level characteristics, such as percent minority, commuting behaviors, household poverty, and educational status, to assess whether any of these characteristics are associated with car seat check participation or car seat misuse rates.

This information will enable child passenger safety agencies to better target their education and outreach programs to focus on underserved areas. This research also includes a discussion of strategies for disseminating information based on administrative data, such as the NDCF, to audiences with varying levels of data literacy.

Recent Differences in Survey and Administrative Measures of Payroll Job Growth

Steve Mance, *Bureau of Labor Statistics*
Kate Eckerle, *Bureau of Labor Statistics*
Chris Grieves, *Bureau of Labor Statistics*

Estimates of monthly nonfarm payroll employment change from the Bureau of Labor Statistics (BLS) are based on a large, establishment survey (CES) and serve as a Principal Federal Economic Indicator. The data are later benchmarked using administrative counts on the population of jobs covered by the unemployment insurance tax system (QCEW). The CES survey data are subject to error from sampling and from difficulties capturing business births and deaths, while both CES and QCEW contain other forms of nonsampling error. Differences in over-the-month employment changes reported from the two data sources can be large and, at one publication vintage, exceeded 1 million nationally from May-to-June 2022. This differential narrowed with further revisions. This paper builds on previous work examining differences in reporting procedures between CES and QCEW, decomposes recent divergences, and documents a pattern in the early vintages of the administrative data where establishments report zero employment in the third month of the quarter but positive employment in adjacent months. Many "month three zeros" subsequently revise to positive values and tend to lessen recent gaps between survey and administrative sources at an aggregate level.

Survey-First vs. Administrative Record-First Approaches to a Census

J. David Brown, *U.S. Census Bureau*
Marta Murray-Close, *U.S. Census Bureau*

About 3.2 percent of the population in the 2020 Census was enumerated using administrative records. The records were used after soliciting self-responses and making at least one Nonresponse Followup (NRFU) contact attempt, and then only if the records were estimated to be of sufficiently high quality. The existence of the people could not be validated for over 54 million of the records, introducing uncertainty into the counts. We construct alternative population statistics starting with administrative records with vintages near the 2020 Census reference date, only including records for people whose identities could be validated. Survey-collected data from the 2020 Census are used to update the locations and demographic characteristics of people found in administrative records. Coverage gaps are filled by including validated people from the 2020 Census who are missing from the administrative data. We construct different versions that include only administrative records, add self-response, add NRFU household interviews, add proxy responses, and add group quarters enumerations. We compare these sets of population statistics and the 2020 Census. This allows us to document the relative contributions of each survey operation to changes in the statistics.

Return to Schedule Overview

Tuesday, October 22, 2024 **2:00 PM**

Session B-3: Open Source Software in the Federal Statistical System

Organizer: Chris Marcum, *Office of the Chief Statistician of the United States*
Chair: Chris Marcum, *Office of the Chief Statistician of the United States*
Discussant: Mike Walsh (invited), *U.S. Census Bureau*
Patuxent

Details Pending

Return to Schedule Overview

Tuesday, October 22, 2024 **2:00 PM**

Session B-4: Delivering Solutions for Evidence Building in a National Secure Data Service Demonstration

Organizer: May Aydin, *National Center for Science and Engineering Statistics*
Chair: May Aydin, *National Center for Science and Engineering Statistics*
Discussant: Eddie Thomas, *Department of Veteran Affairs*
Room 2110

An Overview of the National Secure Data Service Demonstration (NSDS-D): Where are we now?

Heather Madray, *National Center for Science and Engineering Statistics*

On August 9, 2022, President Joe Biden signed the CHIPS and Science Act of 2022 into law. The legislation authorized the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation to establish a National Secure Data Service Demonstration Project, or NSDS-D. An NSDS would provide a platform of shared services to streamline and innovate data access, data linkage, and privacy-protections to support expanded data use for evidence building across the nation. This five-year demonstration project will inform decisions about whether an NSDS is established and the form it will take, determining what shared services are needed and are most effective, as well as what innovations can support an NSDS. This presentation will share a vision for an NSDS, activities, and an overview of the lessons learned from the first two years of the demonstration project. It will also discuss next steps as we explore the future of a potential NSDS.

Considerations for Successful Implementation of the National Secure Data Service Demonstration: Stakeholder Engagement and Lessons Learned

May Aydin, *National Center for Science and Engineering Statistics*

The National Secure Data Service (NSDS) Demonstration Project was authorized under Section 10375 of the 2022 CHIPS and Science Act by the Director of the U.S. National Science Foundation (NSF) and operated directly under the National Center for Science and Engineering Statistics (NCSES), a principal statistical agency within the U.S. Federal Statistical System. To support successful implementation of the NSDS Demonstration, the legislation also outlined key engagement touch points and transparency requirements to ensure coordination

with other stakeholders in its establishment. These requirements included consultation with the Office of Management and Budget (OMB) and the National Artificial Intelligence Initiative Act of 2020 Interagency Committee (NAIRR), engagement with federal and state government agencies, and the establishment of an NSDS website to showcase progress on supported projects.

In this session, we will provide an overview of these engagement efforts, the strategies used to solicit input, and perspectives from a diverse group of stakeholders to inform NSDS demonstration projects and activities. We will also discuss the lessons learned to inform future efforts given the complexity and national importance of establishing a set of shared services within the NSDS to support evidence building.

Utilizing Privacy Preserving Record Linkage to Link Data from Two Federal Statistical Agencies

Cordell Golden, *National Center for Health Statistics*
Lisa B. Mirel, *National Center for Science and Engineering Statistics*
Rob Zybrick, *HealthVerity*
Rui Wang, *Mathematica, Inc.*

Many federal agencies in the U.S. have begun exploring the use of privacy enhancing technologies, such as privacy preserving record linkage (PPRL), to integrate data from disparate sources to support evidence building and high-quality research. Recent legislation, including the Foundations for Evidence-Based Policymaking Act of 2018 and the CHIPS and Science Act, and recommendations from the Advisory Committee on Data for Evidence Building support the establishment of a shared service environment, namely the National Secure Data Service (NSDS), to facilitate this work. This talk will provide an overview of a NSDS Demonstration projects that utilizes PPRL to integrate data covered by different legal provisions. The project aims to link data between two federal statistical agencies. The talk will focus on the development of the required data sharing agreement, legal provisions that needed to be considered, security and privacy measures that had to be addressed, the software and IT infrastructure required, and the PPRL tool utilized for the project. The talk will conclude with a presentation of preliminary linkage results, lessons learned, and a summary of how the project will inform the establishment of a future NSDS.

Expanding Equitable Access to Confidential Data Through the Federal Statistical Research Data Centers: A Demonstration Project

Nate Ramsey, *U.S. Census Bureau*
Michael Castro, *U.S. Census Bureau*

For 30 years, the Federal Statistical Research Data Center (FSRDC) network has been a gateway for qualified researchers to access valuable federal statistical data for demographic and economic research, as well as evidence-building. Though the program was always intended to deliver a wide-reaching public benefit through offering this access, a goal we share with the future National Secure Data Service (NSDS), the evolution and growth of the network has left us facing an equity challenge.

Current FSRDC users are predominately faculty members and graduate students from R1 universities geographically proximate to, or directly hosting, an FSRDC. As a result, many researchers across the country go largely unserved by the FSRDC program including researchers from Minority Serving Institutions (MSIs), Historically Black Colleges and Universities (HCBUs), and other academic institutions. This also includes many researchers outside academia, such as those in State and local government, non-profit institutions, businesses, independent researchers, and others. To ensure that data from the FSRDC network can be used for evidence-building and research across every sector of society, the pool of data users must stretch beyond Federal agencies and traditional R1 academic researchers.

In this session we will give an overview of current efforts that are underway to identify underserved prospective users, as well as understand and address barriers that have led to inequity in the FSRDCs through a series of demonstrations. And, aside from addressing current needs, how we can apply lessons learned from these efforts to avoid perpetuating these inequities in the future NSDS.

Data Protection Toolkit Use Case Analysis: Leveraging and Enhancing Existing Tools within the Federal Statistical System to Promote a Shared Service Model

Michael Hawes, *U.S. Census Bureau*
Martha Stapleton, *NORC at the University of Chicago*

As part of the National Secure Data Service Demonstration (NSDS-D) project, the National Center for Science and Engineering Statistics (NCSES) implemented a use case analysis of the Federal Committee on Statistical Methodology's Data Protection Toolkit (DPT). The DPT outlines strong standards for data confidentiality and privacy, supports compliance with federal law, and provides guidance for upholding trust in federal statistical agencies' ability to protect individuals' privacy. The purpose of the research was to identify both successful uses and potential enhancements to the Toolkit that could expand access to federal datasets while protecting data confidentiality. The research team conducted 15 interviews with staff from federal and non-federal organizations, representing a diverse target audience for the DPT who could share how they have implemented the Toolkit, inform improvements, and provide ideas for additional content. An initial finding is that the Toolkit is recognized as a relevant and valuable resource by those who are familiar with it and those whose first exposure to it was during our interviews. At the same time, none of those interviewed had yet made the Toolkit integral to their data access and protection processes. This presentation will provide a brief overview of the interview methods followed by an in-depth discussion of findings and recommendations. The latter encompass a broad range of topics, from the techniques and resources organizations draw on to ensure privacy protection, to creative ideas for expanding awareness of the Toolkit's value.

Return to Schedule Overview

Tuesday, October 22, 2024 2:00 PM

Session B-5: Improvements to Model-Based Small Area Estimations

Organizer: Jennifer Rhorer, *National Agricultural Statistics Service*
Chair: Jennifer Rhorer, *National Agricultural Statistics Service*
Room 2100

Enhancements to the Modified Kalman Filter for Small Domain Estimation

Lauren Rossen, *National Center for Health Statistics*
Makram Talih, *National Center for Health Statistics*

Assessing health disparities can be challenging because direct estimates for small subgroups may be unreliable due to small sample sizes. As a result, analysts often aggregate data over larger areas, time periods, or groups to obtain reliable estimates, jeopardizing timeliness and public health relevance. The Modified Kalman Filter (MKF) is a statistical modeling tool developed in 2011 for small domain estimation, borrowing strength over time and across groups to produce more reliable model-based estimates for small subgroups. Recent enhancements to the MKF macro have been made to allow for additional flexibility in modeling non-linear trends, improving transparency and usability of the tool.

Simulated data based on the National Health and Nutrition Examination Survey and the National Health Interview Survey were used to assess the performance of the enhanced MKF approach relative to direct survey estimates based on the root mean squared error (RMSE).

The enhanced MKF tool resulted in marked improvements in RMSE relative to direct estimates, with larger improvements seen for smaller sample sizes. Improvements were seen across a wide array of scenarios, including outcomes with higher or lower prevalence, trends that varied from linear to cubic, trends that were shared or varied by group, and trends that involved unequally spaced time points. Gains in equivalent sample size of up to 420% were observed. The enhanced MKF macro can be used to produce model-based estimates of health outcomes for small subpopulations, which will improve the availability of data for the assessment and monitoring of disparities in small groups.

Forest Carbon Accounting with Model-based Estimation from Remote Sensing Data and Nationwide Forest Inventory Plots

Andrew Lister, *USDA Forest Service*
Laura Duncanson, *University of Maryland*
John Hogland, *USDA Forest Service*
Neha Hunk, *University of Maryland*
George Hurtt, *University of Maryland*
Lei Ma, *University of Maryland*
Barry T. Wilson, *USDA Forest Service*

Carbon accounting stakeholders have invested in producing fine resolution carbon maps as interest in terrestrial carbon accounting has grown. These maps, which are developed with remotely sensed data like LiDAR and satellite imagery, provide information on carbon stocks and dynamics at a fine scale. The USDA Forest Service's Forest Inventory and Analysis (FIA) program, which generates estimates of carbon dynamics at the sub-state to national scales, has incorporated maps through probability sampling and model-assisted estimation strategies, such as post stratification and regression estimation. However, stakeholders that do not have a valid probability sample, or are interested in estimates for small areas, require strategies that draw from the strength of models and ancillary data. In the model-based paradigm, the assumption is that estimates associated with the elements of the population are realizations of random variables, and the aggregation of these estimates at the population level comprises the population-level estimate. In the current study, we present results of a workflow that uses FIA data and maps to produce model-based estimates for the state of Maryland. Two Landsat satellite-based maps were used, one produced by FIA, and one produced by researchers associated with NASA's Carbon Monitoring System (CMS). We develop the cloud-based workflow for generating the estimates, and compare these with estimates generated from traditional, model-assisted estimation conducted using the same maps and FIA ground plots. The goal of the study is to establish a framework that FIA and other carbon accounting stakeholders can use in cases where model-based estimation is appropriate.

Measuring the Civilian Noninstitutional Population for Small Areas: Producing Estimates from 2000 to Present with Applications to Official Labor Force Statistics

Andrew C. Forrester, *Bureau of Labor Statistics*

The Local Area Unemployment Statistics (LAUS) program at the U.S. Bureau of Labor Statistics produces estimates of employment and unemployment for states, counties, and cities with a population of at least 25,000. Recent work by Forrester (2023) developed key measures of the civilian noninstitutional population — the population likely to be engaged in the labor force — by county. In this article I describe demographic methods used by LAUS research to estimate the civilian noninstitutional population for counties and county equivalents. This work aims to develop historical time series data back in time to April 2000, harmonize geographic boundaries over time, and incorporate more recent input data. The resulting data are raked to the statewide population series from the Current Population Survey to ensure additive consistency with official labor force statistics. Finally, I combine official estimates from LAUS to demonstrate experimental labor force participation rates and employment-population ratios. These data aim to provide policymakers and data users with novel measures of local labor market activity over a long historical period.

Predicting CPI in Small Areas from Sparse Survey Data

Vladislav Beresovsky, *Bureau of Labor Statistics*
Terrance D. Savitsky, *Bureau of Labor Statistics*

The Consumer Price Index (CPI) survey is designed to measure inflation by collecting quotes in sampled Core-Based Statistical Areas (CBSA) of the U.S. The current design provides for reliable estimation of relative price changes with uncertainty measures in a limited number of large self-representative (SR) CBSAs and Census Divisions. To produce estimates in other localities (i.e. states), we use area level modeling to mass impute inflation measures in all CBSAs in the U.S. Our project faces multiple challenges, including approximately estimated variances of direct estimates, data sparsity, and sampled CBSAs being poorly representative of the population. We co-model point and variance estimates in small areas to mitigate the effect of unreliably

estimated variances, employ global-local variable selection priors, and spatial modeling to compensate for sparsity and lack of representativeness of the available sample. We investigate effect of model assumptions on imputed inflation characteristics in small domains.

Validating model-based Small Domain Estimates for Occupation with Small or No Samples: A Simulated Population Approach

Xingyou Zhang, *Bureau of Labor Statistics*
Daniel Friel, *Bureau of Labor Statistics*
Erin McNulty, *Bureau of Labor Statistics*
Yu Zhang, *Bureau of Labor Statistics*

Model-based small domain (area) estimation produces reliable estimates for small domains by integrating information from both surveys and other data sources via area-level or unit-level models. How to validate model-based estimates is always challenging. A common validation strategy involves comparing statistical consistency between model-based estimates and those survey estimates for small domains with large samples. However, this approach could not really validate model-based estimates for domains with small or no samples. The validation dilemma for small domain estimation is: we don't have statistically reliable estimates for small domains, but we wanted to validate their model-based estimates directly. This study uses Occupational Requirements Survey (ORS) and explores a simulated population approach that allows us to validate model-based estimates for small domains with small and no samples.

The objective of ORS small domain estimation is to produce reliable estimates of job requirements for 844 occupations. ORS721-725 (five sample groups) had small job quotes samples (n<30) for 329 occupations. The simulated population approach for validating small domain estimates takes three basic steps. First, The entire ORS721-725 dataset is treated as a population with known job requirements, and a subsample is drawn this population according to original ORS sampling design. Second, multilevel unit-level models are developed to generate model-based estimates for all 830 occupations sampled by ORS. Third, since we know the population estimates for the job requirements from ORS721-725, we could compare the statistical accuracy (bias, precision, and correlation) of population estimates, subsample-based direct survey estimates and model-based estimates.

Return to Schedule Overview

Tuesday, October 22, 2024 2:00 PM

Session B-6: Maximizing Response Rates: Practical Solutions for Higher Engagement

Organizer: Rebecca Morrison, *Bureau of Labor Statistics*
Chair: Rebecca Morrison, *Bureau of Labor Statistics*
Room 0105

Automatically Mining for Contact Information to Improve Response Rates: A Case Study for the Quarterly Survey of Plant Capacity Utilization

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

The Quarterly Survey of Plant Capacity Utilization (QPC) is a joint effort amongst the Federal Reserve Board (FRB), the Defense Logistics Agency (DLA), and the Census Bureau to estimate plant capacity utilization rates across manufacturing industries. Some sampled businesses do not have email addresses in the QPC records and therefore do not receive important communications such as a survey notification email (usually concurrent with a physical letter), a reminder notification email, and a follow-up notification email. The goal of this project was to leverage other sources of contact information to improve response rates. For each business without an

email address for QPC, we searched hundreds of eligible email addresses in each business survey across the Census Bureau, using the most promising email for the next survey cycle based on qualities such as the recency of contact information, the relevance of the contact information survey source to QPC, the quality of contact information, and other subject matter expertise. To replicate this manual process, we built a tool that's less time and effort intensive, more systematic, and more scalable. Results suggest that finding and using email contact information from other sources can be a powerful tool to efficiently improve response rates and thus the overall statistical quality of data products.

Do you Agree to Participate? Evaluating Informed Consent Materials for the National Longitudinal Survey of Youth

Robin Kaplan, *Bureau of Labor Statistics*
Tywanquila Walker, *Bureau of Labor Statistics*
Rebecca L. Morrison, *Bureau of Labor Statistics*
Safia Abdirizak, *Bureau of Labor Statistics*

The Bureau of Labor Statistics (BLS) National Longitudinal Surveys (NLS) interview the same individuals every year or two for several decades and ask them detailed questions about their employment and other aspects of their lives including education, health, and family life. BLS is planning to begin a new youth cohort, with individuals born between 2011 to 2016, to enable studies of how the events of their lives and their labor market outcomes affect each other. One crucial but under-researched piece of the survey process is obtaining informed consent from parents and guardians and assent from youth. Effective informed consent materials facilitate participation by making the goals of the survey clear and providing key information to both youth and their parents/guardians. This research focused on improving the informed consent materials to ensure that parents, guardians, and youth understand why they were sampled for the survey, the survey's commitment to their privacy and confidentiality of their data, survey topics and procedures, survey duration and frequency, and information about linking data to other sources. Using qualitative and quantitative approaches – including web surveys, interviews, and intercept testing – we evaluated the consent/assent language with youth, parents, and guardians iteratively over multiple testing rounds. We discuss parent/guardian expectations of the informed consent process, comprehension of the form language, willingness to participate and link answers to other data sources, and differences in results between younger (11-14) and older (15-17) youth. We provide recommendations for the language and design of informed consent materials in surveys with youth respondents.

Does it Pay to Send Multiple Pre-Paid Incentives? Evidence from a Randomized Experiment

Andrew C. Chang, *Board of Governors of the Federal Reserve System*
Joanne W. Hsu, *University of Michigan*
Eva Ma, *Board of Governors of the Federal Reserve System*
Kate Bachtell, *NORC at the University of Chicago*
Micah Sjoblom, *NORC at the University of Chicago*

To encourage survey participation and improve sample representativeness, the Survey of Consumer Finances (SCF) offers an unconditional pre-paid monetary incentive and separate post-paid incentive upon survey completion. We conducted a pre-registered between-subject randomized control experiment within the 2022 SCF, with at least 1,200 households per experimental group, to examine whether changing the pre-paid incentive structure affects survey outcomes. We assess the effects of: (1) altering the total dollar value of the pre-paid incentive ("incentive effect"), (2) giving two identical pre-paid incentives holding the total dollar value fixed ("reminder effect"), and (3) offering multiple pre-paid incentives of different amounts holding the total dollar value fixed ("slope effect") on survey response rates, interviewer burden, and data quality. Our evidence indicates that a single $15 pre-paid incentive increases response rates and maintains similar levels of interviewer burden and data quality, relative to a single $5 pre-paid incentive. Splitting the $15 into two pre-paid incentives of different amounts increases interviewer burden though lengthening time in the field without improving response rates, reducing the number of contact attempts needed for a response, or improving data quality, regardless of whether the first pre-paid is larger or smaller than the second.

Improving Outreach to Underserved Populationsfor Survey Data Collection

Marina Vornovitsky, *Centers for Medicare & Medicaid Services*
Meagan Khau, *Centers for Medicare & Medicaid Services*
Jennifer Vanicek, *NORC at the University of Chicago*
Grace Maurer, *NORC at the University of Chicago*

Members of underserved populations, including people who identify as Hispanic, Black, or Asian, are historically under-represented in health survey research and expanding collection, reporting, and analysis of data on these groups is a priority area for the federal statistical community. Collecting data on these populations often requires tailored outreach and a more varied set of gaining cooperation techniques. In Fall 2023, the CMS Office of Minority Health (OMH) and the Office of Enterprise Data and Analytics (OEDA) along with NORC at the University of Chicago conducted enhanced outreach to Medicare beneficiaries sampled for the Medicare Current Beneficiary Survey (MCBS) who identify as Hispanic, Black, or Asian. As a nationally representative longitudinal survey, the MCBS is a valuable source of data on the Medicare population, including underserved populations. The goal of the data collection was to increase the number of interviews completed with beneficiaries of color, thereby increasing the sample sizes available for disparities research using MCBS data.

CMS and NORC successfully increased the total number of interviews across all three groups and achieved a response rate 10 percentage points higher than in the previous year. This presentation will describe the outreach methods used and examine their effectiveness in contributing to a higher number of completed interviews with underserved populations. The presentation will also describe materials on culturally responsive outreach that CMS and NORC developed as part of this effort.

Variation in Response to an Increased Incentive by Characteristics of National Health and Nutrition Examination Survey Households

Te-Ching Chen, *National Center for Health Statistics*
Matt Jans, *National Center for Health Statistics*
Lara Akinbami, *National Center for Health Statistics*
Damon Ogburn, *National Center for Health Statistics*
David Woodwell, *National Center for Health Statistics*
Jessica Graber, *National Center for Health Statistics*

The National Health and Nutrition Examination Survey (NHANES) samples US households and invites randomly sampled residents to participate in a health exam at a Mobile Exam Center (MEC). During the August 2021-August 2023 data collection, the exam incentive increased from $85 to $125 for participants aged 16 and older in a nonrandom subset of sampled counties. The higher incentive increased MEC participation on average. This study re-examines that finding two ways. First, we estimated multilevel logistic regression models predicting MEC completion that include a random effect for county, a fixed effect for incentive level ($85 v. $125), and various person-level and county-level controls. A positive effect of the $125 incentive on MEC participation remained after controlling for these characteristics. Second, we looked at variability in MEC response rates within the two incentive groups as a proxy for nonresponse bias risk (more variability in subgroup response rates suggests more nonresponse bias on the dimension used for subgrouping). We calculated MEC response rates for several demographic and health subgroups (e.g., for men and women separately) and observed whether response rate variance was larger in the $85 or $125 group. The variance of response rates was lower in the $125 (i.e., higher incentive) group for all characteristics we looked, except diagnosed diabetes status and interviewed mode. These findings are discussed in the context of incentive effects broadly and implications for survey practice.

Return to Schedule Overview

Session C-1: Innovative Methods and Tools for Assessing Nonresponse Bias and Survey Refusal

Organizer: Shelley Hyland, *Bureau of Justice Statistics*
Chair: Shelley Hyland*, Bureau of Justice Statistics*
Vessey 1

New Insights on the Nonresponse Bias Analysis For the National Assessment of Educational Progress

Sabrina Zhang, *Westat*
Tom Krenzke, *Westat*

A nonresponse (NR) bias analysis has been implemented in the NAEP study to identify potential sources of nonresponse bias associated with the survey estimates. It compares weighted distributions before and after the NR adjustment on key school- and student-level characteristics, separated by school type, grade, and subject. The results from recent years show that there is potential NR bias that is not addressed by the NR adjustment, which motivates us to explore other potential covariates that could be used in the NR bias analysis and potentially identify variables to improve the NR adjustment process.

In this research, several sources of data were sought and linked to NAEP data, including area-level data from NCES' Education Demographic and Geographic Estimates and CDC's Environmental Justice Index. We calculate the correlations between the outcome and each potential covariates using Pearson and ANOVA tests, and identify new covariates to evaluate nonresponse bias. We also investigate a new metric called the Explained Variation in Outcome. This model-assisted metric was computed for each of the new covariates to evaluate how much variation in the outcome is "explained" by weight adjustment variables and newly considered variables. We find that some area-level variables are correlated to the outcome. We will report on variables that could be used in the NR bias analysis to help identify the potential bias due to nonresponse, and consider their inclusion when forming NR cells to better address the NR bias.

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

Benjamin Schneider, *Westat*
Tamara Nimkoff, *Westat*
Andy Cruse, *Westat*
Anthony Fucci, *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. Federal statistical standards and guidelines increasingly mandates the analysis of potential nonresponse bias in data collection programs, and recent reports from the FCSM Nonresponse Bias Subcommittee show that federal statistical agencies have produced hundreds of nonresponse bias reports since these standards went into effect. However, many federal statistical programs rely on state agencies to collect data and assess its quality, and state-level staff often lack training and tools in how to conduct nonresponse bias 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.

Changes in Refusal Reasons Over Time in the National Health and Nutrition Examination Survey (2017-2023)

Steven Fink, *National Center for Health Statistics*
Matt Jans, *National Center for Health Statistics*
Jill Fleming, *National Center for Health Statistics*
Denise Schaar, *National Center for Health Statistics*
Andrew Caporaso, *Westat*
Jason Clark, *Westat*
George Dixon, *Westat*
Susan Genoversa, *Westat*
Minsun Riddles, *Westat*

Understanding why survey refusals occur is important for improving response rates and data quality. The National Health and Nutrition Examination Survey (NHANES) collects refusal reasons for sampled households and residents. Screener interviews are attempted with any resident age 18+ in sampled dwelling units (DUs). Randomly sampled residents are invited to complete face-to-face survey participant (SP) interviews and health exams. We examine DU screener and SP interview refusal reasons (age 18+) across five years (2017 to 2022, excluding 2020 when NHANES paused for COVID-19). We assessed: 1) are there statistically significant linear trends in dwelling unit (DU) screener and survey participant (SP) interview refusal reasons over time (2017-2022), 2) do any DU screener or SP interview reason trends change after 2019, and 3) at the SP interview level, do reasons differ across demographic characteristics? The most common reason for both DU screener and SP interview refusals was "not interested" (41% to 49% over years [DU]; 37% to 53% [SP]). Many DU (about 30%-35%) and SP (15%-35%) refusals provided no reason, underscoring difficulty obtaining meaningful reasons. The largest increase in meaningful reasons was "general distrust of government or surveys", increasing among both DU screener and SP interview refusals after 2019. SP interview reasons were associated with the age and race of the sampled resident, and other demographics to a lesser degree. Results are discussed in the context of social psychological nonresponse theory and interviewer best practices.

Analyzing Nonresponse Sample Characteristics for Contact Improvement

Liz Jeninga, *U.S. Census Bureau*
Gritiya Tanner, *U.S. Census Bureau*

The Commodity Flow Survey (CFS) is the primary source of information on the intra- and interstate movement of goods in the United States. The CFS is a shipper-based survey and collects shipment data from a variety of sampled companies across the country. While the number of sampled establishments increased 60% and the amount of data received increased greatly compared to the previous cycle, the number of establishments responding to the survey did not increase at a comparable rate. This presentation will describe the characteristics of the companies that did not respond to the 2022 CFS and will focus on analyses of the characteristics of the business, its geographic location, demographics of that geography as well as survey invitation and follow up strategies. These characteristics will be compared to responding locations of companies. The goal is to identify potential opportunities for improvements to contact strategies and may have implications for future CFS cycles and other surveys.

Exploring the Efficacy of Live Survey Methods at the National Science Foundation

James McCall, *Westat*
Kelsey Gray, *Westat*
Matthew Ring, *Westat*
Breanna Wakar, *Westat*
Richard Griffiths, *Westat*
Rahul Shrivastava, *Westat*
Yiting Long, *Westat*
Robin Ferg, *Westat*
Karmen Perry, *Westat*

The National Science Foundations has a need to collect information on the experiences of principal investigators (PIs) and reviewers who engage with the merit review process. Live survey methods, sometimes referred to as real-time data collection, aimed to collect information close to when PIs and reviewers submitted and reviewed proposals. This study explores the efficacy and challenges of live survey methods to collect data from PIs and reviewers of proposals submitted to the Division of Chemistry at NSF between September 1, 2022, and August 31, 2023. The live survey methods used in this study attempted to close the gap in temporal proximity between proposal submission or review and when respondents received their first survey request. To accomplish this, PIs received survey requests within two weeks of submitting a proposal. Reviewers received requests on the last day of a review panel session or within two weeks of submitting an ad hoc proposal review. Response rate and nonresponse bias analyses suggest live survey methods provide effective means to collect reliable survey data from the NSF PI and reviewer populations. Supplemental analysis of temporal proximity effects further suggests closing the gap between key action events and survey requests may be associated with reduced minutes to complete a survey, fewer days between survey invitation and survey initiation, and survey completion. These results have implications for improving survey data collection methods and data quality at NSF.

Return to Schedule Overview

Tuesday, October 22, 2024 3:45 PM

Session C-2: Designing an Economy-Wide Survey to Produce Accurate Estimates at Varying Levels Without Overly Burdening Respondents: The Annual Integrated Economic Survey

Organizer: Colt S. Viehdorfer, *U.S. Census Bureau*
Chair: Nikki Czaplicki*, U.S. Census Bureau*
Vessey 2

Estimating Variances for the U.S. Census Bureau's Annual Integrated Economic Survey: Challenges in Estimation with a Low Entropy Sample Design

Katherine J. Thompson, *U.S. Census Bureau*

The Annual Integrated Economic Survey (AIES) is an economy-wide survey designed to integrate and replace a suite of seven existing annual business surveys. It is designed to be easier for businesses to complete, result in better and more timely data, and allow the Census Bureau to operate more efficiently to reduce costs. The AIES sample is an unequal probability sample selected via sequential random sampling – Chromy sampling. This low entropy, unequal probability sample design presents challenges in estimating variances, particularly for replication methods. This paper will briefly introduce the AIES sampling method and then focus on the challenges of variance estimation. We consider an approximate sampling formula approach as well as a replication method (a bootstrap variation). Performance of these approaches to variance estimation are evaluated via an extensive simulation study utilizing repeated samples in selected industries from the original sampling frame.

Small Area Estimation for the Annual Integrated Economic Survey

Stephen J. Kaputa, *U.S. Census Bureau*

The U.S. Census Bureau will introduce the Annual Integrated Economic Survey (AIES) in 2024, an economy-wide survey that replaces seven independently designed annual surveys. The AIES is designed to produce national detailed industry estimates and limited industry by state estimates. Given the multi-purpose design of the sample, it is not possible to devote enough sample to every state for reliable detailed states estimates while reducing response burden. With the need for more granular estimates, the Census Bureau is researching small area models to produce detailed industry estimates for all states. This research explores the use the Fay-Herriot area-level models, which are fit using design-based domain survey estimates and frame covariates. We will discuss the model selection process which draws from covariates created from historic administrative records and Economic Census data. We conclude with an empirical example for select industries.

A Matrix Sample Design to Reduce Burden on Companies in the Manufacturing Sector of the Annual Integrated Economic Survey

Colt S. Viehdorfer, *U.S. Census Bureau*

With the introduction of the U.S. Census Bureau's Annual Integrated Economic Survey (AIES), there will be an ongoing focus on finding the right balance between satisfying user demands for expansive (and detailed) data without overburdening the sampled companies during collection. A matrix sampling design is a commonly used survey methodology technique designed to reduce respondent burden while allowing unbiased estimation (with a measurable "cost" of increased sampling variance). In a matrix (or split panel) sample, sets of questions are randomly assigned to different sampled units, so that each question has a known inclusion probability. The AIES will pilot this approach in a single sector in its first year of data collection. Companies in the AIES sample with establishments conducting business in the manufacturing sector will be asked to provide information on a suite of items for each establishment. To reduce burden on noncertainty companies in the AIES sample, mostly single-location companies, we introduce a matrix sampling design. This paper provides details on the matrix sampling design, the associated estimation and variance estimation methods, and results from a simulation study. Finally, there is discussion on how to expand this practice to other sectors and sets of items in the survey instrument for future AIES collections.

Imputing Responses for Manufacturing Establishments Using a Bayesian Mixed Model under a Matrix Sub-Sample Design

Yeng Xiong, *U.S. Census Bureau*

The U.S. Census Bureau’s new Annual Integrated Economic Survey (AIES) is an economy-wide survey that primarily collects establishment data on payroll, receipt, and employment. However, within certain industries, such as manufacturing, additional data items may be required to provide an in-depth measure of economic activity. To alleviate response burden, only a matrix sample (i.e., a random subset) of the manufacturing companies in the AIES sample is given a long-form survey. Beside the core AIES items, the long-form survey will also include data items on inventories and cost of materials, among others.

Design-based estimates of these items can be noisy, particularly for detailed domains that were not explicitly incorporated into the AIES sample design. We instead propose building a Bayesian imputation model to predict values for the missing (by design) responses to obtain estimates. This paper describes our candidate models, which make use of frame variables and survey responses from the matrix sample as potential covariates and detailed classification of industry and geography as random effects. In a simulation study of a select few manufacturing industries, our model-based estimates are generally more precise and have a smaller mean squared error. We will discuss generalizing our research to include prediction for a larger set of data items and approaches on improving our estimation.

Return to Schedule Overview

Session C-3: Data Science Methodologies and Applications

Organizer: Yang Cheng, *National Agricultural Statistics Service*
Chair: Linda J. Young, *National Agricultural Statistics Service*
Patuxent

Drift vs Shift: Decoupling Trends and Changepoint Analysis

David S. Matteson, *Cornell University and National Institute of Statistical Sciences*

We introduce a new approach for decoupling trends (drift) and changepoints (shifts) in time series. Our locally adaptive model-based approach for robustly decoupling combines Bayesian trend filtering and machine learning based regularization. An over-parameterized Bayesian dynamic linear model (DLM) is first applied to characterize drift. Then a weighted penalized likelihood estimator is paired with the estimated DLM posterior distribution to identify shifts. We show how Bayesian DLMs specified with so-called shrinkage priors can provide smooth estimates of underlying trends in the presence of complex noise components. However, their inability to shrink exactly to zero inhibits direct changepoint detection. In contrast, penalized likelihood methods are highly effective in locating changepoints. However, they require data with simple patterns in both signal and noise. The proposed decoupling approach combines the strengths of both, i.e.\ the flexibility of Bayesian DLMs with the hard thresholding property of penalized likelihood estimators, to provide changepoint analysis in complex, modern settings. The proposed framework is outlier robust and can identify a variety of changes, including in mean and slope. It is also easily extended for analysis of parameter shifts in time-varying parameter models like dynamic regressions. We illustrate the flexibility and contrast the performance and robustness of our approach with several alternative methods across a wide range of simulations and application examples.

On an Empirical Likelihood-Based Solution to Approximate Bayesian Computation Problem

Sanjay Chaudhuri, *University of Nebraska-Lincoln and National University of Singapore*
Subhroshekhar Ghosh, *National University of Singapore*
Pham Thi Kim Cuc, *National University of Singapore*

For many complex models studied in natural, engineering, and environmental sciences, it is nearly impossible to specify a likelihood for the observed data. Approximate Bayesian Computation (ABC) methods try to estimate such model parameters only by comparing the given observation and some replicates generated from the model for various input parameter values. No explicit relationship between the parameters and the data is postulated. In this article, we propose an empirical likelihood (EL) based solution to the ABC problem. By construction, our method is based on an interpretable likelihood (i.e. the EL) which is computed using estimating equations completely specified by the observed and the replicated data and a few well-chosen summary statistics. The proposed method can be justified through information projections on a specified class of densities. We further show that the posterior is consistent and discuss several of its favorable large sample and large replication properties. Illustrative examples from various real-life applications will also be presented.

Capture-Recapture in the Age of AI

Robert Emmet, *National Agricultural Statistics Service*
Luca Sartore (Presenter), *National Institute of Statistical Sciences and National Agricultural Statistics Service*
Habtamu Benecha, *National Agricultural Statistics Service*
Bruce A. Craig, *Purdue University and National Agricultural Statistics Service*

The United States Department of Agriculture's (USDA's) National Agricultural Statistics Service (NASS) maintains a list frame containing all known farms and potential farms in the U.S. During Census years, the list frame is frozen and becomes the Census Mailing List (CML) for the Census of Agriculture. The CML is incomplete, and NASS uses the June Area Survey (JAS), which utilizes an area frame, to adjust estimates for undercoverage, nonresponse, and misclassification. Capture-recapture (dual-system estimation) is used to

estimate weights assuming conditional independence of the CML and the JAS. The USDA's Farm Service Agency (FSA) maintains a list of potential farms enrolled in FSA programs, which is used in building the CML. After accounting for this dependence, the FSA list can be used with the CML and JAS in a triple-system estimator. However, NASS collects hundreds of control variables, and linear models do not capture coverage probabilities' complex relationships between covariates. Methods to model coverage using AI may improve the estimates. A triple-system estimator using neural networks, which would account for undercoverage, nonresponse, and list dependence is proposed.

Securing Confidentiality in Machine Learning Models

Ellen Galantucci, *Federal Maritime Commission*

In the realm of machine learning projects, ensuring confidentiality is paramount, particularly when dealing with sensitive data. This presentation delves into the intricacies of safeguarding confidentiality through a discussion of a machine learning initiative where stringent confidentiality protocols were implemented. I will explore the challenges encountered, the strategies employed, and the lessons learned in securing confidential data throughout the project lifecycle. From data acquisition and model development to deployment and maintenance, I will uncover practical insights and best practices for implementing effective confidentiality protocols in machine learning projects.

On A Machine Learning Framework for Studying Imbalanced Spatio-Temporal Data

Snigdhansu (Ansu) Chatterjee, *University of Maryland, Baltimore County*

Details Pending

Return to Schedule Overview

Tuesday, October 22, 2024 3:45 PM

Session C-4: Visualizing Insights: Unveiling the Power of Data

Organizer: Ian L Thomas, *U.S. Census Bureau*
Chair: Ian L Thomas*, U.S. Census Bureau*
Room 2110

A Combined Natural Language Annotation and Visualization Tool for the Exploratory Analysis of Federal Survey Response and Note Documents

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

Federal surveys produce large quantities of free text documents in the form of participant responses and operational case notes. To process this data efficiently, automated natural language processing (NLP) tools are necessary to extract or standardize information in these documents. NLP tasks can include language detection, sentiment analysis, topic modeling, and named entity recognition. Each of these tasks can help to understand a corpus of texts in a more automated fashion as opposed to reading the whole texts manually and extracting information from them. In survey specific data there are free text questions that might need further analysis by a clerical worker to analyze information that is provided by the respondents. For example, this analysis can include extracting name and address information from the texts or categorizing and finding themes in responses. However, choosing optimal annotation packages and modeling parameters, as well as integrating the results of multiple annotations to derive conclusions, is a difficult task. We develop a user-friendly, modular, and generalizable NLP annotation pipeline with integrated visualization tools that can aid users in selecting, tuning, and interpreting the results of multiple NLP annotations on a corpus of documents. Providing a visualization tool that can help investigate the results of such tasks can be beneficial to extract information with the optimal balance of automation and human interaction. This talk will go over a visualization of NLP tasks like topic and sentiment analysis as an example of how such a visualization tool can help enhance and automate survey text response analysis.

A Data-Driven Approach to Promoting Transparency in Diversity, Equity, Inclusion, and Accessibility (DEIA)

Anika Sharma, *Department of State*
Constance M. Mayer, *Department of State*
Kumar Yogeeswaran, *Department of State*

As the U.S. State Department represents America on a world stage, it is imperative to curate a workforce that draws from the nation's rich diversity and thereby strengthen foreign policy. Conducting diplomacy must involve people of all backgrounds to maximize the ability to handle the unique challenges and opportunities of the 21st Century. Subsequently, the push towards integrating Diversity, Equity, Inclusion, and Accessibility (DEIA) best practices across the federal sector has become a central feature of federal operations. However, to better understand the current DEIA landscape within an organization, it is crucial to assess data-based patterns and observations of the workforce and over time. This paper introduces the Department of State's (DoS) DEIA Demographic Baseline Report (DBR), which establishes a baseline of the demographic makeup of the Department workforce for which future progress towards creating a representative workforce can be assessed. The DBR relays bureau-by-bureau information on the current state of the workforce broken down by demographic categories such as race, ethnicity, sex, disability, grade/rank, and job series/skill codes via an interactive dashboard which provides the user a range of ways to digest the data. The DBR is groundbreaking for transparency by a leading federal agency where we further discuss challenges faced during the development and vetting stages and provide useful lessons learned. Overcoming these difficulties ultimately led to the release of the detailed DoS workforce data to the American public to help ensure the Department will hold itself accountable in recruiting a workforce which honors and includes all Americans.

So You Want To Build A Dashboard: Choosing The Right Digital Dissemination Tool

Alex Harding, *RTI International*

As data processes mature and modernize, the end stage is often dissemination. Digital modernization moves us from static reports and downloadable tables to interactive web dashboards and query/analysis tools. Choosing a technology for your data dissemination tool is a multifaceted decision – each platform, programming language and deployment destination comes with a complex web of costs and benefits. Additionally, the market for data tools is increasingly crowded, and understanding the capabilities of each tool can be a frustrating exercise in translating marketing documents into technical requirements. This presentation will include a landscape analysis of tools and methodologies for developing web-based data dissemination products; stratify them into categories based on common attributes; and provide a rubric for data stewards to guide technology selection.

Supply Chain Insights Platform (SCIP): Synthesizing Economic Data

Carla Medalia, *U.S. Census Bureau*
Krista Chan, *U.S. Census Bureau*
Kevin Li, *U.S. Census Bureau*
Jonathan Morgan, *U.S. Census Bureau*
Christian Moscardi, *U.S. Census Bureau*
Cecile Murray, *U.S. Census Bureau*

Private and public sector stakeholders demand relevant information about risk to supply chains of various goods. While the Census Bureau has a wide variety of valuable data pertaining to the supply chain, data are difficult to harmonize, contain gaps, and must be presented in ways that meet diverse stakeholder needs. To address these challenges, we developed the Supply Chain Insights Platform (SCIP), a dashboard visualization tool which allows users to quickly find information about key supply chain topics. In this presentation, we will highlight the process of engaging stakeholders from across government to transform a prototype set of data visualizations into an indispensable tool that unlocks the nation's supply chain data infrastructure. We will share the challenges we encountered and suggest strategies that other agencies can use to take a user-centered approach to combining data from siloed programs into relevant and cross-cutting products.

Return to Schedule Overview

Tuesday, October 22, 2024 3:45 PM

Session C-5: Data Insights Across Federal Statistics

Organizer: Joseph Rodhouse, *National Agricultural Statistics Service*
Chair: Joseph Rodhouse, *National Agricultural Statistics Service*
Room 2100

A Reconsideration of the Gibbs Sampler for Gaussian Mixed Linear Models

William R. Bell, *U.S. Census Bureau*

The substantial development of Bayesian inference over recent decades has owed much to the development of Markov Chain Monte Carlo algorithms, for which the Gibbs sampler has played an important role. The Gibbs sampler has been popular for Bayesian treatment of Gaussian mixed linear models, for which it seems to be a natural approach. This includes models commonly used for small area estimation such as the Fay-Herriot model. Relatively little attention seems to have been paid, however, to how well the Gibbs sampler performs for such models. We will examine the Gibbs sampler for Fay-Herriot models and show via two numerical illustrations that its performance on Gaussian mixed linear models can be quite poor, and that much better alternatives are readily available.

Annual Review Process for the Seasonal Adjustment of Metropolitan Statistical Areas

Jennifer Oh, *Bureau of Labor Statistics*

The Bureau of Labor Statistics produces monthly seasonally adjusted estimates of employment and unemployment for over 400+ Metropolitan Statistical Areas (MSA) via the Local Area Unemployment Statistics (LAUS) program using the Census Bureau's X-13ARIMA-SEATS software. After the end of the current year, when the not seasonally adjusted MSA data are revised, LAUS updates the seasonally adjusted series and reevaluates their quality.

In this talk, I will provide background information on the LAUS program, describe the annual processing procedure, and highlight some of the helpful diagnostics used to evaluate the quality of the seasonal adjustment.

Comparing Seven Approaches to Poverty Measurement in Terms of Their Relevance to Wellbeing

Kathryn O'Neill, *University of Pennsylvania*

A variety of poverty measures are frequently used across different literatures in the social sciences as well as by social service programs, government, and media. Available measures consider different resources and thresholds to define poverty. As a result, there are many different estimates of the prevalence and characteristics of poverty in the U.S., which complicates attempts to draw relationships between poverty and outcomes such as physical and psychological well-being. Measures include the Official Poverty Measure (OPM), the Supplemental Poverty Measure (SPM), relative or anchored poverty measures, and consumption- or wealth-based measures. A recent National Academies report proposed the adaptation of the SPM into a 'Principal Poverty Measure,' and recommended that this measure replace the OPM. It is well established in social science that economic deprivation, broadly defined, is negatively associated with wellbeing. In this paper, I examine which poverty measures are most closely associated with wellbeing. I compare seven different poverty measures using data from the Panel Study of Income Dynamics. These measures are assessed in terms of the strength of their relationship with several measures of wellbeing. I examine within-person, within-age, and within-cohort variation via three-way fixed effects. Household income is measured using the Cross-National Equivalent File supplement to the PSID to incorporate taxes and transfers. I find that the most

restrictive poverty measures, reflecting the greatest depth of poverty, are most strongly associated with several measures of wellbeing, including chronic health conditions. In contrast, higher thresholds and different measurement approaches better reflect wellbeing measures such as life satisfaction.

Matrix Decompositions: A Powerful Tool for Data-Driven Topic Modeling in Federal Surveys

Irina Belyaeva, *U.S. Census Bureau*

This study investigates the use of matrix decomposition techniques, namely Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF), for robust topic modeling in large-scale federal surveys. Traditional methods often struggle with the high dimensionality and noise inherent in such datasets. We propose that matrix decomposition can reduce data complexity while retaining key structures, leading to more effective topic identification. We demonstrate this approach using U.S. Census survey data. The results show that matrix decomposition techniques not only improve data interpretability but also enhance the efficiency and accuracy of topic discovery compared to traditional methods. These findings suggest that matrix decomposition can be a valuable tool for extracting nuanced and actionable insights for policy-making and social science research. This paper presents a comprehensive framework for applying matrix decomposition in topic modeling. We discuss the implications of these techniques for analyzing survey data, highlighting their potential to revolutionize data-driven decision-making in public administration.

On Several Recent Quasi-randomization Approaches to Estimation from Non-probability Samples

Julie Gershunskaya, *Bureau of Labor Statistics*
Vladislav Beresovsky, *Bureau of Labor Statistics*
Terrance D. Savitsky, *Bureau of Labor Statistics*

Randomization based sampling has been widely used as a preferred way for conducting surveys. The recent proliferation of computers and the internet has opened new opportunities for collecting and processing information. However, samples obtained without a well-defined randomization mechanism cannot be automatically regarded as representative of the population of interest. Quasi-randomization methods to inferences from non-probability samples are based on the assumption that the sample selection is governed by an underlying random mechanism that can be uncovered by combining non-probability survey data with a "reference" probability-based sample, obtained from the same target population. This talk will provide an overview and comparison of several recently developed methods for estimation of non-probability survey participation probabilities and present a simulation study.

The U.S. Census Bureau Index of Economic Activity (IDEA)

Elizabeth Marra Viehdorfer, *U.S. Census Bureau*
Valerie Pianin, *U.S. Census Bureau*
Jose Asturias, *U.S. Census Bureau*
William R. Bell, *U.S. Census Bureau*
Rebecca Hutchinson, *U.S. Census Bureau*
Tucker McElroy, *U.S. Census Bureau*
Rebecca L. Weaver, *U.S. Census Bureau*

The U.S. Census Bureau Index of Economic Activity (IDEA) is an experimental data product that was first released in February of 2023. It is constructed from 15 of the Census Bureau's primary monthly economic time series, providing a single time series reflecting the variation of the full set of component series over time. The component series are monthly measures of activity in retail and wholesale trade, manufacturing, construction, international trade, and business formations. Building on methodology successfully implemented for other economic indexes, the IDEA is constructed as the first principal component of the standardized monthly growth rates of the seasonally adjusted component series. It is a monthly measure in the sense that all the input components are monthly time series, but it is calculated daily, incorporating the most recent values from the component series for the current release month. For component series whose values have not yet been released for the current month, we predict (nowcast) their values using a multivariate autoregressive time series model.

The index is calculated using open-source software and the component series are sourced from the Census Bureau's publicly available Application Programming Interface (API) data. This presentation discusses the purpose of the index, the construction of the index by Principal Component Analysis (PCA), analysis of the span used for the PCA weight calculation including how pandemic outliers were dealt with, the nowcasting procedure and potential improvements to nowcast models, and the implementation of the IDEA as an Experimental data product.

Return to Schedule Overview

Tuesday, October 22, 2024 3:45 PM

Session C-6: Transforming Data Collection: Advancements in Field Operations and Quality Monitoring

Organizer: Melissa Cidade, *Bureau of Justice Statistics*
Chair: Melissa Cidade, *Bureau of Justice Statistics*
Room 0105

Evaluation of Field Operation Training Programs: A Systematic Review

Maxwell Hope, *U.S. Census Bureau*
Crystal Hernandez, *U.S. Census Bureau*
Lin Wang, *U.S. Census Bureau*

Enumerator training on the Nonresponsive Follow-up (NRFU) operation is a critical element for data collection quality. It is equally essential to assess the effectiveness of such training programs. We present a systematic review which examined the evaluation methodologies of adult educational interventions and training programs from 2014 to 2024, aiming to understand how effectiveness of training programs is measured and what methodologies are used. Employing the search terms "Adult training program" or "adult education program" alongside "evaluation," "assessment," "learning outcomes," "learning assessment," or "appraisal," we conducted a systematic search utilizing Google Scholar. The evaluation process involved subjective, expert judgments to ascertain the relevance of training programs vis-à-vis our research objectives. Consequently, certain domains of training, notably in the medical field, were excluded, while a wide spectrum of training programs encompassing diverse sectors such as education, parenting, and professional domains were included for analysis. Each reviewer independently reviewed the first 50 relevant articles and then discrepancies were resolved between reviewers with a 90% agreement rate. Preliminary findings reveal diverse evaluation methods including pre/post-test questionnaires, interviews, surveys, and observations, assessing criteria such as knowledge impact, program implementation, trainee satisfaction, and educational outcomes. Overall, a combination of quantitative and qualitative methodologies was frequently used to provide comprehensive assessments of training program effectiveness. Results of this systematic review will be presented, with implications extending to the enhancement of NRFU enumerator training.

Field Quality Monitoring: Findings and Expansion from the Reinterview Pilot Program

Rachel Huang, *U.S. Census Bureau*
Laura Hergert, *U.S. Census Bureau*
Elizabeth Mahoney, *U.S. Census Bureau*
Sydney England, *U.S. Census Bureau*

Reinterview is a process conducted by the U.S. Census Bureau to monitor an interviewer's quality of work and verify that the original interview was conducted and followed proper procedures. The current reinterview process uses a combination of a pre-selected sample and a supplemental selection of reinterview cases. This current approach does not ensure all interviewers are monitored through reinterview. This inconsistency can limit the effectiveness of quality control.

In 2023, the Office of Survey and Census Analytics was responsible for implementing a reinterview pilot program for two surveys. Reinterview was conducted in two ways; using supplemental reinterview for the entire reinterview workload and using a combination of a pre-selected sample of cases with supplemental reinterview. Supplemental reinterview case selection criteria was based on workload size and the presence of data anomalies. The combination of these criteria and monitoring of reinterview case outcomes helped ensure that almost all interviewers had at least one case selected and successfully reinterviewed with a particular focus on interviewers with potentially problematic data. This presentation will begin by outlining the methodology used in selecting cases for the pilot reinterview programs, followed by a detailed explanation of the implemented new process. Subsequently, the presentation will disclose the results and conclusions derived from these pilot programs.

Harnessing Paradata with Machine Learning to Inform Data Collection

Jack Zhou, *Westat*
Gizem Korkmaz, *Westat*
Ting Yan, *Westat*
Jill Carle, *Westat*
Ryan Hubbard, *Westat*
Rick Dulaney, *Westat*
Brad Edwards, *Westat*

Despite the proliferation of contact mode options such as text and email, survey response rates continue to decline, exacerbating the escalating cost of data collection. Paradata -process data collected as part of survey data- contain a vast amount of information on when and how sampled persons are contacted and the outcome of each contact attempt and serves as a valuable tool to inform responsive or adaptive designs. Leveraging the analysis of paradata holds the potential to inform more strategic contact strategies, leading to increased efficiencies in data collection processes.

In this study, we use paradata associated with the Medical Expenditure Panel Survey (MEPS) and employ unsupervised and supervised machine learning techniques to understand the relationship between contact mode sequences and outcomes, and the role of household characteristics. Specifically, we use k-means clustering to group similar households. Notable features include metrics related to contact attempts, such as the number of contacts and the distribution of contact modes, and features representing household composition such as household size, language spoken, the age and race of the head of the household. We identified distinct household groups with unique temporal patterns and demographic variations. These clusters showed potential in identifying challenging cases, unusual interviewer behavior, and effective contact strategies. Additionally, we are training supervised machine learning models, such as decision trees, to predict successful contact outcomes based on paradata. These models have demonstrated predictive powers and can provide recommendations for contact plans.

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

Elizabeth 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*
Karen Pennie, *U.S. Census Bureau*

As the nation's premier data collector, the Census Bureau bears the crucial responsibility of ensuring the accuracy and quality of its data. This intricate task entails monitoring the work of thousands of interviewers across numerous surveys, meticulously verifying adherence to established procedures. To address this challenge, the Field Quality Monitoring (FQM) Program was conceived in 2021, implementing a near real-time monitoring system utilizing paradata to detect, mitigate, and rectify potential data quality issues.

The FQM Program's reach has grown far beyond its initial purpose of monitoring a single survey. This dynamic, cross-survey system now leverages a diverse toolkit to enhance data quality across the entire data collection process. FQM's expanded role goes beyond monitoring – it actively shapes data integrity by revamping reinterview programs and working towards implementing interviewer quality ratings. This presentation delves into the progress of the FQM program, including introducing novel metrics developed for outlier detection and exploring the accompanying investigative tools. Furthermore, it evaluates the efficacy of these tools 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 Census Bureau's data, ensuring its reliability and trustworthiness for generations to come.

The Field Data Collector Labor Force: Lessons from the Pandemic

Jill Carle, *Westat*
Tammy Cook, *Westat*
Rick Dulaney, *Westat*
Brad Edwards, *Westat*
Jeannine Barget, *Westat*

In-person data collection is critical for the success of many large government-sponsored surveys. Despite response rate declines and increasing costs, the mode remains the gold standard for meeting the most rigorous survey requirements for federal survey programs, particularly as part of a multimode data collection strategy (Schober 2018). However, over the last ten years critical labor market and workforce changes, exacerbated by the pandemic, have made in-person data collection efforts prohibitive for all but the largest survey organizations. Shifting ideas about job flexibility and job satisfaction alongside the increasingly technical role and demanding nature of the job have impacted recruitment and retention for survey organizations across the U.S. and Europe (Charman et al 2024). Indeed, some smaller European countries have abandoned in-person data collection altogether for lack work to sustain a field data collector labor force.

This presentation will summarize trends in U.S. field data collector employment over the past decade. It will outline key challenges and promising practices in recruiting, training, and retaining high quality field data collectors. Additionally, it will consider broader ways to structure the field data collector labor force for continued success, including supplementing field data collection with multimode alternatives such as video interviewing and updating value propositions for respondents.

Abstract Listings for Wednesday, October 23

- **Concurrent Sessions D** **8:30 am – 10:00 am**
- **Concurrent Sessions E** **10:30 am – 12:00 pm**
- **Concurrent Sessions F** **1:45 pm – 3:15 pm**
- **Concurrent Sessions G** **3:30 pm – 5:00 pm**

Session D-1: AI-Ready Data: Promise, Practices, Policies from the Department of Commerce Experience

Organizer: Brian Quistorff, *Bureau of Economic Analysis*
Chair: Oliver Wise*, Department of Commerce*
Discussant: Oliver Wise*, Department of Commerce*
Chesapeake A

Building AI-Ready Best Practices: The Work of the Commerce Data Governance Board

Sallie Ann Keller, *U.S. Census Bureau*

In early 2024, the Commerce Data Governance Board launched the AI and Open Government Data Assets Working Group to develop guidelines for publishing Commerce data that can be consumed by GenAI technologies. The working group is particularly interested in (1) how GenAI can democratize access to Commerce data and (2) ways to mitigate risks that come when AI systems produce erroneous or fabricated results. This presentation provides an overview of the working group, progress to date, and next steps.

Unleashing the Power of Open Data: NOAA's AI Readiness Journey

Tyler Christensen, *National Oceanic and Atmospheric Administration*

The extent to which open government data can be discovered and ingested by AI technologies depends on many factors, including data quality, access, and documentation. For the last several years, NOAA has been developing and deploying solutions to make its massive repository of open data easier to use in AI applications. The agency has also built a vibrant community of practice for sharing best practices and creating new knowledge to advance AI and machine learning. This presentation highlights NOAA's AI readiness journey with lessons learned that apply to the scientific and statistical data communities.

AI-Ready Data: The Census Track to Machine Understandable Data

Kenneth Haase, *U.S. Census Bureau*

Recent breakthroughs in artificial intelligence, especially GenAI, have the potential to significantly democratize the ability to understand and draw insights for published statistical data products. Unfortunately, the foundational technology behind recent breakthroughs, Large Language Models (LLMs), consistently and constitutionally struggle with accuracy, consistency, and precision. These are exactly the qualities which are paramount to statistical data publishers. This presentation highlights work at the Census Bureau on technologies and standards for machine understandable data where the context, background, and caveats of statistics are explicitly part of the product. This will allow AI systems to present, explain, and analyze data responsibly.

Look Before You Leap: Exploring BEA Data Capture by GenAI Technologies

Paul Iwugo, *Bureau of Economic Analysis*

More and more Americans are using readily available GenAI technologies to guide their everyday decisions. It's critical that the responses from these models reflect accurate, unbiased information. As the nation's trusted source of data on the U.S. economy, BEA is exploring ways to make its statistics more useful and used by GenAI. An important first step is figuring out how these tools capture and report BEA data. This presentation presents the results from a recent project that examines how GenAI applications leverage BEA's products; highlights successes, challenges, and lessons learned; and outlines plans for building on this work.

Return to Schedule Overview

Session D-2: What's going on in there? Applications of Formal Privacy

Organizer: Doug Williams, *Bureau of Labor Statistics*
Chair: Doug Williams*, Bureau of Labor Statistics*
Chesapeake B

The Statistical Mechanics of Formal Privacy

Mark Fleischer, Ph.D., *U.S. Census Bureau*

This paper proposes a simple and general approach for modeling privacy rooted in statistical mechanics and based on the application of the the simulated annealing algorithm (SA). A histogram vector along with a weight matrix maps to a scalar objective function. By defining a specific objective function for SA, the stationary distribution of the corresponding Markov Chain corresponds to a multivariate discrete Gaussian distribution. The optimal value of this objective function corresponds to the true histogram. By using a non-zero temperature parameter, the SA algorithm moves away from the optimal (true) configuration (histogram) so that suboptimal configurations have higher probabilities. In effect, instead of adding random variates to histogram bin counts to protect privacy, we create a stochastic process that moves from the true configurations to suboptimal configurations where counts are perturbed. Candidate solutions are generated by a SWAPPER algorithm that changes a person's attributes (bin) to another attribute (bin).

This approach allows control of different levels of privacy by appropriately defining the weight matrix and the temperature control parameter in SA. Furthermore, we show that the corresponding Markov Chain satisfies the definitions of {\em differential privacy} (DP) and reveals a relationship between $(\epsilon,\delta)-$DP and $\epsilon-$DP where $(\epsilon,\delta)-$DP converges to $\epsilon-$DP as a function of iterative steps. Sampling issues regarding the stationary distribution, applications, variations, experiments and theorems are presented.

An Integer Programming Cell Suppression Algorithm – Providing Company Level Protection in One Optimization

Bei Wang, *U.S. Census Bureau*

Economic Census publications have been using cell suppression to protect sensitive information since 1992. Over the years, we made changes to the model and algorithm to achieve significant improvements limiting under-suppression and reducing over-suppression.

From a network minimum cost flow model in Fortran to a linear programming (LP) model in C++, the program transitioned from a local to a global optimization model while adding a few features in the algorithm. However, the need for improvement is a work in progress. In recent years, we have been looking into an integer programming (IP) model to obtain an overall global optimum cell suppression pattern.

Our integer programming model for cell suppression, in this research, is built with the overall data and structure in mind. It accomplishes the reduction of over-suppression in two aspects. First, it is a simultaneous model versus a sequential linear programming model. Second, it allows incorporation of company level protection in the algorithm that removes a "super cell" check from the usual linear programming approach. We also look into ways to reduce the Integer programming model to relieve computational burden.

We build an integer programming cell suppression model that provides company level protection and finds the global optimization pattern in one optimization. We use Economic Census data to examine this pattern and compare its performance with the results from the existing sequential linear programming

Optimizing Disclosure Control Methods for Data Products with Uncommon Characteristics: A Case Study on the Census of Fatal Occupational Injuries

Danny Friel, *Bureau of Labor Statistics*
Alyssa Gillen, *Bureau of Labor Statistics*
Julie Krautter, *Bureau of Labor Statistics*
Yvangelista Saastamoinen, *Bureau of Labor Statistics*

The Census of Fatal Occupational Injuries (CFOI) generates complete counts of workplace fatal injuries in the United States across sixteen characteristics. Because many CFOI cell counts are small, and because a one-unit difference in a CFOI cell count could lead to substantially different policy and research conclusions, there is a special interest in publishing only exact counts for CFOI cells. This informs a preference for suppression-based approaches to CFOI disclosure control over methods that involve noise infusion or model-based estimation.

BLS uses a hypercube algorithm to identify cells that could reveal the values of sensitive cells flagged for suppression. With billions of cells to screen across the sixteen CFOI characteristics, it is a statistical and logistical challenge to maximize the utility of the published dataset without sacrificing confidentiality protections.

Building on previously presented work focused on zero-count cells, we present new advancements in our research efforts to optimize the hypercube. Our findings demonstrate the flexibility of a hypercube approach to consider multiple quantitative and qualitative dimensions of utility, including maximizing the magnitude of published cell counts, minimizing the overall number of cell suppressions, and prioritizing certain categories among the characteristics that define cells. This blended approach also informs how disclosure control methods are evaluated.

We discuss CFOI methods in the broader context of disclosure policy, including with respect to the CIPSEA. We highlight reasons why data products like CFOI require custom solutions for disclosure control and discuss the special considerations for applying CIPSEA requirements to data products with uncommon characteristics.

What Impact Do the 2020 Census Differential Privacy Methods Have on the National Survey on Drug Use and Health Sample?

Katie Morton, *RTI International*
Rong Cai, *SAMHSA/CBHSQ*
Ahmed Khago, *SAMHSA*
Peilan C. Martin, *RTI International*
Matthew R. Williams, *RTI International*

To protect respondent confidentiality, the U.S. Census Bureau applies disclosure avoidance methods to its decennial Census data products. Beginning with the 2020 Census, the disclosure avoidance system was modernized to use differential privacy methods. These methods involve adding statistical noise to every published estimate so that no person or household can be identified with certainty. Many researchers are concerned that the differential privacy methods may limit the ability to target specific demographic subgroups in survey samples. To address these concerns, the U.S. Census Bureau released 2010 Census demonstration products that were treated with 2020 Census differential privacy methods. This paper summarizes an analysis which used the demonstration data to evaluate the impact of the differential privacy methods on the National Survey on Drug Use and Health sample. The findings suggest the treated data will have little impact on national sample yields by age and race.

Return to Schedule Overview

Session D-3: Expanding Gender and Sexuality Measurement: Insights from Comparative Survey Formats and Community Engagement

Organizer: Jennifer Truman, *Bureau of Justice Statistics*
Chair: Jennifer Truman*, Bureau of Justice Statistics*
Chesapeake C

A Community Based Approach for Creating Business Owner Demographic Questions, Including SOGI and Disability, on the Annual Business Survey

Hillary Steinberg, *U.S. Census Bureau*
Katie Beardall, *U.S. Census Bureau*

The Annual Business Survey is an important tool for understanding the demographics of business owners in the United States. However, workplace dynamics and proxy reporting may create challenges to data quality and equity. Respondent researchers at the Census Bureau conducted cognitive interviews to identify two topics that were especially complicated to report for in a workplace setting: 1) sexual orientation and gender identity (SOGI) identities and 2) disability identities of business owners. Particularly, proxies were often uncomfortable with answering these questions on behalf of owners. These topics often do not translate to a business setting. We also determined that incorporating more feedback directly from disabled business owners and LGBTQ+ business owners would benefit the survey.

We identified a need to engage with the communities of focus to ensure data quality. This would allow us to better uncover the social processes that may lead to barriers in reporting both sets of identities. Thus, we proposed community-based methods. We first engaged with internal experts at the Census Bureau who work on these topics. From there, convened two technical review panels, one on SOGI in the workplace and one on disability. These were interagency panels featuring federal employees, scholars, and advocates from community organizations. Finally, we recruited through community organizations to interview disabled business owners and LGBTQ+ identified business owners. These approaches will better inform how to collect data on both populations of business owners, and also directly engage the communities of focus to ensure more accurate counts.

A Safe and Inclusive Approach to Disseminating Statistical Information about the Non-binary Population in Canada

Claude Girard, *Statistics Canada*
France-Pascale Ménard, *Statistics Canada*

In 2022, Canada became the first country to release statistical information about its transgender and non-binary populations based on data collected from the 2021 Census of Population. Moreover, following a recent government-wide directive, Statistics Canada's surveys have begun to collect and disseminate information about gender rather than sex at birth.

Due to the small size of the transgender and non-binary populations – according to the 2021 Census, these represent 0.3% of individuals aged 15 or older in Canada – disseminating safe statistical information about them at detailed sociodemographic or geographical levels poses a challenge.

The dissemination strategy adopted for the 2021 Census, which was subsequently adapted and recommended for surveys, is centered on a new two-category gender variable (Men+ and Women+) that include non-binary individuals, and which is to be used at all but the highest dissemination levels. In this talk, we retrace the methodological considerations that have gone into creating and adopting this novel approach, which is deemed both inclusive of non-binary people and statistically safe.

Alignment Between Free Text and Close-ended Responses for Gender and Sexuality

Christina Pao, *Princeton University*
Dashram Pai, *Princeton University*
Yan Zhen Zhu, *Princeton University*

In recent years, there has been increased interest in creating expansive measures for gender identity and sexual orientation in large-scale social surveys; many of these expansive measures include either a mixed-format question (e.g., with free-text write-in option combined with closed-ended responses) or a fully open-ended option. With the diversification in question formats, we seek to answer the following question: How well do open-ended text boxes for gender identity and sexual orientation correspond to traditional (closed-ended) categorical measures?

To answer this question, we use original survey data collected in the US and UK in February 2022 (N=2518) to evaluate the within-respondent correspondence rates between free-text and multiple-choice measures of gender and sexuality. These data were collected using Lucid and had demographic quotas for race (in the US sample), age, region, education, and sex. First, using basic text-cleaning script (i.e., text sequence patterns) and hand-codes for places of ambiguity in my free-text responses, we evaluate the extent to which survey designers and pollsters should be concerned about harmonization and respondent confusion against the potential benefits of greater response diversity. Second, using the US National Center for Health Statistics' AI tool, SANDS, we predict which respondents are more likely to provide invalid responses (or nonresponses). This paper adds more broadly to the study of social measurement of gender and sexual orientation measures used in large-scale surveys.

Demographics and Proxy Reporting on Economic Surveys: Thinking Through SOGI and Race Measures on a Survey of Youth in Residential Placement

Hillary Steinberg, *U.S. Census Bureau*
Kristin Stettler, *U.S. Census Bureau*

Demographic measures are critical to understanding the diversity of the United States. However, many demographics variables are vulnerable to underreporting. In 2024, respondent researchers at the Census Bureau tested Sexual Orientation and Gender Identity (SOGI) measures for the first time, along with a revised race measure, for the Census of Juveniles in Residential Placement. This survey, conducted on behalf of the Office of Juvenile Justice and Delinquency Prevention at the Bureau of Justice Statistics, produces statistics about youth in residential placement across the United States, including those who are incarcerated. The survey is completed by staff as proxy respondents, most often using records.

We tested three questions on SOGI identities of youth: a sex assigned at birth question, a gender identity question, and a sexual orientation question. We also tested a new version of the race question, which integrated Hispanic/Latino identity, added Middle Eastern or North African identity (MENA), and allowed participants to select multiple races.

We conducted two rounds of cognitive moderated interviews with 39 typical survey respondents. We also collected data on these questions from 139 unmoderated interviews.

This presentation will discuss our findings, highlighting how proxy reporting impacted how respondents reacted to, and reported for, these SOGI and race questions. In general, we found that most participants were comfortable answering the SOGI and race questions and most had access to this information. However, it may take time for data quality to stabilize since some facilities do not currently keep these demographic characteristics in their records.

Epidemiologic Theoretical and Methodological Considerations for Transgender Population Health Research Using Youth Risk Behavior Surveillance System Data

Izraelle I. McKinnon, *Centers for Disease Control and Prevention*
Katie Labgold, *Centers for Disease Control and Prevention*
Kathleen H. Krause, *Centers for Disease Control and Prevention*

Michelle N. Carman-McClanahan, *Centers for Disease Control and Prevention*
Loredona Arrey, *Centers for Disease Control and Prevention*
Michael Goodman, *Emory University*
Nancy Brener, *Centers for Disease Control and Prevention*
Nicolas A. Suarez, *Centers for Disease Control and Prevention*

Background. The 2023 Youth Risk Behavior Surveillance System (YRBSS) presents the first nationally representative estimates of transgender identity among high school students. We provide theoretical and methodological guidance for analyzing and interpreting YRBSS data for transgender population health research.

Methods. We will discuss epidemiologic considerations for incorporating transgender data into analyses, including potential misclassification biases of biological sex and sexual identity by transgender status and implications for adjusting for these characteristics in studies comparing students who are and are not transgender.

Results. Transgender students may answer questions on sex and sexual identity in a way that is systematically different from cisgender students. In 2023, 7% of transgender students did not provide a response to the question, "What is your sex?", compared to less than 1% of cisgender students. When adjusting for grade and race, transgender students were three times more likely to experience forced sex compared to cisgender students (PR = 2.98, 95% CI [2.38-3.73]); however, this association is substantially attenuated when adjusting for sex and sexual identity, separately or together (PR = 1.5, 95% CI [1.1-2.1]). Adjustment for sex and sexual identity might induce bias due to the potential differential misclassification of these variables by transgender status.

Conclusions. Considerations and recommended approaches for using the transgender identity item in YRBSS, including incorporating sex and sexual identity, are discussed to reduce bias in analyses. We encourage thoughtful use and interpretation of YRBSS data for transgender population health research.

Return to Schedule Overview

Wednesday, October 23, 2024 8:30 AM

Session D-4: Defining Relevance for Statistical Collections and Data Products

Organizer: Lexi Shankser, *National Center for Education Statistics*
Chair: Lexi Shankser, *National Center for Education Statistics*
Vessey 1

As federal agents at Recognized Statistical Units and Statistical Officials, we are required to produce "relevant and timely" statistical information to inform decision-makers in governments, businesses, institutions, and households. However, we lack an agreed-upon framework for assessing the relevance of our work. What should such a framework look like? What are the attending data needs of that framework? This panel will bring perspectives on these questions from multiple agencies with diverse missions, yet the discussion will center on approaches or frameworks that are applicable across the federal recognized statistical agencies and units.

Panelists:

- Sharon Boivin, *National Center for Science and Engineering Statistics*
- Grace Kena, *Bureau of Justice Statistics*
- Jacob Malcom, *Department of the Interior*
- Rolf Schmitt, *Bureau of Transportation Statistics*
- Lexi Shankster, *National Center for Education Statistics*

Return to Schedule Overview

Wednesday, October 23, 2024 **8:30 AM**

Session D-5: Using Record Linkage of Administrative Records to Improve Federal Statistics

Organizer: Mark Motivans, *Bureau of Justice Statistics*
Chair: Ryan Kling, *Bureau of Justice Statistics*
Vessey 2

Data Matching in Department of Homeland Security (DHS) Enforcement Lifecycle Tool

Hongwei Zhang, *Department of Homeland Security*

The Office of Homeland Security Statistics (OHSS) Enforcement Lifecycle Tool matches transactional immigration DHS enforcement records and DOJ EOIR case information by using person-level or event identifiers from different source data. The key innovation behind the DHS OHSS Enforcement Lifecycle Tool is its ability to link records across all relevant DHS and DOJ data systems, to provide new insights into how non-citizens move through the end-to-end immigration enforcement system, and to support highly granular reporting and analysis for DHS and administration policy planning and public messaging in volatile border crisis situations. The records-matching methodology and algorithm of the tool will be presented in the panel.

Using Record Linkage to Assess Data Quality in Demographic Information Collected Across Federal Justice Agencies

Mark Motivans, *Bureau of Justice Statistics*

The Federal Justice Statistics Program (FJSP) is managed by the Bureau of Justice Statistics (BJS) and serves as the national clearinghouse of administrative federal criminal case processing data. Data are received from six agencies each year and are standardized, linked, and archived. Identifiers are used to create linked files—a set of crosswalks that provide the researcher with the ability to track persons forward (e.g., from arrest to subsequent processing stages) and backward (e.g., from sentencing to earlier stages). This presentation uses linked data to investigate the quality of person-level demographic information collected by agencies. It is recognized that data quality assessments are generally less developed for integrated and secondary-use data therefore record linkage is used in this presentation as a strategy to identify threats to data quality for the FJSP. The goal is to apply the structure and terminology of the "objectivity" dimension (accuracy, reliability, and coherence) from the FCSM's A Framework for Data Quality (2020) to an integrated and secondary-use data file and to gauge the "fitness of use" of similar demographic information collected by FJSP agencies.

Leveraging Linked Data to Investigate the Consistency of Legal Information Collected by Federal Justice Agencies

Chris Cutler, *Abt Global*
Isaiah Gerber, *Abt Global*
George Ebo Browne, *Bureau of Justice Statistics*

Although federal justice agencies often collect comparable information on federal defendants, such as type and length of sentence and offense information, it is not known how compatible or complete this information is. The Federal Justice Statistics Program (FJSP), collects data from six federal justice agencies each year, spanning arrest and booking through incarceration. Consecutive stages are linked together, permitting researchers to track individuals as they are processed through the criminal justice system. This presentation uses the Dyad Link Files to assess the consistency and completeness of sentence type and length using information recorded in the attorney (Executive Office for US Attorneys), court (Administrative Office of United States Courts), and prison (Bureau of Prisons) data with sentencing information from the United States Sentencing Commission. Furthermore, offense information is compared across linked records to assess consistency and completeness

of agency data when compared across comparable cohorts. This research highlights the need to better understand the data consistency and completeness across federal criminal justice agencies as it pertains to understanding the strengths and weaknesses of each source on conclusions drawn about sentence length and offense type.

Linking Federal Immigration Court and Federal Criminal Justice System Data

Ryan Kling, *Bureau of Justice Statistics*

A frequently discussed question about migrant individuals is how often they overlap with the criminal justice system. However, there are few comprehensive administrative datasets about migrant experiences in the justice system. One way to overcome this disconnect is to merge criminal justice data with sources of administrative data containing known migrant legal status. In this paper, we investigate methods of linking administrative data from immigration court proceedings provided by the Department of Justice (DOJ) Executive Office of Immigration Review (EOIR) with administrative federal criminal justice data collected and reported by the Bureau of Justice Statistics (BJS) Federal Justice Statistics Program (FJSP). Using linking techniques developed by the FJSP to link individuals and case information across different federal justice agencies, including arrest, prosecution, adjudication, sentencing, and imprisonment, we demonstrate how data overlap between immigration court and the federal criminal justice system and identify potential gaps in our understanding of the justice response to immigration violations.

Return to Schedule Overview

Wednesday, October 23, 2024 8:30 AM

Session D-6: Toward a More Perfect Sample: Case Studies in Sample Redesign

Organizer: Jonathan Mendelson, *Bureau of Labor Statistics*
Chair: Jonathan Mendelson*, Bureau of Labor Statistics*
Stas Kolenikov*, NORC at the University of Chicago*
Room 0105

Demographic Surveys Sample Redesign at the US Census Bureau: Update, Lessons Learned, and Next Steps

Richard Levy, *U.S. Census Bureau*
Matthew Herbstritt, *U.S. Census Bureau*

The Demographic Surveys Sample Redesign program has served historically to update samples and sampling methods for six major demographic surveys conducted by the U.S. Census Bureau: the American Housing Survey, Consumer Expenditure Surveys, Current Population Survey, National Crime Victimization Survey, National Health Interview Survey, and the Survey of Income and Program Participation. These surveys collect and disseminate information on topics such as employment, health insurance, crime victimization, participation in social programs, and housing tenure and characteristics. Updated samples reflect changes in population, not only in composition but also geography. In addition, we update sampling methods to introduce efficiencies and bring in state-of-the-art methods as feasible. Over the past decade, the program has expanded scope to address various issues such as coverage, measuring nonresponse error and bias, and assessing the business environment for the information life cycle. This presentation will give an update on the 2020 Sample Redesign, share research findings on topics related to non-sampling error, and offer research and development ideas for the 2030 Redesign cycle.

Exploring the Potential Redesign of a Traditionally In-Person Federal Survey after Pandemic-related Mode Changes and Evolving Priorities

Nicholas Davis, *NORC at the University of Chicago*
Whitney Murphy, *NORC at the University of Chicago*
Holly Cast, *NORC at the University of Chicago*
Chrystine Tadler, *NORC at the University of Chicago*
Nathaniel Poland, *NORC at the University of Chicago*

The COVID-19 pandemic forced many historically in-person federal surveys to shift data collection modes and impacted response patterns, providing incentive to reevaluate the design of sampling frames and methodologies. The Medicare Current Beneficiary Survey (MCBS), conducted by the Centers for Medicare & Medicaid Services (CMS), is a longitudinal survey with a geographically clustered sample design that aims to obtain a full picture of beneficiaries' health care services and expenditures. Since 2020, most MCBS interviews have been completed by telephone, although the sample continues to be constrained to geographically-limited sampling units – an artifact of historical in-person data collection. Increasing interest in oversampling subpopulations (e.g., racial and ethnic minorities) has further strained these sampling limitations, while efforts to reduce data collection costs have provided incentive to reexamine the survey's in-person design. This presentation discusses original research into alternative designs, including variously stratified national telephone frame designs, potential oversampling designs, and dual-frame and multimode designs. We present findings on 1) the impact of a national telephone frame on sampling, response rates, and estimate precision; 2) the suitability of direct and geography-based oversampling designs within the national framework; and 3) the potential for leveraging both telephone and in-person interviewing to enhance data quality and increase participation. The unique aspects of the MCBS provide opportunities to explore multiple considerations and trade-offs between sample design priorities. Our findings may help researchers developing other surveys understand alternative approaches to sampling amid pandemic-related and other challenges.

Innovations in Sample Design: A Comparison of Address-Based Sample and Panel Frame Sample for Federal Transportation Statistics

Sarah Kahl, *Ipsos*
Joann Lynch, *Ipsos*
Jared Coopersmith, *Ipsos*

Representative data is critical for equitable planning and investment. However, some demographics such as limited-English speakers and low-income households, are often underrepresented in travel surveys used for transportation policy. These surveys demand comprehensive travel and demographic information which, due to its personal nature and respondent burden, necessitates extensive recruitment and incentive strategies. Nonetheless, certain groups remain underrepresented, and overall response rates are declining. Increasing pressure for better representation together with the rising cost for address-based sample (ABS) recruitment have led researchers to look for other probability-based sample sources or blended sample frame approaches as a replacement or supplement to traditional single-instance ABS methods. The US Department of Transportation's National Household Travel Survey (NHTS) has pioneered new sampling strategies with its NextGen program. The 2022 NextGen NHTS, serving as a case study for innovative sample design, included a traditional address-based sample (ABS) of 7,500 households and a parallel sample of 7,500 households from the Ipsos KnowledgePanel, a probability-based panel frame sample (PFS). PFS can offer cost-savings on recruitment, better coverage of hard-to-survey households, and faster data collection. These parallel samples allow us as to directly compare the ABS and PFS results. We analyze the two sources in terms of demographic representation, data quality (e.g., trip rates, non-response), and cost. These results will inform the use of probability-based PFS for future NHTS, state and local travel surveys, as well as other national surveys conducted to generate federal statistics.

Redesign of the National Highway Traffic Safety Administration Occupant Protection Surveys

Lacey Werth, *National Highway Traffic Safety Administration*

NHTSA is updating the sample designs for two key surveys: the National Occupant Protection Use Survey

(NOPUS) in 2024 and the National Survey of the Use of Booster Seats (NSUBS) in 2025. NOPUS gathers data on restraint use, motorcyclist helmet use, and driver electronic device use, while NSUBS focuses on restraint use for children under age 13 across the United States. These updates are necessary to ensure that the survey estimates accurately reflect current population characteristics. The 2024 NOPUS will retain its 2015 Primary Sampling Unit (PSU) frame, with major stratification by Census region and urbanicity, updated using 2020 Census data. Proportional sampling, based on 2018 Vehicle Miles Traveled (VMT), will yield 60 PSUs via optimization and sequential Poisson sampling, maintaining overlap with the 2015 design. Secondary Sampling Units (SSUs) will be stratified by road type, with local rural remote roads excluded for efficiency. Similarly, the 2025 NSUBS will maintain its PSU frame and exclusions, but its stratification will be updated from Census region and child restraint use law to Census region and urbanicity. Additional South Census region stratification by child restraint use law since 2022 is added. The Measure of Size (MOS) will shift to children ages 0-9, aligning with the American Community Survey. Optimal Neyman allocation will be used to select PSUs in a way that minimizes variance while staying within cost constraints. The SSU frame will be selected before the 2025 survey, with stratification and sampling of SSUs expected to resemble the 2015 design to ensure continuity and reliability.

Return to Schedule Overview

Wednesday, October 23, 2024 10:30 AM

Session E-1: Cracking the Code to Higher Response Rates: Fresh Evidence and Innovative Approaches

Organizer: Kristin Tennyson, *Bureau of Justice Statistics*
Chair: Kristin Tennyson, *Bureau of Justice Statistics*
Chesapeake A

Which Mode Leads to a Complete? Findings from a Contact Mode Analysis

Katie Johnson, *NORC at the University of Chicago*
Quentin Brummet, *NORC at the University of Chicago*
Leah Christian, *NORC at the University of Chicago*
Natalie Wang, *NORC at the University of Chicago*
Anthony Washburn, *NORC at the University of Chicago*

Surveyors are increasingly using multiple modes of contact to reach potential survey participants, which can be especially important in longitudinal surveys where the goal is to maximize participation across multiple survey waves. However, more research is needed on which modes of outreach lead to survey completion and how this varies by demographics. The National Longitudinal Surveys of Youth (NLSY) follow respondents' labor market activities and other significant life events over time, with cohort members receiving multiple modes of contact attempts including phone calls, texting, emailing, mail, in-person visits, and outreach to family members. This presentation discusses a contact mode analysis using the 1979 and 1997 NLSY cohorts to determine the most important modes that led to interview completion. A random sample of cases were selected, stratified by prior round completion history, and interviewer comments were reviewed by coders to determine which modes were most responsible for interview completion.

Based on preliminary analysis from the latest round of the NLSY97, results indicate that modes vary in importance by respondents' completion histories. Phone outreach was the most important contact mode, likely because most interviews were conducted by phone, but was especially important for respondents who have completed interviews in every prior round. Text outreach was the next most important mode. Additionally, for cases who have missed at least the prior round, in-person outreach and contacting relatives were important contact modes. Additional analyses will explore differences in contact mode importance across cohorts by through comparisons from the NLSY79.

It's in the Bag: Using a Non-Monetary Incentive to Encourage Consistent Response in a Panel Survey

Erin Tanenbaum, *National Center for Education Statistics*
Andrew Zukerberg, *National Center for Education Statistics*
Rebecca Bielamowicz, *National Center for Education Statistics*
Ryan Iaconelli, *National Center for Education Statistics*

In this presentation, we will discuss the results of an experiment designed to increase unit response rates that was conducted as part of the 2023-24 School Pulse Panel (SPP), a monthly school-based survey sponsored by the National Center for Education Statistics. The 2023-24 SPP collection sampled approximately 4,000 public K-12 schools and surveyed sampled school leaders monthly between August 2023 and June 2024 for approximately 30 minutes each month. We conducted an experiment on hard-to-reach respondents that either never responded to the 2023-24 SPP between the September 2023 and February 2024 collections ("never responders") or that responded at least once but not to all collections during this time period ("sometimes responders"). A subsample of these hard-to-reach respondents was mailed a non-monetary incentive: a tote bag and advance letter package during the April 2024 data collection, while the remaining hard-to-reach respondents did not receive the tote bag/advance letter package. We analyze impacts on school response rates between those mailed the tote bag/advance letter package compared with the remaining never and sometimes responders that did not. Response rate findings overall as well as by school characteristics will be presented. Results from this experiment will inform strategies to increase participation among occasional respondents in panel surveys.

Hold the Phone: Examining the Factors Driving Individual and Household Nonresponse in a Mobile App Diary Survey

Lauren Miller, *Economic Research Service*
Joseph B. Rodhouse, *Economic Research Service*

The National Household Food Acquisition and Purchase Survey (FoodAPS) is a multi-day, multi-module diary survey that collects data on all foods households obtain during a typical week, as well as myriad personal data such as health information, income, and food program participation from every member of the household. Household surveys often collect data from a single person who answers on behalf of the whole household due to the difficulty of getting every member of the household to participate. The difficulty in getting individual-level responses is exacerbated when the survey is a multi-day, multi-module diary-type survey where there are multiple points of drop off. However, in 2022 USDA Economic Research Service conducted a field test of the FoodAPS survey where data was collected via native mobile app, in part to make the respondent experience better and participation from all household members easier in this complex survey. In this paper, we discuss factors that impacted response via the mobile app at the individual level and how that related to household level response across survey modules and as a whole. We also compare response rates for the primary respondent versus other household members and explore whether the design of the FoodAPS mobile app diary study reduced instances of proxy reporting and increased total participation. As one of the first federal household diary studies conducted via mobile app, this research has broad implications for statistical agencies that may be considering using mobile devices to collect responses.

Citizenship Question Effects on Noncitizen Household Response

J. David Brown, *U.S. Census Bureau*
Misty L. Heggeness, *University of Kansas*

Several small-sample studies have predicted that a citizenship question in the 2020 Census would cause a large drop in self-response rates. In contrast, minimal effects were found in Poehler et al.'s (2020) analysis of the 2019 Census Test randomized controlled trial (RCT). We reconcile these findings by analyzing associations between characteristics about the addresses in the 2019 Census Test and their response behavior by linking to independently constructed administrative data. We find significant heterogeneity in sensitivity to the citizenship question among households containing Hispanics, naturalized citizens, and noncitizens. Response drops the most for households containing noncitizens ineligible for a Social Security number (SSN). It falls more

for households with Latin American-born immigrants than those with immigrants from other countries. Response drops less for households with U.S.-born Hispanics than households with noncitizens from Latin America. Reductions in responsiveness occur not only through lower unit self-response rates, but also by increased household roster omissions and internet break-offs. The inclusion of a citizenship question increases the undercount of households with noncitizens. Households with noncitizens also have much higher citizenship question item nonresponse rates than those only containing citizens. The use of tract-level characteristics and significant heterogeneity among Hispanics, the foreign-born, and noncitizens help explain why the effects found by Poehler et al. were so small. Linking administrative microdata with the RCT data expands what we can learn from the RCT.

Assessing a Nonresponse Follow-Up Protocol for Household Probability Panel Recruitment

Michael Jackson, *SSRS*

Probability panels recruited via the address-based sampling (ABS) frame have become increasingly important vehicles for household survey research in the U.S., including for Federally sponsored studies. In panel studies, the response rate to the recruitment phase is usually the most important driver of the cumulative response rate, so the optimal recruitment protocol is the subject of considerable research interest.

One approach to increasing weighted recruitment response rates while controlling costs is to implement an intensive nonresponse follow-up (NRFU) protocol for a subsample of nonrespondents to initial contacts. Subsampled NRFU designs are deeply rooted in the non-panel literature, but their use for probability panel recruitment raises unique benefit-cost tradeoffs.

For example, are respondents reached via NRFU contacts more or less likely to join a panel? To what extent do they actually respond to subsequent panel surveys? And do NRFU contacts change behavioral and attitudinal estimates obtained from panel surveys, above what could be achieved via weighting?

To address these questions, this paper will evaluate an experimental recruitment NRFU protocol tested in spring 2024 for the SSRS Opinion Panel, a nationally representative ABS probability panel. The protocol tested involved additional mailings in priority envelopes, escalated pre- and postpaid incentives, the introduction of a hardcopy registration survey, outbound dialing to matched phone numbers, and refusal conversion efforts for non-panel-joiners. Aside from the above questions, attention will be given to the relative effectiveness of each of these elements at recruiting active panelists with distinct characteristics from those recruited via non-NRFU contacts.

[Return to Schedule Overview]

Wednesday, October 23, 2024 10:30 AM

Session E-2: Accessible Data for Young Learners: Innovative Paths to Make Federal Data Relevant to Students

Organizer: Josue DeLaRosa, *National Center for Education Statistics*
Chair: Josue DeLaRosa*, National Center for Education Statistics*
Chesapeake B

Empowering Tomorrow's Data Detectives: Promoting Statistical Literacy and Engagement among Youth

Ryne Paulose, *National Center for Health Statistics*
Juliana McAllister, *Centers for Disease Control and Prevention*
Trudi Ellerman, *Centers for Disease Control and Prevention*
Jody Derezinski Williams, *University of Maryland*
Rebecca Nichols, *American Statistical Association*

Josue DelaRosa, *National Center for Education Statistics*

There are an estimated 54 million children ages 5-17 years in the United Sates (ACS, 2022). Federal statistical agencies are tasked with disseminating relevant statistical information. To engage students and promote statistical literacy, new strategies may be needed, including STEM-focused enrichment opportunities.

The Data Detectives Camp is a week-long summer camp for rising 6th and 7th grade students to promote data science literacy through hands-on, project-based learning, enabling students to explore and understand the real-world implications of data.

The camp is conducted by the National Center for Health Statistics (NCHS) in collaboration with the American Statistical Association, Bureau of Justice Statistics, Bureau of Labor Statistics, CDC Museum, National Center for Education Statistics, and University of Maryland. The camp's curriculum integrates data from NCHS and other partners to familiarize students with essential federal statistical data and other resources. By working together, these organizations leverage their expertise and resources to effectively educate and engage students in statistical concepts.

By nurturing an interest in data and statistics, the camp strives to inspire future data scientists and build a pipeline of skilled professionals who can contribute to the field of data science and use federal statistical information effectively.

This presentation will discuss the camp's role as a unique method for promoting data literacy, increasing student engagement with federal statistical data, and inspiring future statisticians and data scientists. Through such initiatives, agencies can fulfill their mandate of disseminating relevant statistical information to all users while enhancing public engagement, particularly among youth.

Enhancing Access to Postsecondary Information: The Impact of NCES College Navigator on Student Decision-Making

Tara Lawley, *National Center for Education Statistics*

The National Center for Education Statistics (NCES) plays a key role in collecting, analyzing, and disseminating data related to the condition of education in the United States. A key part of NCES's mission is to make data accessible and useful for a wide range of stakeholders, including students and families making critical decisions about postsecondary education. The NCES College Navigator represents a significant advancement in fulfilling this mission by providing NCES's Integrated Postsecondary Education Data System (IPEDS) information in a family and student friendly, online platform that offers comprehensive information about U.S. colleges and universities in an easily navigable format.

This presentation will explore how the College Navigator supports NCES's goal of expanding the use of data by enabling students and families to make informed decisions based on a variety of IPEDS data points, including tuition costs, retention and graduation rates, financial aid availability, and program offerings. By analyzing usage patterns, we will demonstrate the tool's effectiveness in aiding students in their postsecondary selection process, thus contributing to better educational outcomes and alignment with their career aspirations. Moreover, we will discuss the innovative features of the College Navigator, such as its interactive maps, comparison tool, and personalized search capabilities, which enhance the user experience and provide tailored information that meets the diverse needs of prospective college students.

From Classroom to Community: Using Census Data to Build Statistical and Civic Literacy

Victoria R Glasier, *U.S. Census Bureau*

Knowing how to use and interpret data are the keys to making informed decisions. To be successful today, students need to have the skills to better understand the data-driven world around them. Developing statistical literacy is the key to this understanding.

In most schools, more emphasis is placed on other advanced math subjects than on statistics even though statistics education enhances critical thinking and can be applied across disciplines. Statistics skills empower students of any level and subject to analyze data about topics in their world, such as sports, fitness goals and

business success.

As the nation's leading source of data about our people and economy, the Census Bureau knows the importance of understanding these data when leaders use them to inform decisions affecting us all. But even on the micro level, understanding the work of the Census Bureau and developing civic and statistical literacy will help students tackle big and small decisions for the rest of their lives. Data are everywhere.

Students learn best when they can connect with material in a real way. Activities from the Statistics in Schools (SIS) program use the wealth of information that the Census Bureau collects to help students understand the importance of that data and connect with their communities and nation. This presentation will discuss SIS resources and outreach methods to bring real-world data and statistics into the K-12 classroom.

Tailoring BLS Information for Educational Audiences

Maya Brandon, *Bureau of Labor Statistics*
Tracy Jack, *Bureau of Labor Statistics*

The Bureau of Labor Statistics (BLS) produces key economic data, yet the technical nature of products for researchers may limit accessibility for students, educators, and families. This presentation highlights a BLS case study focused on making such data relevant and accessible for educational stakeholders, particularly through products like the Occupational Outlook Handbook (OOH) and the BLS K-12 pages.

The OOH provides detailed career information in plain language, facilitating young audiences' understanding of job prospects and educational requirements. Concurrently, the BLS K-12 pages promote economic literacy by explaining basic economic concepts and the role of statistics in everyday life. These resources are designed to make complex economic data accessible and relevant to a young demographic.

Usability testing and feedback loops assess the effectiveness of these products, ensuring they meet the educational needs of their users. This approach not only demystifies economic information but also enhances its educational value, allowing students and educators to effectively explore and understand labor market trends.

This case study exemplifies BLS's commitment to disseminating data all segments, including students, showcasing the potential of federal statistical data to inform and educate future generations. The insights from this initiative provide practical lessons for other agencies looking to improve the accessibility and impact of their data, underscoring the importance of user-centered design in government data dissemination.

Return to Schedule Overview

Wednesday, October 23, 2024 10:30 AM

Session E-3: Establishing the Establishments: Innovations in Establishment Data

Organizer: Emily Schondelmyer, *U.S. Census Bureau*
Chair: Emily Schondelmyer, *U.S. Census Bureau*
Chesapeake C

A Decade in Review: The Evolution of Online Data Collection of Establishment Surveys

Temika Holland, *U.S. Census Bureau*
Rebecca Keegan, *U.S. Census Bureau*
Melissa Cidade, *Bureau of Justice Statistics*

Establishment surveys tend to collect data that require the use of records and other data sources in order to report. They may also require that respondents convert existing data into acceptable formats or classifications.

This additional effort has implications on how business respondents interact with survey instruments, the data they report, and potentially, the quality of data obtained. With this in mind, over the past decade, the U.S. Census Bureau has made incredible strides in its online instrument designs to accommodate changes in business respondents, record keeping practices, and survey needs.

Emerging technologies like machine learning, responsive design, and automation, provide unique opportunities in survey research to keep up with growing trends. They allow surveys to introduce features that may be beneficial in reducing burden associated with reporting and improving data quality by taking out much of the guesswork. Citing examples from recent surveys at the U.S. Census Bureau, including the Annual Capital Expenditures Survey, Commodity Flow Survey, Economic Census, and the Annual Integrated Economic Survey, this presentation seeks to outline considerations and guidelines for the incorporation and evaluation of new technologies in establishment surveys. This presentation will provide a framework for effectively communicating new/unique functionality to survey respondents and tools for tracking the user experience through each survey iteration. Additionally, this presentation will discuss some of the risks associated if newly introduced technologies do not meet the needs, wants, and/or expectations of our respondents.

Federal Government Wage Indices

Ted To, *Bureau of Labor Statistics*
Travis Cyronek, *Bureau of Labor Statistics*

For nearly 50 years, the Employment Cost Index (ECI) has been providing the public with estimates of the change in employer labor costs for Private and state and local government establishments. We explore the practicality of constructing federal wage indexes, in the spirit of the ECI, using Office of Personnel Management (OPM) salary data. To accomplish this task, we aggregate OPM records into occupation and industry groups. Although these salary data have a crosswalk for mapping OPM occupation codes into the Standard Occupational Classification system, no corresponding crosswalk exists for industries. A key hurdle, therefore, involves creating a crosswalk that assigns industry codes to OPM establishments. We create this crosswalk by developing an algorithm that uses Quarterly Census of Employment and Wages data and machine-learning tools to match agencies with a unique industry. With this agency-North American Industry Classification System crosswalk, we calculate annual Laspeyres, Paasche, and Fisher wage indexes for several aggregations. The resulting wage inflation rates are plausible but may deviate substantially from the corresponding private industry and state and local wage inflation rates. Given the size of the US Government workforce, the addition of a Federal ECI would improve both coverage and accuracy of the ECI.

Presenting EHRs Data Alongside Survey Data: Insights and Challenges

Christine Caffrey, *National Center for Health Statistics*
Manisha Sengupta, *National Center for Health Statistics*
Shannon Kindilien, *National Center for Health Statistics*

Aiming to automate data sharing and integrate systems in new and innovative ways, the Division of Health Care Statistics in CDC's National Center for Health Statistics (NCHS) is using data from electronic health records (EHRs) to automate survey reporting across various health care settings. The National Post-acute and Long-term Care Study (NPALS) is a biennial study of major sectors of paid, regulated long-term care services in the United States and includes primary data collection from adult day services centers (ADSCs) and residential care communities (RCCs). To address reduced response rates and increased respondent burden, NCHS has been assessing the EHR environment in RCCs and exploring the possibility of extracting NPALS data elements from commercial EHR platforms. Although EHR adoption by RCCs is increasing, interoperability is lagging compared to other settings due to lack of incentives and an industry standard for data collection and reporting. This presentation will include (1.) a description of the use of EHRs in the residential care sector over the years; (2.) preliminary findings comparing selected NPALS estimates with estimates from a commercial EHR platform; (3.) discuss insights and challenges that came about during the process. Findings will inform the feasibility of using EHRs data to supplement survey data.

Starting up a New Survey of Government Facilities: Challenges and Lessons Learned

Ronda Britt, *National Center for Science and Engineering Statistics*

Sherri Mamon, *ICF*
Randy ZuWallack, *ICF*

The Federal Facilities Research & Development (R&D) Survey is a new survey of U.S. federal R&D facilities asking for their annual R&D expenditures and number of personnel. This survey will complete NCSES's collection of performer-reported R&D data in each sector (government, business, academia, and nonprofits). The survey was pilot tested in the fall of 2022 and a full rollout to all eligible facilities will be conducted in the fall of 2023.

The presentation will focus on the unanticipated challenges during the frame development and pilot test. We began the pilot test envisioning contacting each facility directly and requesting their individual participation. When facilities were reluctant to respond without their parent agency's involvement and/or approval, we had to pivot to establish central points of contact at each agency headquarters and work our way down from that level. We also experienced many unanticipated challenges in defining the appropriate reporting units ("facilities") across the government. Each agency has a unique structure and different ways of organizing their research-performing units. Some organize their R&D by program and the work is spread across many different facilities. There are also cases of separate laboratories or offices rolling up to larger facilities, and it was challenging to identify the best level for reporting for all agencies to achieve some degree of comparability across reporting units. Our initial vision of a survey of stand-alone, clearly identified, and comparable federal facilities had to morph into something a little more customized, while still keeping our end goals in sight.

Using Open-Ended Text to Evaluate "Other" Disposition Codes

Victoria R. Narine, *Bureau of Labor Statistics*
Josh Langeland, *Bureau of Labor Statistics*

The Occupational Employment and Wage Statistics (OEWS) program at the U.S. Bureau of Labor Statistics (BLS) uses an advance letter to inform sampled establishments of the upcoming survey request and collect contact information for the establishment's preferred respondent. In 2023, the authors reported results from an experiment conducted in a single state indicating that placing a telephone call to respondents after an advance letter is sent helps with collecting contact information. A second study replicated findings in a different state and took advantage of enhanced paradata collection that captured disposition information from the telephone calls (e.g., "left message," "spoke with business," "refusal," etc.). One category in the disposition information is "Other," which consists of open-ended text responses with descriptions of results from the telephone calls. These open-ended responses contain rich information about the business and the analysts' interaction with it; we suspected it could be leveraged to determine whether the standard disposition codes are sufficient or in need of improvement. Using text clustering to explore the content of the open-ended fields on a bigger scale, focusing on whether a) cases could be coded into the standard categories, and b) the codes themselves might be in need of future revision. In addition to sharing results from the text analysis, we will discuss how analysis of unstructured data can complement quantitative analyses in survey methodology research by providing insight into contextual factors that may not be captured in a categorical variable.

Return to Schedule Overview

Wednesday, October 23, 2024 10:30 AM

Session E-4: PANEL: Roles and Responsibilities of Statistical Officials

Organizer: William Wiatrowski, *Bureau of Labor Statistics*
Chair: William Wiatrowski, *Bureau of Labor Statistics*
Discussant: Jacob Malcom, *Department of Interior*
Vessey 1

This session will discuss the roles, responsibilities, and related aspects of Statistical Official activities, from different perspectives, possibly including:

- Current roles and responsibilities of a Statistical Official who also heads a statistical agency.
- Current roles and responsibilities of a Statistical Official in an agency that does not have a statistical agency/unit.
- Communicating the roles and responsibilities of the Statistical Official within an agency.
- Training new Statistical Officials.
- Budget and staff needed to be an effective Statistical Official.
- Possible organizational structures for the Statistical Official.
- Activities of Statistical Official around data quality, confidentiality, and data access across the agency?
- Policies and procedures needed across the agency to ensure that all agency data are relevant, timely, credible, and accurate.
- Specific areas where the Statistical Official can get involved in agency/unit activities, such as development/review of Learning Agendas.

Panelists

- William Wiatrowski, *Bureau of Labor Statistics*
- Jacob Malcom, *Department of Interior*
- Amy Ritualo, *Department of State*
- Alex Marten, *Environmental Protection Agency*
- Kevin Scott, *Bureau of Justice Statistics*

Return to Schedule Overview

Wednesday, October 23, 2024 10:30 AM

Session E-5: Creating an Integrated System of Data and Statistics on Household Income, Consumption, and Wealth: Progress on Building

Organizer: David Johnson, *National Academies of Sciences, Engineering and Medicine*
Chair: Timothy Smeeding, *University of Wisconsin, Madison*
Discussant: Katharine Abraham, *University of Maryland*
Vessey 2

Creating an Integrated System of Data and Statistics on Household Income, Consumption, and Wealth: Highlights from the Report

David Johnson, *National Academies of Sciences, Engineering and Medicine*
Timothy Smeeding, *University of Wisconsin, Madison*
Connie Citro, *National Academies of Sciences, Engineering and Medicine*

Building upon the great work that has already been done in the federal agencies, the report, Creating an Integrated System of Data and Statistics on Household Income, Consumption, and Wealth: Time to Build, highlights the potential for using multiple data sources, including surveys, tax records, state and federal administrative records, and commercial data. The report provides guidance on the appropriate definitions of household, family, and individual income, consumption, and wealth, and variations in definitions that would be useful for particular purposes, and recommends a variety of new consistent statistics on income, consumption and wealth that should be produced by the federal agencies. In addition, the report identifies the appropriate breakouts of income, consumption, and wealth statistics, population groups, level of geographic granularity, and frequency and timeliness of updated estimates. Given the variety of research and data blending that is currently in progress, the report provides multiple ways (or paths) to build this system, recognizing that different research and policy requires different data and statistics. Finally, the report documents the legal and administrative barriers to creating integrated, high-quality estimates and a data infrastructure for researchers to evaluate the impacts of public policy.

The National Experimental Wellbeing Statistics Project

Jonathan Rothbaum, *U.S. Census Bureau*

The National Experimental Wellbeing Statistics project aims to produce the best possible estimates of income and poverty given all available survey and administrative data. We link survey, decennial census, administrative, and commercial data to address measurement error in income and poverty statistics. We estimate improved income and poverty statistics (both pre-tax and post-tax inclusive of the value of non-health insurance in-kind transfers) by addressing several possible sources of bias documented in prior research. We address biases from (1) unit nonresponse through improved weights, (2) missing income information in both survey and administrative data through improved imputation, and (3) misreporting by combining or replacing survey responses with administrative information. Reducing survey error substantially affects key measures of wellbeing: addressing downward biased income and upward biased poverty estimates.

Consumption Inequality During and After the COVID-19 Pandemic

Thesia I. Garner, *Bureau of Labor Statistics*
Brett Matsumoto, *Bureau of Labor Statistics*
Jake Schild, *Bureau of Labor Statistics*

In this paper we study consumption inequality before, during, and after the COVID-19 pandemic. Our consumption measure uses data from the U.S. Consumer Expenditure Surveys augmented with additional data and imputations. This builds on the work over the past couple of years by researchers at the U.S. Bureau of Labor Statistics who have been involved in the development a series of consumption measures. We compare consumption to a measure of expenditures defined by outlays. We find that consumption inequality declined from 2019 to 2020 before increasing in 2021 and 2022. By 2022, consumption inequality was similar to 2019 levels. Outlay inequality follows a similar pattern but is less equal in 2022 compared to 2019. To better understand the drivers behind this change, we decompose the Gini coefficient by category of consumption and decompose the Theil index by household demographic characteristics. We find that the decline in overall inequality in 2020 was due to shifts away from the consumption categories that were most impacted by the pandemic, which reversed in 2021 and 2022.

What Money Can Buy: A Joint Distribution of Personal Income and Personal Consumption Expenditures

Marina Gindelsky, *Bureau of Economic Analysis*
Robert Martin, *Bureau of Labor Statistics*

The U.S. Bureau of Economic Analysis (BEA) and Bureau of Labor Statistics (BLS) have recently constructed a new joint distribution of personal income and personal consumption expenditures. As a prototype exercise for 2017 (with recently updated results to include 2004-2022), this is the first such joint distribution for the U.S., consistent with national accounts. By developing a new methodology to link these distributions, BEA and BLS will allow users to analyze the distributions of income and consumption expenditures for the same quantiles and create a distribution of savings. Though such linkages are challenging and have limitations, this exercise presents an important step in bridging that gap for national accounts.

Return to Schedule Overview

Session E-6: Sexual Orientation and Gender Identity (SOGI) Data Action Plans: Delivering on Executive Order 14075

Organizer: Christina Dragon, *National Institutes of Health*
Chair: Renee Ellis*, U.S. Census Bureau*
Room 0105

The U.S. National Science Foundation's Sexual Orientation and Gender Identity Data Action Plan

Greg Anderson, *National Science Foundation*

The U.S. National Science Foundation's (NSF's) Sexual Orientation and Gender Identity (SOGI) Data Action Plan was released publicly on December 7, 2023. The SOGI plan aligns with Executive Order 14075 on "Advancing Equality for Lesbian, Gay, Bisexual, Transgender, Queer, and Intersex (LGBTQI+) Individuals", NSF's statutory mission to "promote the progress of science; to advance national health, prosperity and welfare; and to secure the national defense", NSF's Strategic Goal to "empower STEM talent to fully participate in science and engineering", and with NSF's core value of Diversity and Inclusion. NSF's SOGI Data Action Plan describes three Learning Questions that represent foci for the NSF mission: the internal NSF workforce; public access to and engagement in NSF programs, activities, and opportunities; and the U.S. STEM community supported by NSF. Seven Evidence-Building Activities, including workforce surveys and listening sessions, data and program inventories, and new funding opportunities, will help NSF answer to the Learning Questions. The evidence gathered will improve NSF's ability to (1) make evidence-based decisions and (2) undertake Evidence-Use Activities related to its programs, policies, and/or operations. Insights from activities under NSF's SOGI Data Action Plan may also prove useful to the NSF's federal partners as they develop evidence-informed decisions pertaining to their programs.

The US Census Bureau's SOGI Data Action Plan

Amy Symens Smith, *U.S. Census Bureau*

On 4/3/2023 the Census Bureau submitted its Sexual Orientation and Gender Identity (SOGI) Data Action Plan in response to the Federal Evidence Agenda on LGBTQI+ Equity. The Plan is consistent with the Census Bureau's mission to serve as the leading source of quality data about the nation's people and economy. The Census Bureau's Household Pulse Survey (HPS) includes SOGI questions in addition to sponsored surveys such as the National Crime Victimization Survey (NCVS) and the National Health Interview Survey (NHIS). The Plan outlines a variety of evidence-building activities including testing the feasibility of adding SOGI questions on the American Community Survey (ACS), engagements with stakeholders, working groups and the international community. Evidence-building activities will be accomplished by working across the Census Bureau, with the Department of Commerce and federal partners.

US Department of Education's SOGI Data Action Plan

Elise Christopher, *National Center for Education Statistics*

The Federal Evidence Agenda on LGBTQI+ Equity included a section on Economic Security and Education topics that were used to inform the development of the SOGI Data Action Plan for the U.S. Department of Education (ED). ED's SOGI Data Action Plan contains five Learning Questions that align with the ED Learning Agenda and its Strategic Plan. These focus on: attainment; adverse experiences; supports and barriers; educator training; and educator experiences. Four Evidence-Building Activities have been identified thus far.

TBD

George Carter, *Department of Housing and Urban Development*
Gretchen Armstrong, *Department of Housing and Urban Development*

The Sexual Orientation and Gender Identity (SOGI) Action Plan for the Department of Housing and Urban Development was released publicly on August 17, 2023. HUD's SOGI Data Action Plan aligns with Executive Order 14075 on "Advancing Equality for Lesbian, Gay, Bisexual, Transgender, Queer, and Intersex (LGBTQI+) Individuals," HUD's mission to "to create strong, sustainable, inclusive communities and quality affordable homes for all," and HUD's Strategic Plan. HUD is committed to affirmatively furthering fair housing, and works to redress the Nation's long history of discriminatory housing practices "and systemic barriers to safe, accessible, and affordable housing for people of color, immigrants, individuals with disabilities, and lesbian, gay, bisexual, transgender, gender non-conforming, and queer (LGBTQ+) individuals" (86 FR 7487; HUD FY 22-26

Strategic Plan 2022: 14). HUD administers programs and develops policies to promote housing stability and housing security and to mitigate housing-related discrimination. HUD's SOGI Data Action Plan documents HUD's plans to assess, improve and monitor the housing experiences of LGBTQI+ people over time. Guided by 10 learning questions, HUD's SOGI Data Action Plan documents HUD's plans to collect SOGI data in the American Housing Survey (AHS), Homeless Management Information Systems (HMIS), and in administrative forms for recipients of HUD subsidies as well as efforts to assist HUD technical assistance (TA) providers in collecting SOGI data. In addition, the plan documents plans to incorporate SOGI measures into HUD program evaluations and housing discrimination studies. The presentation will discuss HUD's plan and progress to date on implementing the plan.

Action on SOGI Data & Delivering on EO 14075 at Department of Health and Human Services

Christina N Dragon, *Office of the Assistant Secretary for Health & National Institutes of Health*
Adrian Shanker, *Office of the Assistant Secretary for Health*
Karen Parker, *National Institutes of Health*

The Department of Health and Human Services (HHS) released their Sexual Orientation and Gender Identity (SOGI) Data Action Plan in December 2023, after it was approved by Secretary Becerra and in response to the June 2022 Executive Order 14075 issued by president Biden and the January 2023 Federal Evidence Agenda on LGBTQI+ Equity. HHS took a broad approach to addressing gaps and opportunities for more robust SOGI data collection to help drive health equity. Organized into three major sections, evidence-building activities, evidence-building infrastructure, and evidence-use activities, the HHS plan first focuses on establishing a baseline for what is already happening regarding voluntary SOGI demographic data, and where demographic measures need to be updated. Within evidence-building activities there are four major action items, the first of which was developing division level work plans from each HHS component as the primary focus for the first year's efforts. With workplans in hand, HHS components will be able to prioritize those data systems that can update their demographic data collections, including more expansive gender data beyond binary categories and adding sexual orientation questions, if appropriate. Evidence-building infrastructure includes an emphasis policies, processes, staff, and resources. HHS acknowledges that as a department we hold a significant amount of data that is collected across our components and that data is often integral to understanding the health and well-being of everyone in these United States.

Return to Schedule Overview

Session F-1: Using AI and ML to Improve Data Analytics

Organizer: Erin Babich, *Office of Homeland Security Statistics*
Chair: Erin Babich*, Office of Homeland Security Statistics*
Chesapeake A

A Semi-Supervised Active Learning Approach for Block-Status Classification

Atul Rawal, *U.S. Census Bureau*
James McCoy, *U.S. Census Bureau*
Andrew Duvall, *U.S. Census Bureau*
Elvis Martinex, *U.S. Census Bureau*

The Census Bureau, as a part of its decennial census must maintain and update all the addresses present within the United States. For the 2020 Census, in-office staff manually canvassed address coverage in every block. While this process was effective, it also brought about challenges associated with cost and time. To help aide the census bureau in labelling and classifying coverage at the block level, we propose a machine learning approach via semi-supervised learning. We present a robust machine learning solution to improve both data labeling and classification of block data to enable new data-driven insight while reducing costs and effort for data assessment. Towards this goal, we have employed an active-learning scheme to make accurate and precise classifications using <50,000 labelled bocks out of the 8,000,000+ within the country. We utilized multiple machine learning and deep learning models to make predictions on unlabeled data by training the model on the smaller set of labelled data. Predictions from all the models are then compared to pinpoint the blocks where there is a mismatch between the different models. Upon validation, the predicted data is then added to the training data before making predictions on the next subset of the data. We also discuss the different challenges associated with working on real-world data at this scale such as class-imbalance and data completeness, integrity.

Discovering STEM Researchers' Trajectories through AI-aided Data Linkage

Eric Livingston, *Capitol Technology University / Elsevier*
Wan-Ying Chang, *National Center for Science and Engineering Statistics*

Novel data assets can inform the career trajectory, productivity, and impact of STEM doctoral researchers to advance science and policymaking. The Survey of Doctorate Recipients (SDR), conducted by the National Center for Science and Engineering Statistics within the U.S. National Science Foundation, is an international longitudinal survey of U.S.-trained STEM research PhDs. Multiple innovative linkage efforts are underway to link data from SDR survey respondents to alternative sources and construct person-level research profiles that will further advance evidence-based science policymaking,

A two-phase AI-aided approach links data from SDR respondents to the Scopus abstract and citation database. Phase one employs the Apache Solr search engine to find candidate matches from Scopus Author Profiles. The Solr search engine results in 10 candidate authors per SDR respondent and detailed similarity measures. To determine whether each cluster of candidates contains any correct matches, and to address the potential for multiple Scopus profiles to be correct matches for a single respondent, Phase one data are further processed.
Phase two uses a custom AI Transformer model that learns the within-cluster patterns of similarity measures to predict the final linked data. The model is trained on a sample comprising of gold standard, human-curated data and automatically generated training data. The results are evaluated and compared with prior machine learning approaches on the same data sources to assess the strengths and limitations of this new model. Coupled with AI generated linked bibliometrics data, SDR's comprehensive education, demographics, and employment data will open a new chapter for research.

Explainable Artificial Intelligence for Bias Identification and Mitigation in Demographic Models

Atul Rawal, *U.S. Census Bureau*
Sandy L. Dietrich, *U.S. Census Bureau*
James McCoy, *U.S. Census Bureau*

Artificial Intelligence (AI) and Machine Learning (ML) have made tremendous advancements in recent decades. AI/ML models have been used in demographic research at the Census Bureau to gain insights for specific populations and research focuses. While these advanced models are certainly capable of providing novel and in-depth analysis, challenges related to bias and fairness remain a major issue. To address this issue, AI/ML models must be designed with fairness and trustworthiness as a core component of the model. Towards the fairness and trustworthiness of AI/ML models, explainable AI (XAI) has garnered interest in filling the gaps where traditional AI/ML models fall short. In this paper, we highlight the use of XAI to identify bias within AI/ML models and the datasets used for these models. We present two use-case examples for applying post-hoc explainability to traditional AI/ML classification models to highlight the bias in the models. One use-case example highlights the disparities in income based on the language spoken at home for the US population. Using an open-source dataset, we created multiple classification models, which were then evaluated for performance based on accuracy, precision, recall and F1 score. The top performing model was then chosen to apply XAI via SHAP. Using SHAP, we generated feature relevance and beeswarm plots to highlight the bias within the model and data. We conclude with a robust discussion on how XAI can be used by policy makers and practitioners to mitigate bias by taking a data driven explainable approach to policy decision making.

Using Machine Learning Methods to Identify Potential Construct Validity and Measurement Error Disparities in Health Outcomes from U.S. National Surveys

Morgan Earp, *National Center for Health Statistics*
Lauren Rossen, *National Center for Health Statistics*
Sarah Forrest, *National Center for Health Statistics*
Trent Buskirk, *Old Dominion University*

Measurement equity is critical for the accurate assessment of sex disparities in health outcomes. Estimated disparities in the prevalence of a given health outcome can be affected by whether the outcomes are measured in a health exam or self-reported in a survey based on a diagnosis from a health professional. Using data from the National Health and Nutrition Examination Survey (NHANES) which contains both measured in a health exam and self-reported survey health data, we used machine learning models to assess potential measurement error (e.g. differences between the true health exam measured outcome versus the self-reported health outcome) inequities across several chronic health conditions and behaviors. Using data on four health outcomes (e.g., diabetes, hypertension, high cholesterol, and current smoking) from NHANES (2015 through March 2020), we assessed differences between self-reported survey and health exam measured outcomes by various sociodemographic characteristics (e.g., age, sex, race/ethnicity, education, marital status, health insurance, and poverty). We used linear regression trees via the 'rpms' package in R to identify demographic subgroups with larger differences between self-reported and measured health outcomes. Identifying subgroups where measurement error may be larger or smaller could help inform future work on improving the estimation of the prevalence of chronic conditions and related health disparities.

Who Are the Careless Web Respondents Identified by Machine Learning

Ting Yan, *Westat*
Gizem Korkmaz, *Westat*
David Cantor, *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 detect 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 classify 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 one round of a web panel study. We will examine the demographic characteristics of respondents who were identified as careless respondents by at least one cluster method. We will also examine their response behaviors and their survey answers in other waves of data collection. Implications of findings will be discussed.

Return to Schedule Overview

Wednesday, October 23, 2024 1:45 PM

Session F-2: New Bayesian Methods for Statistical Data Privacy

Organizer: Jingchen (Monika) Hu, *Vassar College*
Chair: Jingchen (Monika) Hu, *Vassar College*
Chesapeake B

Bayesian Data Synthesis and the Utility-Risk Trade-Off for Mixed Epidemiological Data

Joe Feldman, *Duke University*
Daniel Kowal, *Cornell University*

Much of the micro data used for epidemiological studies contain sensitive measurements on real individuals. As a result, such micro data cannot be published out of privacy concerns, and without public access to these data, any statistical analyses originally published on them are nearly impossible to reproduce. To promote the dissemination of key datasets for analysis without jeopardizing the privacy of individuals, we introduce a cohesive Bayesian framework for the generation of fully synthetic high dimensional micro datasets of mixed categorical, binary, count, and continuous variables. This process centers around a joint Bayesian model that is simultaneously compatible with all of these data types, enabling the creation of mixed synthetic datasets through posterior predictive sampling. Furthermore, a focal point of epidemiological data analysis is the study of conditional relationships between various exposures and key outcome variables through regression analysis. We design a modified data synthesis strategy to target and preserve these conditional relationships, including both nonlinearities and interactions. The proposed techniques are deployed to create a synthetic version of a confidential dataset containing dozens of health, cognitive, and social measurements on nearly 20,000 North Carolina children.

Generating Spatially Referenced, Differentially Private Synthetic Data Using a Poisson-lognormal Approach

Harrison Quick, *University of Minnesota*

The recently proposed Poisson-gamma approach for generating differentially private data operates by modeling sensitive count data with a Poisson distribution and using a gamma distribution prior that is infused with external information to estimate the underlying expected counts. Unlike more traditional approaches for satisfying differential privacy – where privacy protections are often provided via increasing levels of additive noise – the Poisson-gamma framework satisfies differential privacy by increasing how informative the prior distribution is relative to the data. While the Poisson-gamma framework is transparent in how the model's "informativeness" can be quantified, recent work in the field of disease mapping has extended this notion to the Poisson-lognormal model specification and the commonly used conditional autoregressive model. That said, whereas the notion of model informativeness in the privacy literature is aimed at concealing the underlying data, the focus in the disease mapping literature has, to this point, been the opposite – i.e., to prevent the model from overpowering the data and thereby avoid oversmoothing the estimates. In this talk, we reverse

this objective and demonstrate how requirements for satisfying differential privacy in the Poisson-gamma framework can be extended and applied to the Poisson-lognormal framework. In doing so, we hope to ease the Poisson-gamma framework's reliance on external information by virtue of leveraging dependencies in the data (e.g., spatial structure) to produce synthetic data with high utility. To illustrate this work, we will analyze county-level heart disease death data from the state of Minnesota.

Mechanisms for Global Differential Privacy under Bayesian Data Synthesis

Terrance Savitsky, *Bureau of Labor Statistics*
Jingchen (Monika) Hu, *Vassar College*
Matthew Williams, *RTI International*

This paper introduces a new method that embeds any Bayesian model used to generate synthetic data and converts it into a differentially private (DP) mechanism. We propose an alteration of the model synthesizer to utilize a censored likelihood that induces upper and lower bounds of [exp(−/2), exp(/2)], where denotes the level of the DP guarantee. This censoring mechanism equipped with an epsilon−DP guarantee will induce distortion into the joint parameter posterior distribution by flattening or shifting the distribution towards a weakly informative prior. To minimize the distortion in the posterior distribution induced by likelihood censoring, we embed a vector-weighted pseudo posterior mechanism within the censoring mechanism. The pseudo posterior is formulated by selectively downweighting each likelihood contribution proportionally to its disclosure risk. On its own, the pseudo posterior mechanism produces a weaker asymptotic differential privacy (aDP) guarantee. After embedding in the censoring mechanism, the DP guarantee becomes strict such that it does not rely on asymptotics. We demonstrate that the pseudo posterior mechanism creates synthetic data with the highest utility at the price of a weaker, aDP guarantee, while embedding the pseudo posterior mechanism in the proposed censoring mechanism produces synthetic data with a stronger, non-asymptotic DP guarantee at the cost of slightly reduced utility. The perturbed histogram mechanism is included for comparison.

Privacy Preserving Autocoders

Matthew Williams, *RTI International*
Robert Chew, *RTI International*
Terrance Savitsky, *Bureau of Labor Statistics*

In this work, we propose a new privacy preserving machine learning method for stochastic gradient descent (SGD) algorithms based on a Bayesian approach that formulates a privatizing mechanism for the algorithm, as a whole, so that it is independent of the iterations of the SGD solver. This approach provides an asymptotic differential privacy guarantee by combining the Bayesian Pseudo Posterior Mechanism with the Gaussian Stochastic Weight Averaging (SWAG) approximation to the posterior distribution The pseudo posterior mechanism uses a vector of by-record weights, where each weight is associated with a data record and falls within the range [0, 1] where the model contributions (e.g., to a model loss function) for more highly risky cases/records (that would disproportionately drive model estimation) are down-weighted. This method selectively reduces the influence of records that pose a higher privacy risk during training. Our motivating application is the private training of machine learning (ML) autocoders, for example using pretrained large language models and fine tuning on confidential data provided to a statistical agency.

Exploring the Use of Synthetic Data to Reduce Disclosure Risks in Municipal Health Surveys

Wen Qin Deng, New York City Department of Health and Mental Hygiene
Tashema Bholanath, New York City Department of Health and Mental Hygiene
Stephen Immerwahr, New York City Department of Health and Mental Hygiene
Fangtao He, New York City Department of Health and Mental Hygiene
Nneka Lundy De La Cruz, New York City Department of Health and Mental Hygiene
Jingchen Hu, Vassar College

Dissemination of public-use data files for health surveys by government agencies holds immense value for public health and science. While government agencies have a long history in protecting the confidentiality and

privacy of survey respondents, due to the increasing availability of other data sources and technological advancement in recent years, a rigorous and systematic method for disclosure risk control of health surveys should be carefully developed and implemented before the public release of micro-level data. Using the population-level estimates through a combination of key variables, we identified observations subject to the high risk of re-identification in the confidential dataset. We explored two different solutions to mitigate the risk of re-identification – synthesis using Classification and Regression Trees and synthesis via Bayesian models – and assessed their impact on both risk and loss of utility of the resulting protected data. We evaluated the methods through an application to the 2021 New York City Community Health Survey. Synthesis via Bayesian models was chosen due to its satisfactory balance between risk reduction and data utility preservation and the final data product is a mixture of real and synthetic data. We further proposed solutions to avoid over-protecting and potentially obscuring estimates for underserved and vulnerable groups. In this presentation, we will share our experience and offer recommendations to other data curators who may encounter similar challenges when seeking to make micro-level health survey data publicly available.

Return to Schedule Overview

Wednesday, October 23, 2024 1:45 PM

Session F-3: In the Business of Understanding the Business Landscape – the Annual Business Survey

Organizer: Audrey Kindlon, *National Center for Science and Engineering Statistics*
Chair: Brian Headd*, Small Business Adminstration*
Chesapeake C

Capturing the Annual Business Survey in Synthetic Microdata: Construction and Use Cases of a Public Use File

Jorge Cisneros Paz, *Oak Ridge Institute for Science and Education*
Audrey Kindlon, *National Center for Science and Engineering Statistics*
Tim Wojan, *Oak Ridge Institute for Science and Education*
Matt Williams, *RTI International*
Jennifer Ozawa, *RTI International*
Christine Task, *Knexus Research Corporation*
DJ Streat, *Knexus Research Corporation*

Public use files (PUF) of federal survey microdata on individuals have been available for decades. However, large increases in computing power and the greater availability of Big Data have dramatically increased the probability of re-identifying anonymized data, potentially violating the pledge of confidentiality given to survey respondents. The same data science tools that increase the risk of disclosure can also be used to produce synthetic data that preserve critical moments of the empirical data but do not contain the records of any existing individual or business respondent. These synthetic data tools open new possibilities for producing microdata that will allow producing informative descriptive or multivariate analyses using PUFs that currently require access to confidential microdata. Developing public use establishment data from surveys presents unique challenges from demographic data, because there is a lack of anonymity and certain industries can be easily identified in a given geographic area. The presentation will briefly describe an algorithm used to construct a synthetic public use file based on the 2019 Annual Business Survey (ABS) and discuss various quality metrics. The ABS collects data on R&D expenditures, innovation-related data, globalization, and business owner characteristics from businesses operating in the U.S. Various use cases—either to substitute or supplement accessing confidential data—will be discussed in the context of tiered access. An objective of the presentation is to get feedback from session attendees regarding the ABS variables of greatest research interest to be considered in the public use file that aims to balance privacy and utility.

Refocusing on What We Don't Know: A Sample Redesign to Leverage Administrative Data

Sandy Peterson, *U.S. Census Bureau*
Daniel Cordes, *U.S. Census Bureau*
Stephen Hardy, *U.S. Census Bureau*
Audrey Kindlon, *National Center for Science and Engineering Statistics*

The Annual Business Survey, conducted by the Census Bureau and sponsored by the National Center for Science and Engineering Statistics, measures business owner demographics and other business characteristics such as types of customers, innovation activities, and research and development activities. Each independently selected annual sample uses administrative data sources to stratify firms by expected demographics of owners with the primary goal of estimating the total number of firms by demographic characteristics of owners. Recent advances in the availability of administrative data will allow the survey to measure demographic characteristics by directly assigning demographics from administrative data for many respondents. This research examines the implications of this improvement on the sampling methodology. We will create a simulated population and examine alternative sample designs focusing on the non-demographic characteristics measured by the survey aiming to reduce respondent burden and increase sample efficiency.

Rural Innovative Firms and Credit: Findings from the Annual Business Survey

Anil Rupasingha, *Economic Research Service*
Luyi Han, *Penn State University*
Timothy Wojan, *Oak Ridge Institute for Science and Education*
Stephan Goetz, *Penn State University*

In this paper we assess the prevalence and role of innovation in rural America as it relates to financial capital use, uptake, and availability among rural businesses, with particular emphasis on small businesses. We seek to assess whether there are systematic variations in applying for financial capital between rural and urban areas and whether these variations are related to innovation. The paper uses data from the 2018 Annual Business Survey in combination with secondary county-level data on the supply of financial capital. The survey data allow us to compare different sources of credit use by rural and urban businesses, controlling for firm specific characteristics including innovation. Results show that while access to finance from commercial banks in particular and traditional sources in general is negatively associated with innovative firms in nonmetro counties, this association is positive in metro counties.

Can It Work for Employers? The Expansion of Administrative Records Use beyond Nonemployer Demographic Statistics

Adela Luque, *U.S. Census Bureau*
Valeska Araujo, *U.S. Census Bureau & George Mason University*
Michaela Dillon, *U.S. Census Bureau*
John Earle, *U.S. Census Bureau & George Mason University*
Lorenz Eckerd, *U.S. Census Bureau*
James Noon, *U.S. Census Bureau*
Vitaliy Novik, *U.S. Census Bureau & George Washington University*
Jared Wold, *U.S. Census Bureau & George Mason University*
Samuel Young, *U.S. Census Bureau & Arizona State University*

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 embarked in the evaluation of assigning demographic characteristics to U.S. employer businesses using administrative records. In the presentation we will discuss the latest findings as well as the background, methodology, ongoing challenges, 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.

ABS Ownership Diversity and its Association with Patenting and Venture Capital Success

Gary Anderson, *National Center for Science and Engineering Statistics*
Audrey Kindlon, *National Center for Science and Engineering Statistics*
Timothy Wojan, *Oak Ridge Institute for Science and Education*
Emily Schondelmyer, *U.S. Census Bureau*

The Annual Business Survey (ABS) as the replacement for the Survey of Business Owners (SBO) serves as the principal data source for investigating business ownership of minorities, women, and immigrants. As a combination of SBO and the innovation questions formerly collected in the Business R&D and Innovation Survey (BRDIS), ABS opens new research opportunities investigating how ownership demographics are associated with innovation. One critical issue that ABS is uniquely able to investigate is the role that diversity among ownership teams plays in facilitating innovation or intermediate innovation outcomes. Earlier research using ABS identified both demographic and disciplinary diversity as strong correlates to new-to-market innovation. This research investigates the extent to which the various forms of diversity also impact tangible innovation related intermediate outcomes such as the awarding of patents or securing venture capital financing for R&D. The other major difference with the earlier work is the focus on R&D performing microbusinesses that are an essential input to radical innovation through the division of innovative labor. Evidence that disciplinary and/or demographic diversity affect the likelihood of receiving a patent or securing venture capital financing by small, high-tech start-ups may have implications for higher education, affirmative action, and immigration policy.

Return to Schedule Overview

Wednesday, October 23, 2024 1:45 PM

Session F-4: TBD

Organizer: TBD
Chair: TBD
Vessey 1

TBD

Return to Schedule Overview

Wednesday, October 23, 2024 1:45 PM

Session F-5: Appraise, Assess, & Apply: Advancing Alternative Data Use

Organizer: Jeramiah Yeksavich, *Energy Information Administration*
Chair: Jeramiah Yeksavich, *Energy Information Administration*
Vessey 2

Potential Value of Data and Free Access to These Data

Spiro Stefanou, *Economic Research Service*
Amos Golan, *American University*

In this study we develop measures of the potential value of information with an emphasis on observed information – data. Though value is a relative concept, developing approximate and applicable measures is essential. Such a measure (or set of measures) allows us to evaluate the potential value of public and privately available datasets, and the value of accessing each. There are several benefits to having such measures. First, providers of data can perform a cost-benefit analysis. Second, policy makers can better determine the benefits of different data when deciding whether to invest in its collection, production and release. The proposed measures are derived from information-theoretic principles as well as other statistics, in conjunction with relative measures based on semantic arguments. These measures are functions of attributes that can be aggregated into three basic blocks: (i) data reliability, integrity and accuracy, (ii) data quality, and (iii) potential value. Detailed empirical examples of applying these measures to three data sets, each of which is different in context, size and complexity, are provided as well.

Developing Fitness for Purpose Guidelines for Alternative Data Sources

Sarah Konya, *U.S. Census Bureau*
Rebecca Hutchinson, *U.S. Census Bureau*
Jenny Thompson, *U.S. Census Bureau*

As more alternative data sources are becoming available, survey teams are considering if and how they should be incorporated with (or replace) survey data to reduce respondent burden and traditional survey costs. To ensure a consistent approach for evaluating these alternative data sources, a team at the U.S. Census Bureau is developing Fitness for Purpose guidelines. These guidelines provide program areas with a list of questions about quality that they should consider in order to use data responsibly. The Fitness for Purpose tool offers a flexible approach that allow the evaluators to determine which aspects of data quality are important for their particular set of data. The guidelines ask questions that provide objective metadata about a data source that can be reviewed by several teams who may be interested in a dataset. They also ask use case-specific questions that are quantitative and qualitative. This presentation will discuss the guidelines, some examples of its use, and next steps.

Progress on Adopting Big Data in the US Consumer Price Index

Brendan Williams, *Bureau of Labor Statistics*

The Bureau of Labor Statistics has made significant progress in expanding its use of alternative data sources to supplement the surveys traditionally used to calculate the Consumer Price Index as originally outlined in Konny, Williams, and Friedman (2022). The adoption of a general, standardized approach to incorporating alternative data into the CPI will facilitate this transition and ensure BLS statistics accurately reflect mixed sampled and non-survey data. Previous efforts, such as crowd-sourced gasoline prices and new vehicle sales data, were implemented as ad hoc development projects that either replaced individual sample observations or entire component indexes. The introduction of an aggregation step between surveyed and non-survey price indexes would allow new data sources to be incorporated more efficiently while maintaining separate, survey-based measures. Shortly thereafter, the index for leased vehicles will be replaced with an index based on transactions data. The BLS is also evaluating the use of weight augmentation based on alternative data sources to improve the accuracy and representativity of estimates of both sampled and non-survey data.

A Diesel Motor Fuel Price Index: Incorporating Secondary Source Data into the US CPI

Sarah Niedergall, *Bureau of Labor Statistics*
Alec Harkins, *Bureau of Labor Statistics*

We introduce a new price index based on secondary source data for diesel motor fuel prices. The current component series for diesel in the Consumer Price Index is calculated from survey-based data. Our work adapts the method already in use for all types of gasoline (regular, midgrade, and premium) in the CPI calculated from a secondary data source. County-level expenditure weights are integrated into the aggregation of price change within each component geographic index area. These county-to-index area population proportions are based on Census data and the Consumer Expenditure Survey expenditure weights for diesel in each index area. The weighted price change now reflects county level differences and provides more detailed geographic estimates of price change and variance. Our work serves as an example of how the CPI is researching alternative data

sources to improve upon its traditional survey-based framework, by using a data source that supports more advanced statistical methods.

[Return to Schedule Overview]

Wednesday, October 23, 2024 **1:45 PM**

Session F-6: Assessing and Maintaining Quality in a Fast-paced Online Probability Panel Survey: Lessons from the National Center for Health Statistics' Rapid Surveys System

Organizer: Talia Kaatz, *National Center for Health Statistics*
Chair: Stephen Blumberg, *National Center for Health Statistics*
Room 0105

Quality in Questionnaire Development for a Rapid Online Panel Survey

Talia Kaatz, *National Center for Health Statistics*
Stephen Blumberg, *National Center for Health Statistics*
Grace Medley, *National Center for Health Statistics*
Paul Scanlon, *National Center for Health Statistics*
Amanda Smith, *RTI International*
Victoria Dounoucos, *RTI International*

The National Center for Health Statistics' (NCHS) Rapid Surveys System (RSS) aims to respond to demand for more timely and flexible data collection by using two commercial probability-based web panels – the NORC at the University of Chicago AmeriSpeak Panel and Ipsos' KnowledgePanel – to produce nationally-representative data on a variety of health topics. NCHS works closely with partners at the Centers for Disease Control (CDC) to quickly develop questionnaire content on a wide range of health topics, including emerging or priority data needs or topics for which other sources are out of date or of poor or unknown quality. Additionally, questionnaires include content from the National Health Interview Survey (NHIS) for use in weight calibration and benchmarking to assess reliability.

This presentation discusses pre-data collection efforts as the foundation of quality for the rest of the survey process – ensuring that questions are clear and meaningful so that resulting data will be interpretable and useful, selecting NHIS content that can provide insight into what the web panels are able to measure well and what they are not, managing quality control across multiple contractors, and implementing procedures during instrument development and programming to avoid errors and discrepancies in data collected from two separate panels. We will also discuss how frameworks of survey quality and error can be applied to the question design process when timelines preclude pre-testing and other gold-standard validation methods.

Ensuring Representativeness of a Sampled Subpopulation from a Probability Panel

Jared Coopersmith, *Ipsos*
Zachary Lewis, *Ipsos*
Randall Thomas, *Ipsos*
Mary Noorzai, *Ipsos*
Matthew Bramlett, *National Center for Health Statistics*
James Dahlhamer, *National Center for Health Statistics*
Katherine Irimata, *National Center for Health Statistics*

While the first four rounds of the National Center for Health Statistics' (NCHS) Rapid Surveys System (RSS) aimed for representativity of the U.S. adult population age 18+, the fifth round was intended to represent the population of U.S. youth aged 0-17, with information reported by a parent or guardian. Because Ipsos' KnowledgePanel and NORC's AmeriSpeak probability panels, which comprise the RSS data collection platform,

are both designed to represent adults age 18+, a non-standard sampling approach was required.

In this paper we present Ipsos's approach to produce a nationally representative sample of children with total survey error minimized — combining a targeted sample of parents of children aged 0-17, identified through panel profile data, with a general untargeted panel sample for which the presence of children 0-17 in the respondent's household, and therefore their eligibility for the survey, is unknown. Since population benchmarks for various types of parental roles (e.g., grandparents, older siblings) are lacking or different from standard high-quality government surveys like the National Health Interview Survey (NHIS), we will also discuss how this approach can help ensure improved representativity and can supplement benchmarks from other NCHS or U.S. Census Bureau surveys for this survey's qualified population (defined by relationship to the child age 0-17 in the household). We will also explore the weighting approach necessary for adjusting for sample design as well as potential non-response and coverage biases.

Navigating the Data Processing Journey from Online Probability Panel Data Collection to Public Use File

Adam Lee, *RTI International*
Emily Terlizzi, *National Center for Health Statistics*

The National Center for Health Statistics' Rapid Surveys System (RSS) utilizes two separate probability panel providers to collect data and seeks to release data as quickly as 60 days after the completion of data collection. As a result, the complexity of the data harmonization between two separate panel providers is compounded by tight timelines. This presentation outlines the post-data collection process to prepare the first round of RSS data for public release and the current process through the 3rd round of study. We will discuss the complexity posed by working with two different data collection panels, outline approaches used to identify disparities and harmonize data across panel providers, and highlight challenges encountered and lessons learned.

Combining Data from Multiple Sources: Performance of Different Classes of Estimators from Monte Carlo Simulations

Stas Kolenikov, *NORC at the University of Chicago*
Soubhik Barari, *NORC at the University of Chicago*
David Dutwin, *NORC at the University of Chicago*
Katherine Irimata, *National Center for Health Statistics*
James Dahlhamer, *National Center for Health Statistics*

As part of the National Center for Health Statistics' (NCHS) Rapid Surveys System (RSS), NORC is conducting methodological research to evaluate and develop estimation methods for combining traditional federal survey data with probability-based online panel data. While large-scale NCHS surveys like the National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES) provide high-quality data for reliable estimation, online probability panels like NORC's AmeriSpeak Panel can deliver faster and cheaper data to supplement NCHS surveys, which is especially valuable in health emergencies such as those presented by the COVID-19 pandemic.

Through Monte Carlo simulations, this research project compares the performance of different classes of estimators based on combined high-response in-person major survey data and online panel survey data. Using NHIS public use data as the finite population to sample from, we compared calibration estimators, propensity-score adjustment (PSA) estimators, small area estimators followed by model calibration, doubly robust estimators, and hybrid estimators on outcomes such as obesity, mental health status, and access to healthcare. We have identified small area estimation with model calibration as the method that produces the least biased estimates, followed by outcome-specific PSA methods. Calibration of the panel to the health outcomes from the major survey have proved to be reasonably robust to the more complex nonresponse and noncoverage scenarios. We continue working on methodological enhancements such as combining the PSA models to produce omnibus weights for the online panel data.

Assessing Bias Across Health Domains from Two Online, Orobability-based Panel Surveys: Examples from the National Center for Health Statistics Rapid Surveys System

James Dahlhamer, *National Center for Health Statistics*
Katherine Irimata, *National Center for Health Statistics*

Launched in 2023, the National Center for Health Statistics (NCHS) Rapid Surveys System (RSS) utilizes two commercial probability-based survey panels, the NORC at the University of Chicago AmeriSpeak Panel and Ipsos' KnowledgePanel, to collect and release relevant and timely data on emerging health topics. The RSS supports fielding of surveys several times a year and utilizes statistical weighting methods to leverage the strength of NCHS' core household surveys to improve the reliability of estimates. For initial RSS rounds, the questionnaires included questions from the National Health Interview Survey (NHIS) to enable benchmarking to NHIS estimates and identification of health outcomes that are measured well with the panel surveys, and those that are not.

Using NHIS estimates as gold standards, this talk presents absolute and standardized bias of individual panel provider and combined estimates for over 70 health outcomes collected during the first two rounds of RSS. Outcomes covered a variety of health topics, including mental and self-rated health, chronic health conditions, chronic pain, health behaviors, healthcare access, healthcare utilization, and social determinants of health. Initial analyses suggest that chronic health conditions and health behaviors are measured with minimal bias, while larger deviations from NHIS estimates are observed for mental and self-rated health, healthcare access, and healthcare utilization measures. We conclude by discussing the implications of our findings for future questionnaire development and possible changes to weighting procedures.

Return to Schedule Overview

Wednesday, October 23, 2024 3:30 PM

Session G-1: Best Practices and Challenges of Disclosure Reviews

Organizer: Jingchen (Monika) Hu, *Vassar College*
Chair: Jingchen (Monika) Hu, *Vassar College*
Discussant: Michael Hawes, *U.S. Census Bureau*
Chesapeake A

Disclosure Review Challenges in a Small Federal Statistical Agency

Alisha Coleman-Jensen, *Economic Research Service*
Carrie Jones, *Economic Research Service*
Molly Burress, *Economic Research Service*
Andrew Kerns, *Economic Research Service*

USDA's Economic Research Service (ERS) is a small federal statistical agency that conducts and publishes research intramurally and in collaboration with external collaborators from various institutions. Some of that research uses confidential and/or proprietary data. As such, outputs must go through disclosure review prior to release. This is a vital task to ensure data are protected and, because it is so important, challenges must be approached with careful consideration. This presentation will provide a broad overview of ERS's process for developing a Disclosure Review Board (DRB). Issues of particular concern for the DRB will also be discussed, such as the cumulative disclosure risk of many publications from a single data source, the need for two levels of disclosure review of preliminary tables and final publications, developing an agency culture around disclosure and training needs for research staff.

Safeguarding Confidentiality: Establishing Data Protection Protocols in a Small Federal Agency Venturing into Public Data Release

Ellen Galantucci, *Federal Maritime Commission*

Embarking on the journey of data dissemination for the first time poses unique challenges for small federal agencies, particularly in maintaining the confidentiality of sensitive information. This talk delves into the imperative need for small federal agencies to establish robust confidentiality protocols, particularly as they venture into the realm of public data release. I explore the challenges and considerations unique to smaller agencies, emphasizing the importance of tailored approaches to data protection. Drawing from industry best practices and regulatory standards, I outline a comprehensive framework designed to safeguard confidentiality while facilitating responsible data dissemination. By implementing these strategies, small federal agencies can confidently navigate the complexities of data disclosure, fostering transparency without compromising security.

Disclosure Review at the Bureau of Economic Analysis

Dan Yorgason, *Bureau of Economic Analysis*

The International Economics Directorate at the Bureau of Economic Analysis (BEA) conducts company-level surveys on foreign direct investment, multinational enterprises, and international trade in services. It administers a program for approved external researchers (primarily academics) to use microdata from these surveys. Because of the relatively small size of its survey collections and its external researcher program, BEA does not maintain a formal disclosure review board. This discussion will discuss BEA's approach to disclosure review of external researcher outputs. Key features of this approach include provision of a disclosure handbook and careful review of research outputs by BEA staff. BEA's disclosure review process also entails close collaboration with researchers to set expectations, educate, and provide suggestions to facilitate the disclosure review process. In addition, the discussion will discuss disclosure review-related challenges BEA faces in the absence of a formal disclosure review board, including those associated with coordinating dispersed review efforts of output of internally produced published statistics, internally produced research, and research by external researchers.

Best Practices and Challenges in Performing Disclosure Reviews at the U.S. Energy Information Administration

David Kinyon, *Energy Information Administration*

The U.S. Energy Information Administration (EIA) conducts almost 50 surveys in which at least selected data items are protected by law under the Confidential Information Protection and Statistical Efficiency Act of 2018 (CIPSEA) or exemptions from the Freedom of Information Act (FOIA). Prior to publication, we perform disclosure reviews of our data products based on protected data, as required by the Office of Management and Budget. These disclosure reviews allow us to keep our promises to survey respondents regarding the protection of their reported data. Although we publish most of our data products based on protected data in tabular format, we also publish public-use microdata files for two of our energy consumption surveys. In this presentation, we discuss best practices in performing our disclosure reviews, which are based on automated, repeatable applications of the statistical disclosure limitation methodologies that are presented in FCSM's Statistical Policy Working Paper 22. We also discuss challenges in performing disclosure reviews, including the potential for future expansion of our data sharing program to include external researchers using our protected data for statistical purposes, as well as our progress to date on establishing a Disclosure Review Board at EIA.

BLS Disclosure Review Board

Joshua Klick, *Bureau of Labor Statistics*

The DRB provides guidance, oversight and approval for disclosure limitation methods used for publication of BLS data products. These efforts are intended to protect BLS respondents from having their information disclosed unintentionally. BLS best practices include conducting formal privacy research with Penn State expert contributors and maintaining a SharePoint site to facilitate reviews and disseminate reports.

Development of a Disclosure Avoidance Process and Tool for the National Center for Science and Engineering Statistics

Shaun Genter, *National Center for Science and Engineering Statistics*
Darius Singpurwalla, *National Center for Science and Engineering Statistics*
Heather Madray, *National Center for Science and Engineering Statistics*
Keigen Rice, *NORC at the University of Chicago*
Jennifer Taub, *NORC at the University of Chicago*
Sydney Bell, *NORC at the University of Chicago*
Alexander Hass, *NORC at the University of Chicago*
Jay Breidt, *NORC at the University of Chicago*
Julie Banks, *NORC at the University of Chicago*
Benjamin Reist, *NORC at the University of Chicago*

The National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation is a federal principal statistical agency that serves as a clearinghouse for data on the state of the STEM workforce, STEM education, and funding for STEM research and development. NCSES sponsors thirteen surveys to accomplish this work, including several surveys that require disclosure protections, such as the Survey of Earned Doctorates (SED) and National Survey of College Graduates (NSCG). Although sensitive products have always been subjected to disclosure review, until recently this process was not standardized. To address this, NCSES recently instituted a new disclosure avoidance process with the guidance of the National Opinion Research Center (NORC). In this talk, we will discuss the factors that prompted NCSES to formalize our disclosure review process, some of the challenges and decision points NCSES faced when implementing the process, and some lessons learned during early implementation. As part of our talk, we will also provide a demonstration of our Disclosure Assessment Tool, a web-based application developed by NCSES in collaboration with NORC to help assess and track the disclosure status of data products.

Return to Schedule Overview

Wednesday, October 23, 2024 3:30 PM

Session G-2: Effective Strategies in the Era of Evidence Building; The Future is Now!

Organizer: Ted Horan, *Social Security Administration*
Chair: Barbara A. Downs, *U.S. Census Bureau*
Chesapeake B

Defining Relevance for Statistical Collections and Data Products

Alexandra (Lexi) Shankster, *National Center for Education Statistics*
Sharon Boivin, *National Center for Science and Engineering Statistics*
Rolf Schmitt, *Bureau of Transportation Statistics*
Grace Kena, *Bureau of Justice Statistics*
Jacob Malcom, *Department of the Interior*

How should federal statistical agencies and units measure 'relevancy' of their data and data products? As

federal agents at Recognized Statistical Units, we are required to produce "relevant and timely" statistical information to inform decision-makers in governments, businesses, institutions and households. However, we lack an agreed-upon framework for assessing the relevance of our work. What should such a framework look like? What are the attending data needs of that framework? This panel will bring perspectives on these questions from multiple agencies with diverse missions, yet the discussion will center on approaches or frameworks that are applicable across the federal recognized statistical agencies and units.

Enhancing School Safety Through the Indicators of School Crime and Safety Program: The Value of Cross-Agency Collaboration

Josue DeLaRosa, *National Center for Education Statistics*
Veronique Irwin, *National Center for Education Statistics*

Beginning as an outgrowth of the Annual Report on School Safety—a report initiated in 1997, in response to a series of violent incidents in schools that highlighted the lack of comprehensive data on school crime and violence—the Indicators of School Crime and Safety Program (ISCS) intends to provide educators, families, and students with detailed information on school crime.

The program is led by the National Center for Education Statistics (NCES) in collaboration with the Bureau of Justice Statistics (BJS). Their annual publication offers a comprehensive review that reports the current condition of K-12 and postsecondary school crime and safety across the United States. The collaboration harnesses diverse data sources, addressing gaps in single-agency capabilities and leading to a richer understanding of school safety trends.

By synthesizing data from several data sources, this partnership provides a richer, more detailed picture of the dynamics of school crime and safety. This collaborative effort also highlights the importance of a unified federal strategy towards data collection on school safety. It has demonstrated how cross-agency collaboration can lead to improved data quality, interoperability, and comprehensiveness. As such, the ISCS program is instrumental in supporting evidence-based policy, informing educational leaders, and enhancing the overall safety of schools across the nation.

This panel will review several aspects of the Indicators of School Crime and Safety Program, including advancements in data dissemination and outreach, strategic data coordination, and a review of school safety trends.

Toward a 21st Century National Data Infrastructure: Managing Privacy and Confidentiality Risks with Blended Data

Jerry Reiter, *Duke University*
Jennifer Park, *Committee on National Statistics*

Significant technical advances and policy changes have increased the availability of data that can be used to inform evidence building. Blended data, that is, combined sources of previously collected data, can improve the quality of analyses, enable new analyses, and reduce burden and cost to the public. Recent federal legislation, regulation, and guidance have broadly described the roles and responsibilities for stewardship of blended data. Yet, questions remain as the country strives to create a modern national data infrastructure.

The Committee on National Statistics of the National Academies of Sciences, Engineering, and Medicine formed a panel to address these concerns. The Panel was directed to identify privacy and confidentiality aspects of sharing and analyzing blended data that can be addressed by technical approaches, to identify research gaps in technical approaches, and to identify aspects that may require policy approaches. In addition, the panel was directed to develop frameworks for designing and evaluating integrated technical and policy approaches that can guide best practices.

The recently released report, Toward a 21st Century National Data Infrastructure: Managing Privacy and Confidentiality Risks With Blended Data, is the result of the panel's work. The report describes the unique risks to managing privacy and confidentiality in blended data, key attributes of sound plans, and a framework to guide responsible decision making.

Unleashing the Promise of Evidence-Based Decision-Making: The Power of the "3 Cs"

Alyssa Holdren, *Bureau of Economic Analysis*
Barbara Downs, *U.S. Census Bureau*
Erica Zielewski, *Office of Management and Budget*

The U.S. federal statistical system exists within a much broader data and evidence ecosystem. As trusted intermediaries between data providers and users, statistical agencies play a key role in the evolution of this ecosystem. And they have wealth of expertise, decades of experience, and new authorities and responsibilities under the Evidence Act upon which to build. By leveraging the 3 Cs—coordination, communication, and capacity building—agencies can help unleash the promise of evidence-based decision-making.

Since the adoption of the Foundations for Evidence-Based Policymaking Act of 2018, U.S. statistical agencies and units have increased their collaborations to further expand access to data for evidence building. For example, these flexibilities enabled the federal statistical system to rapidly pivot to a remote-access model for data analysis in response to the COVID-19 crisis and to develop and launch surveys to measure the impact of the pandemic on individuals and businesses. At the same time, with the Evidence Act and its evidence-building plans, statistical agencies and units also have a newfound opportunity to help answer agencies' priority learning questions and support effective mission delivery.

Key questions this presentation addresses include: How can statistical agencies coordinate better inside and outside government to provide data and evidence for decision-making? How can agencies better communicate what data are available and for what purposes? And how can statistical agencies develop partnerships and combine forces to build capacity across the data and evidence ecosystem?

[Return to Schedule Overview]

Wednesday, October 23, 2024 3:30 PM

Session G-3: Making the Invisible Visible: Employing Intersectionality as Inquiry and Praxis for Illuminating Complex Race, Gender, Class Inequalities in Health, Education and Labor for Equity-Minded Policy

Organizer: Kelvin Choi, *NIH/NIMHD*
Chair: Kelvin Choi, *NIH/NIMHD*
Discussant: Howard Hogan, *U.S. Census Bureau*
Chesapeake C

Quantifying the Wage Gap for Black Women and Its Effect Beyond the Dollar ($)

Keisha Jean Charles, *Bureau of Labor Statistics*

The U.S. gender wage gap between various gender and race groups is evident in the lower salaries of Black and Hispanic women relative to white men, who overwhelmingly have the highest male population. Many scholars have associated the gender wage gap with racial and gender discrimination. The gender wage gap worsened during and after Covid-19 for some race groups. The U.S. Bureau of Labor Statistics and U.S. Census Bureau's Current Population Survey (CPS) data shows that in 2022, white and Asian women earned 83.0 percent and 105.3 percent, respectively, of white men's salary compared to Black and Hispanic women, who earned 71.2 percent and 64.9 percent of white men's salary. This presentation focuses on the wage gap for Black women, who have historically shown the second highest post-secondary educational attainment rates coupled with a steady employment rate. But Black women have one of the largest wage and salary income gaps when compared to white men. Moreover, this wage and salary gap correlates with other inequities faced by Black women such as wealth, housing, food, and health. All these factors have had residual effects on the lives of Black women, as well as the lives of their families.

Association of Unfair Treatment in Health Care Settings with Multi-dimensional Measures of Race Among Hispanic/Latinx Adults

Dulce Gonzalez, *Urban Institute*
Michael Karpman, *Urban Institute*
Geneivive M. Kenney, *Urban Institute*
Karishma Furtado, *Urban Institute*
Marla McDaniel, *Urban Institute*
Doug Wissoker, *Urban Institute*

In this study, we explore the value of multidimensional measures of race for assessing disparities in the rates at which Hispanic/Latinx adults experience discrimination when seeking health care. We use multivariate models to assess whether measures of street race (i.e., the race that people believe others ascribe to them based on their physical appearance) and skin tone among Hispanic/Latinx adults ages 18 to 64 are associated with experiences of unfair treatment because of race in health care settings. The data used for this analysis are drawn from the Urban Institute's Health Reform Monitoring Survey and Well-Being and Basic Needs Survey, two nationally representative surveys of nonelderly adults fielded in 2022. We find that Hispanic/Latinx adults were more likely to report experiencing unfair treatment at a doctor's office, clinic, or hospital based on their race in the past 12 months if they are racialized as Black/African American or Brown (rather than white) and if they have darker skin tones, after accounting for their self-identified race and other characteristics. These findings suggest complementing measures of self-identified race with additional measures that capture the socially constructed nature of race could improve understanding of health inequities within the Hispanic/Latinx population.

Examining Shifts in the Black Latino Population in the Census Bureau's Modified Race Question

Ricardo Lowe, *The University of Texas-Austin*

This study leverages sample data from the American Community Survey (ACS) to assess how changes to the 2020 race question influence the odds of multi-race classification for Black Latinos. Our results suggest that most Black Latino origin groups exhibit higher odds of multi-race classification in the modified race question compared to the pre-2020 race question. The reclassification of single-race Black Latinos as multiracial is particularly notable for Dominicans and Hondurans. This finding suggests that the coding schemes used in the modified race question may be overclassifying single-race Black Latinos as multiracial. In fact, we find substantial drops in single-race Black Latino origin group counts between samples, which coincides with a relative surge in multi-race classifications. Ultimately, this study underscores the ethical consequences of misclassifying single-race Black Latino responses as multiracial while also emphasizing the broader methodological implications in how official changes to question wording and coding schemes lead to artificial population shifts that are not attributed to natural demographic change. These findings are particularly relevant for measurements of changes in racialized poverty within Latinx communities and beyond.

What's Your Street Race? Employing Intersectionality as Inquiry and Praxis in the Collaborative Multiracial Post-Election Survey CMP for Illuminating Inequalities

Edward D. Vargas, *Arizona State University*
Yasmiyn Irizarry, *The University of Texas, Austin*
Nancy López, *The University of New Mexico*

We use the Collaborative Multiracial Post-Election Surveys (CMPS 2016/2020) to analyze differences in outcomes among people who self-identify as Black, Indigenous, Latinx, Asian by their reported street race/perceived race. The CMPS 2016 asked Latinos (n=11,868) respondents and the CMPS 2020 asked (n=17,545) respondents about their street race as well as other detailed information about health, involvement with the justice system. Our findings show that street race Latinx individuals, who report their street race as Latino have better labor market outcomes and have a 30 percent lower odds of being unemployed relative to street race Black and Asian respondents; their unemployment rate is about equal to that of street race white respondents. Street race Latinos are about 40 percent less likely to report being stopped and questioned by the police while in a car relative to street race Black respondents, holding all else constant. When

we disaggregate street race within racial and ethnic samples, we see differences across nationality, gender, and self-reported skin color. These findings suggest that how individuals are seen by others can have real world implications with regard to social inequality despite how they themselves report their race and will be the main contributions of this work.

Racial and Ethnic Categories and the Measurement of Inequality: The Case of U.S. Latinos

Karen Lee, *DePaul University*

In this study we apply a relational approach to the measurement of race and ethnicity to investigate racial-ethnic stratification. Using representative data from the American Community Survey (2015-2019), we advance a multidimensional approach that engages in intersectional inquiry and analysis and integrates measurements that considers the analytical distinction and simultaneity of race, ethnicity, and ancestry to scrutinize labor market and educational outcomes among U.S. Latinos. Contrary to empirical expectations, conventional analyses of national data indicate that Black Latinos have higher college graduation rates and average incomes than White Latinos. However, when we account for ancestral variations within racial categories, our analysis unveils a distinct portrait of inequality. We find that White Latinos reporting Spanish, European, and South American ancestries not only surpass Black Latinos in college graduation rates by over 10% but also exhibit some of the most favorable socioeconomic outcomes across all groups. At the same time, White Latinos who report Mexican and other Central American ancestries report among the lowest socioeconomic outcomes of all groups. These results underscore the importance of intersectional analyses that considers the heterogeneity of Latinos and its implications for disaggregating critical differences and inequalities within the U.S. Latino population. Moreover, this study holds broader relevance to research on measuring and analyzing social difference and inequality.

Bridging Data Gaps in Educational Equity: An Intersectional Analysis Leveraging NPSAS

Josué (Josh) DeLaRosa, *National Center for Education Statistics*

The National Postsecondary Student Aid Study (NPSAS) provides comprehensive data that enables a detailed examination of the intersection between parental income and college enrollment. Due to unique breadth and depth, NPSAS integrates administrative and survey data, offering a robust framework for exploring how socioeconomic factors interplay with race, gender, and geography to shape educational trajectories. This presentation explores NPSAS's innovative data collection to conduct an intersectional analysis, uncovering nuanced insights into educational experiences of postsecondary students. By merging extensive administrative records from federal databases with rich survey data, NPSAS enables a level of precision and depth in analysis previously unattainable. This methodology not only fills a significant gap left by other datasets but also aligns with the Federal Committee on Statistical Methodology's (FCSM) emphasis on leveraging credible and innovative statistical methods for policy-relevant research. Furthermore, the use of an intersectional approach reveals complex dynamics at the nexus of parental income, student identity, and educational access, offering a comprehensive understanding crucial for formulating targeted policy interventions. The unique contribution of NPSAS data in illuminating these interconnections underscores the importance of continued investment in such datasets to inform efforts aimed at promoting equity in higher education. Through this analysis, the study contributes to bridging the knowledge gap, providing policymakers and educational institutions with actionable insights to address the multifaceted nature of educational disparities.

Return to Schedule Overview

Wednesday, October 23, 2024 **3:30 PM**

Session G-4: The Roles of Data Quality and Transparency in Supporting Scientific Integrity for Statistical Agencies

Organizer: Steven Klement, *U.S. Census Bureau*
Chair: John Eltinge*, U.S. Census Bureau*
Discussant: John Eltinge*, U.S. Census Bureau*
Vessey 1

Scientific Integrity
Lisa Mirel, *National Science Foundation*

I will speak about the importance of maintaining scientific integrity as we develop a shared service environment to use data for evidence building. I will also talk about the different National Secure Data Service (NSDS) Demonstration projects and what we are doing to maintain scientific integrity. This will include but not limited to the idea of toolkits, community of practice, the steps we are taking with the linkages and the creation of synthetic data.

Data Quality and Scientific Integrity
Patricia Abaroa, *Bureau of Economic Analysis*

Discussion of factors that degrade data quality (e.g., constraints on time and resources; or use of data sources that have limited documentation), and thus have a negative effect on scientific integrity and quality. In addition, we will discuss practical procedures to implement policies on scientific integrity within federal statistical agencies.

Policies on Scientific Integrity
Amy Branum, *Centers for Disease Control and Prevention*

I will discuss formal policies on scientific integrity within the U.S. government overall; within federal statistical agencies; other governments outside of the U.S.; and from the broader scientific and research communities. In addition, I will discuss ways in which principles and policies of scientific integrity have deep connections with concepts and methods related to the quality of statistical information and transparency

Return to Schedule Overview

Wednesday, October 23, 2024 **3:30 PM**

Session G-5: Evidence-based Bayesian Methods for More Precise Estimates and Useful Inference

Organizer: John Deke, *Mathematica, Inc*
Chair: Scott Cody*, Westat*
Discussant: Thomas Wei*, Institute of Education Sciences*
Vessey 2

A Strong Case for Rethinking Causal Inference
John Deke, *Mathematica, Inc.*

In this paper I examine the root cause of, and suggest a solution to, recurring inferential errors in the field of impact evaluations, both experimental and non-experimental. These errors have been called promising trials

bias, the winner's curse, and Type-M/S errors. I argue the root cause of these errors is that researchers often attempt to draw Bayesian inferences using non-Bayesian inferential statistics. I recommend using a "best of both worlds" framework called BASIE for drawing causal inferences that is part frequentist and part Bayesian. In this talk, I describe BASIE, how it addresses the aforementioned inferential errors, and how it can also be used for statistical inference for non-causal research questions.

Advancing Equity through Enhanced Performance Measurement: A Case Study with the Pennsylvania Department of Education

Lena Rosendahl, *Mathematica, Inc.*
Brian Gill, *Mathematica, Inc.*
Lauren Forrow, *Mathematica, Inc.*
Jennifer Starling, *Mathematica, Inc.*

The Every Student Succeeds Act requires states to identify schools with low performing student subgroups for Targeted Support and Improvement or Additional Targeted Support and Improvement. Measurement error reduces the statistical reliability of the performance measures used to identify schools for these categorizations, introducing a risk that the identified schools are unlucky rather than truly low performing.

This study assessed the extent to which Bayesian hierarchical modeling can increase the reliability of performance measures, specifically academic proficiency rates. The Pennsylvania Department of Education (PDE) provided school-by-subgroup-level proficiency rates for analysis.

The authors fit two Bayesian hierarchical models, mirroring the processes PDE uses to identify schools for targeted supports. The authors extracted stabilized proficiency rates from the fitted models and compared the stabilized proficiency rates to the unstabilized rates. Specifically, the authors assessed the relationship between the number of students in the subgroup and the variation in proficiency rates in the stabilized and unstabilized data, as a proxy for statistical reliability.

In the unstabilized data, a sharp decrease in variation with increasing sample size signals the influence of measurement error on the estimated proficiency rates for smaller schools and subgroups, while in stabilized data, relatively constant variation across sample sizes indicates the expected increase in statistical reliability. These results suggest that stabilizing school performance measures using hierarchical modeling could improve their usefulness for school accountability, while at the same time including smaller schools and subgroups than the processes currently accommodate.

Assessing the Assessment: Reinterpreting Changes in State- and District-Level NAEP Scores Using a Hierarchical Bayesian Approach

Jennifer Starling, *Mathematica, Inc.*
Lauren Forrow, *Mathematica, Inc.*
Jon Gellar, *Mathematica, Inc.*
Brian Gill, *Mathematica, Inc.*

The 2022 release of the National Assessment of Education Progress (NAEP), the first assessment since the COVID pandemic, showed nationwide declines in test scores that raised concerns among educators, policymakers, researchers, and parents. Many news outlets naturally looked for bright spots, focusing on states or districts that did not show statistically significant declines in scores. However, statistical significance is easily misunderstood, and some public reports reflected this widespread communication challenge. The purpose of this study is to reinterpret the changes in state- and district-level NAEP scores between 2019 and 2022 using Bayesian hierarchical modeling, both to stabilize the estimated changes and to take advantage of the Bayesian framework in communicating the results. The authors fit a Bayesian model that partially pools across states or districts, as well as subjects (math and reading) and grade levels (4th and 8th grade) within a state or district, to stabilize the estimated changes in scores. They then interpreted the estimated changes using probability statements anchored in score changes that, according to recent research, correspond to meaningful learning loss for students. The Bayesian estimates reaffirm the prevalence and severity of declines in NAEP scores across states and districts, especially in math. It also highlights cases where declines were not statistically significant but still very likely to be meaningful, such as the drop in reading scores for the District of Columbia Public

School system. The approach provides more robust estimates and allows for a more flexible and policy-relevant way to interpret results, thereby improving decision-making.

Measurement Error Subarea Models: An Application of Farm Labor Parameters

Lu Chen, *National Agricultural Statistics Service*
Balgobin Nandram, *Worcester Polytechnic Institute & National Agriculture Statistics Service*
Linda J. Young, *National Agricultural Statistics Service*

Small area estimation methods typically combine survey direct estimates with available auxiliary information to provide indirect estimates from a model. If the covariate is measured with error, the potential pitfalls of ignoring this error in the modeling process include biased estimates and underestimated variability. Chen, et al. (2022) and Young and Chen (2022) described the hierarchical Bayesian subarea models to estimate the quantities from Farm Labor Survey conducted by USDA's National Agriculture Statistics Service (NASS). Among the covariates in the models, the previous year's official estimates, which are subject to variability, would presumably differ among areas. In this presentation, the measurement error is assessed and accounted for in the small area estimation models that have past official estimates as covariates. The original sub-area model and the measurement error model are examined and compared using Farm Labor wage rates survey data in the case study.

Hierarchical Bayes Small Area Estimation for County-level Health Prevalence to Having a Personal Doctor

Andreea Erciulescu, *Westat*
Jianzhu Li, *Financial Industry Regulatory Authority*
Tom Krenzke, *Westat*
Machell Town, *Centers for Disease Control and Prevention*

The complexity of survey data and the availability of data from auxiliary sources motivate researchers to explore estimation methods that extend beyond traditional survey-based estimation. The U.S. Centers for Disease Control and Prevention's Behavioral Risk Factor Surveillance System (BRFSS) collects a wide range of health information, including whether respondents have a personal doctor. While the BRFSS focuses on state-level estimation, there is demand for county-level estimation of health indicators using BRFSS data. A hierarchical Bayes small area estimation model is developed to combine county-level BRFSS survey data with county-level data from auxiliary sources, while accounting for various sources of error and nested geographical levels. To mitigate extreme proportions and unstable survey variances, a transformation is applied to the survey data. Model-based county-level predictions are constructed for prevalence of having a personal doctor for all the counties in the U.S., including those where BRFSS survey data were not available. An evaluation study using only the counties with large BRFSS sample sizes to fit the model versus using all the counties with BRFSS data to fit the model is also presented.

Return to Schedule Overview

Session G-6: Striking the Right Balance of Old and New: Combining Paper and Web Approaches to Optimize Data Collection

Organizer: Julie Weeks, *National Center for Health Statistics*
Chair: Julie Weeks*, National Center for Health Statistics*
Room 0105

Adding a Paper Questionnaire to a Teen Web Survey: Is it Worth the Effort? An NHIS-Teen Investigation

Benjamin Zablotsky, *National Center for Health Statistics*
Amanda Ng, *National Center for Health Statistics*
Lindsey Black, *National Center for Health Statistics*
Jonaki Bose, *National Center for Health Statistics*
Jessica Jones, *Maternal and Child Health Bureau*
Stephen Blumberg, *National Center for Health Statistics*

In 2021, the National Center for Health Statistics launched National Health Interview Survey – Teen (NHIS-Teen), a pilot follow-back survey designed to collect self-reported health information directly from teenagers aged 12-17. NHIS-Teen was 95 questions long, taking teenagers on average 15 minutes to complete.

Although NHIS-Teen was exclusively a web-based survey when it launched, a paper questionnaire was added in 2023 to determine if providing an alternative method for completing the survey could improve response rates, and, as a result, help reduce potential bias. The paper questionnaire was the first of two physically mailed reminders to nonrespondents to complete NHIS-Teen (eligible teens additionally received e-mail and/or text reminders). Among a subset of teenagers who completed NHIS-Teen in 2023 (n=754), we explored whether mode differences in sociodemographic characteristics, prevalence estimates, item missingness, or survey experiences existed.

In total, approximately 24% of teenagers who completed NHIS-Teen did so by paper. Among the sociodemographic characteristics examined, a higher percentage of boys who were mailed the paper questionnaire completed and returned the questionnaire compared to girls. The average item missingness was 0.9% for the web survey compared to 2.5% for the paper questionnaire. For the majority of health outcomes, no mode differences were found. When differences did exist, teenagers who completed the paper questionnaire were more likely to report a poorer health outcome (e.g. no regular bedtime). Teens who completed NHIS-Teen by paper also reported greater burden, but no differences were found in perceptions of difficulty, sensitivity, or length when compared to the web questionnaire.

Is it Helpful to Include QR Codes on Mail Contact Materials for Self-Administered Web Surveys?

Taylor Lewis, *RTI International*
Nicole Lee, *RTI International*
Dain Palmer, *RTI International*
Naomi Freedner, *RTI International*
Hannah Matzke, *Chicago Department of Public Health*
Nik Prachand, *Chicago Department of Public Health*

The Healthy Chicago Survey (HCS) is a self-administered, multimode survey conducted annually on a random sample of Chicagoans aged 18 years or older. The survey uses an address-based sampling frame and offers both web and paper modes, with a Choice+ data collection protocol offering a higher promised incentives for completing by web. Overall, about 90% of individuals take the HCS via web. Sampled households are sent a series of mailings inviting the adult who is next to celebrate a birthday to complete the survey. In the 2023 HCS administration, an experiment was fielded in which half the mailings contained a QR code enabling immediate survey launch for individuals with smartphones, while the other half of mailings did not include a QR code. This

paper summarizes differences observed across the two experimental groups with respect to response rates, breakoff rates, sociodemographic distributions, and key health indicator distributions.

Testing the Effectiveness of an Advance Data Worksheet and Initial Hard-copy Survey on Response Rates and Data Quality

Sean Goodison, *Bureau of Justice Statistics*

The Bureau of Justice Statistics (BJS), along with RTI International, invited 3,500 law enforcement agencies (LEAs) to participate in the 2023 Law Enforcement Management and Administrative Statistics Post-Academy Training and Officer Wellness (LEMAS PATOW) Survey. The LEMAS PATOW survey is part of BJS's Law Enforcement Core Statistics (LECS) program, which includes multiple establishment surveys focused on policies, procedures, and personnel characteristics of state, county, and local LEAs in the United States. Given concerns about both response rates and data quality across establishment surveys, the LEMAS PATOW was administered using an experimental design. All agencies were divided into one of three groups at the survey invitation stage to include a control group, another group receiving a data worksheet, and a final group receiving a hard-copy survey with the invitation. Previous research provides some evidence on the effectiveness of providing a data worksheet prior to survey administration, which would allow the respondent to consult with colleagues and collect administrative data in advance. Additionally, web-based establishment surveys may not have a hard-copy survey or only provide a paper copy later in the fielding. This presentation describes the research results from conducting an experiment to formally test the effectiveness of an advance data worksheet and initial hard-copy survey in the LEMAS PATOW, with the aim of improving response rates and better directing and preparing respondents to answer critical data points such as the fiscal year operating and training budgets, numeric questions on training hours, and questions on staffing with improved data quality.

Understanding Our Agency's Needs: Preparing to Integrate Web Surveys

Renee Stepler, *U.S. Census Bureau*
Jennifer Sinibaldi, *U.S. Census Bureau*
Jessica Holzberg, *U.S. Census Bureau*

Major Census Bureau surveys, such as the Current Population Survey, are preparing to integrate a web component into their current interviewer-administered data collection efforts. The Center for Behavioral Science Methods (CBSM) is assisting with these transitions and developing a research agenda to help confront challenges that are specific to Census Bureau surveys.

To understand our agency's needs and formulate a relevant research agenda, we conducted qualitative interviews with Survey Directors to explore topics including: the benefits, risks, and motivations for adding a web mode; expected changes to protocols and costs; and challenges for integrating a web mode. While much can be learned from existing literature, these interviews allowed us to uncover issues that may not be thoroughly addressed in the literature. We will present the findings from our interviews and our working research agenda to support Census Bureau surveys who are adopting web survey components. The methodology may be a helpful template for others in the federal statistical system who are also embarking on large survey improvements.

Return to Schedule Overview

Abstract Listings for Thursday, October 24

- **Concurrent Sessions H** **8:30 am – 10:00 am**
- **Concurrent Sessions I** **10:30 am – 12:00 pm**
- **Concurrent Sessions J** **1:45 pm – 3:15 pm**
- **Concurrent Sessions K** **3:30 pm – 5:00 pm**

Session H-1: Innovations in Sample Design: From Theory to Practice

Organizer: Stephanie Coffey, *U.S. Census Bureau*
Chair: Stephanie Coffey*, U.S. Census Bureau*
Chesapeake A

Applying Chi-square Tests to Examine Homogeneity of Proportions between Data Collected with Different Sampling Designs

Li-Yen Hu, *National Center for Health Statistics*
Yulei He, *National Center for Health Statistics*
Katherine E. Irimata, *National Center for Health Statistics*
Rong Wei, *National Center for Health Statistics*
Van L. Parsons, *National Center for Health Statistics*

Chi-square tests are often employed to examine the association of categorical variables, the homogeneity of proportions between two or more samples, and the goodness-of-fit for a specified distribution. In order to account for the complex sampling design (CSD) of survey data, variants of chi-square tests as well as software packages that implement these tests have been developed. While results from different statistical programs employing the same types of tests with similar adjustments accounting for CSD are generally comparable, different adjustments of chi-square tests to account for the CSD could yield variations in resultant p-values. In addition, when one is to examine homogeneity of proportions between different surveys, the procedures used to concatenate nonidentical sampling designs and to adjust for sampling weights between different surveys could somewhat influence statistical results. This presentation will address the performance of different types of chi-square tests that are commonly employed for survey data with CSDs, as well as explore the impact on statistical results due to different approaches of combining datasets with nonidentical sampling designs.

Balanced Sampling: Comparisons between INCA and Cube Method

Yang Cheng, *National Agricultural Statistics Service*
Lu Chen, *National Agricultural Statistics Service & National Institute of Statistical Sciences*
Luca Sartore, *National Agricultural Statistics Service & National Institute of Statistical Sciences*
Valbona Bejleri, *National Agricultural Statistics Service*

In balanced sampling, auxiliary information is used to improve the sample structure for many multipurpose surveys. The USDA's National Agricultural Statistics Service (NASS) has used Multivariate Probability Proportional to Size (MPPS) sampling design, an adaption of Brewer sampling, for multipurpose surveys. In joint work with the National Opinion Research Center (NORC) at the University of Chicago, we have explored balanced sampling with cube method to improve the MPPS design. In this presentation, integer-calibration (INCA) routines that perform discrete optimization over a constrained integer lattice are evaluated as an alternative to the cube method. The effectiveness and efficiency of INCA, which was originally designed for the US Census of Agriculture, and the cube method are compared. The relative errors of the two methods are evaluated in a case study based on data from the 2017 Census of Agriculture.

Extending Cochran's Sample Size Rule to Stratified Simple Random Sampling

Siyu Qing, *Ernst & Young LLP*
Richard Valliant, *Universities of Michigan & Maryland*

Audit samples are selected by businesses, institutions, government agencies, and other organizations to check the accuracy of financial reports and assess the quality of services provided among other reasons. A standard design used in auditing is stratified simple random sampling. Point estimates and confidence intervals for population values are usual products of an audit. Cochran's sample size rule and its extension by Sugden et al. (2000) provided minimum sample size formulas for a simple random sample to ensure effective normal approximation and adequate coverage for nominal 95% confidence intervals of a standardized or Studentized sample mean. The purpose of this paper is to extend Cochran's rule and establish a formula for the minimum

sample size for the normal approximation and the use of traditional one-sided or two-sided confidence intervals to be acceptable for mean estimation in stratified simple random samples. We concentrate on variables that are at least partially continuous. Simulations are used to examine the performance of our sample size formula with a variety of skewed populations based on ones encountered in auditing.

National Web Survey of Korean Americans using Respondent Driven Sampling

Kaidar Nurumov, *University of Michigan*
Associate Research Professor, Dr. Sunghee Lee,

Web-based respondent driven sampling (Web-RDS) is an extension of RDS, which exploits existing social networks for recruiting research participants, specifically, of rare, hidden and/or hard-to-reach groups. The strong social network combined with the high Web access rate among minorities makes Web-RDS an attractive platform for minority data collection. However, there is a notable void in the literature on design aspects of RDS.

This study attempts to fill this gap by providing practical design guidelines and tools for Web-RDS for a goal of improving data quality through successful implementations, where success is measured with sample composition and recruitment propensity. Specifically, we explore two specific design elements (seed selection and coupon design) through randomized experiments. For doing so, Web-RDS will be applied to a national survey of Korean Americans, a rare minority group that comprises less than 1% of the U.S. population. Seeds are selected from two sources: 1) a convenient sample and 2) randomly selected addresses associated with Korean surnames in the commercial data. For the coupon experiment, we use a nested design of three factors for coupons experimented in the main survey: 1) coupon delivery mode (preferred vs. email); 2) number of emails (all coupons in one email vs. one coupon per email) within the email mode; and 3) coupon format (verbal script vs. visual image) within the text mode. We will compare these seed and coupon conditions on the recruitment success and examine our Web-RDS data quality by comparing its estimates against Korean Americans ascertained from the American Community Survey.

Novel Sampling Methodology Combining Probability Sampling with Online Recruitment - A Case Study of an Internal Revenue Service (IRS) National Taxpayer Survey

Peter Enns, *Cornell University and Verasight*
Peter K. Enns, *Cornell University and Verasight*
Yvonne Nomizu, *Pacific Consulting Group*
Jake Rothschild, *Verasight*

This presentation will describe a novel sampling methodology combining probability sampling with dynamic technology-based recruiting, designed to achieve national population coverage and include often underrepresented segments. This approach was recently applied on an annual national taxpayer survey for the Internal Revenue Service who partnered with Pacific Consulting Group (PCG) and Verasight. The study leads from the IRS Statistics of Income and PCG, along with Verasight Co-Founder and Professor of Government at Cornell University Peter K. Enns, will share analysis related to (a) evaluation of the new methodology compared to existing approaches relative to sample demographic composition, population coverage, and non-response bias, (b) analysis of trend data through a methodology shift, and (c) statistical weighting strategies for a sample with multi-modal recruitment.

PSU Random Walk under Three-year Pooled Data

William Waldron, *National Center for Health Statistics*

Several prominent federal surveys have appealed to fixing Primary Sampling Units, typically county clusters, for ten-year periods to save recruitment costs. In addition to incurring long term sample bias, which may be tolerable, we demonstrate the diminished impact of the design on pooled three-year survey releases. Such pools are used for improving state-level estimates for outcomes that are relatively stable over time. We briefly depict how PSU Cycling can improve the estimation under three-year data pools along with other incremental design-based modifications that can improve state-level inference.

Return to Schedule Overview

Thursday, October 24, 2024 8:30 AM

Session H-2: Balancing Data Privacy and Utility

Organizer: Jennifer Rhorer, *National Agricultural Statistics Service*
Chair: Jennifer Rhorer, *National Agricultural Statistics Service*
Chesapeake B

Assessing Utility of Synthetic Data: Applications to the Survey of Doctoral Recipients and Census Transportation Planning Projects

Robyn Ferg, *Westat*
Tom Krenzke, *Westat*
Minsun Riddles, *Westat*

Synthetic data generation is gaining traction as a method for avoiding disclosure in publicly released data products (e.g., microdata, tables, and estimates). For the publicly released synthetic data products to be useful to data users, the conclusions reached using the synthetic data must be similar to those reached using the original data, both overall and for various subgroups. In deciding which synthesized data products to publicly release, a careful balance must be struck between utility and risk of disclosure. In this talk, we address the utility side of the risk-utility tradeoff, discussing several measures used to assess utility of synthetic data sets. These utility measures check that univariate distributions and multivariate relationships are maintained in the synthetic data, while also taking uncertainty into account. Moreover, these utility measures can be used to target specific demographic and geographical subgroups, ensuring equitable analysis results. We apply these utility measures to data generated for the Survey of Doctoral Recipients and Census Transportation Planning Products.

Empirically Measuring Privacy over the NIST CRC Deidentified Data Archive

Gary Howarth, *National Institute of Standards and Technology*
Christine Task, *Knexus Research Corporation*

Data de-identification to provide privacy protection is a rapidly developing research area, and there is broad interest in understanding how varying techniques impact the utility and privacy of data releases. The National Institute of Standards and Technology (NIST) is approaching these questions through a program called the Collaborative Research Cycle (CRC), an effort to benchmark the performance of deidentification approaches. CRC asks the public to deidentify diverse and interesting tabular benchmark data of real people drawn from the American Community Survey (ACS) data. The program has collected nearly 500 deidentified instances drawing from many different privacy approaches, including differentially private techniques, synthetic data, and traditional statistical disclosure control. NIST uses a custom software suite, SDNist, to provide extensive data fidelity and utility evaluation of the de-identified data to facilitate comparison of approaches.

This talk reviews the outcomes of a symposium hosted by NIST focusing on empirical privacy evaluation of the deidentified data collected by the CRC project. Specifically, we will cover a representative range of privacy frameworks and definitions, comparisons of privacy performance by deidentification approach, effects of privacy metric configuration (e.g., selection of quasi-identifiers, record similarity definition, attacker assumptions), and the effect of data distribution properties on privacy for diverse subgroups. This talk will provide traceable lessons on conducting privacy analysis for public data release programs.

Enabling Third-Party Audits of Algorithmic Systems with Privacy Enhancing Technologies

Michael Walton, *U.S. Census Bureau*
Tomo Lazovich, *U.S. Census Bureau*

As the push to regulate algorithmic decision systems grows, a persistent roadblock has emerged: model owners' inability to safely share model weights, outputs, and training datasets with outside auditors. There are often two reasons for this. First, owners may be legally restricted from sharing demographic information with others, severely hampering the types of bias audits that can be conducted. Second, owners may be reluctant to share full models because of the dangers of data reconstruction attacks, even in cases where they only share a black box API. In this short position paper, we propose privacy enhancing technologies (PETs) as a potential solution to enable privacy-preserving third-party auditing. We specifically examine the benefits and drawbacks of two approaches: differential privacy and secure multiparty computation. We then outline potential directions for future research to further elucidate the practicality of using these techniques as part of a third-party audit methodology. In the process, we hope to elucidate how agencies within the Federal Statistical System can make their datasets accessible via PETs to enable third-party model auditing in industry and across the federal government.

Feasibility Study on the Per-Record Differential Privacy for the Census of Agriculture

Michael Jacobsen, *National Agricultural Statistics Service*
Casey Meehan, *Tumult Labs*
Ashwin Machanavajjhala, *Tumult Labs*
Yang Cheng, *National Agricultural Statistics Service*

The USDA National Agricultural Statistics Service (NASS) is exploring a new methodology to modernize the statistical disclosure limitation techniques for the Census of Agriculture. Currently, NASS uses the cell suppression methodology, which was developed in the 1990s, to protect sensitive tabular data from disclosure to the public. For this feasibility study, NASS and Tumult Labs have proposed a relaxed version of differential privacy (DP) called per-record differential privacy (pr-DP) to offer DP-style guarantees in skewed-data settings. pr-DP sets a privacy threshold parameter, T_a, for each farm attribute a. Finally, we present a case study based on a set of synthetic microdata of Michigan farms from the 2017 Census of Agriculture.

Slowly Scaling Per-Record Differential Privacy

Brian Finley, *U.S. Census Bureau*
Anthony Caruso, *U.S. Census Bureau*
Justin Doty, *U.S. Census Bureau*
Ashwin Machanavajjhala, *Tumult Labs*
Mikaela Meyer, *MITRE*
David Pujol, *Tumult Labs*
William Sexton, *Tumult Labs*
Zachary Terner, *MITRE*

We develop formal privacy mechanisms for releasing statistics from data with many outlying values, such as income data. These mechanisms ensure that a per-record differential privacy (Seeman et al., 2023) guarantee degrades slowly in the protected records' influence on the statistics being released.

For context, formal privacy mechanisms generally add randomness to published statistics. If the statistics' distribution changes little with the addition, deletion, or alteration of a single record in the underlying dataset, an attacker looking at these statistics will find it plausible that any particular record was present or absent or took any particular value, preserving the records' privacy. More influential records - those whose absence, presence, or alteration would change the statistics' distribution more - typically suffer greater privacy loss. The per-record differential privacy framework quantifies these record-specific privacy guarantees, but existing mechanisms let these guarantees degrade rapidly (linearly or quadratically) with influence. While this may be acceptable in cases with some moderately influential records, it results in unacceptably high privacy losses when records' influence varies widely, as is common in economic data.

We develop mechanisms with privacy guarantees that instead degrade as slowly as logarithmically with influence. These allow for the accurate, unbiased release of statistics, while providing meaningful protection for highly influential records. As an example, we consider the private release of sums of unbounded establishment data such as payroll, where our mechanisms extend meaningful privacy protection even to very large establishments. We evaluate these mechanisms empirically and demonstrate their utility.

Return to Schedule Overview

Thursday, October 24, 2024 **8:30 AM**

Session H-3: Federal Advances in Sexual Orientation and Gender Identity Measurement

Organizer: Jennifer M. Ortman, *U.S. Census Bureau*
Chair: Jennifer M. Ortman, *U.S. Census Bureau*
Chesapeake C

What's Next for Measuring Sex and Gender on the National Health Interview Survey

Grace Medley, *National Center for Health Statistics*
Jim Dahlhamer, *National Center for Health Statistics*
Catherine Simile, *National Center for Health Statistics*

The National Health Interview Survey (NHIS) collects information from respondents on a variety of health measures. Measuring gender identity allows for the description of the health of the gender minority population and comparisons across gender identity groups. Since 2022, a measure for collecting gender identity has been on the NHIS, first as a series of tests, and then a final measure on the 2024 NHIS. The addition of a gender identity question, however, led to a re-examination of the NHIS roster measure that captures "sex".

For 2024, the NHIS measures has been revised to better capture the conceptual relationship between "sex" and "gender identity" while still being able to navigate through the survey instrument. The household roster sex question includes an introductory phrase to make it clear that the household roster question is asking about biological sex so respondents are asked the correct version of sex specific questions throughout the interview. The 2024 NHIS gender identity question is unchanged from 2023, while the sex assigned on the birth certificate and confirmation questions were dropped because it was clear these items were not helping respondents accurately report their gender identity.

Are You the Same People? Exploring Two-Step Gender Identity Question Options

Christina Dragon, *National Institutes of Health*
Rae Ellis, *U.S. Census Bureau*

Building on our work to assess Pride in Federal Service members willingness to disclose sexual orientation and gender identity demographic information on employment forms, this research explores how transgender people prefer to identify themselves. Our first study offered two different options of a two-step gender identity question. A long-standing and well tested two-step gender identity question included a current gender and sex assigned at birth question. Offering an alternative two-step question that replaced sex assigned at birth with a transgender experience question provided information about both comprehension and validation against the traditional two-step question. Analysis of the 472 validated responses showed the same people identifying as gender minorities from both question paths, however there were nuances in how people identified in each pathway. This presentation will explore response variation coupled with the qualitative follow-up responses providing additional context for answer choices and preferences. Some possible limitations might also include confusion about the way one of the questions was worded, but also undercounting transgender respondents who only identify with their current gender and not as "transgender" or "having transgender experience." In our 2024 follow-up we followed up with additional questions to better understand how the way the transgender identity question was phrased and if that impacted response decisions.

Cognitive Testing Results for Proxy SOGI items about Children and Adults in English and Spanish

Michelle McNamara, *National Center for Education Statistics*
Maria Payri, *American Institutes for Research*
Danielle Battle, *American Institutes for Research*

Though surveys have increasingly included questions asking about sexual orientation and gender identity (SOGI) resulting in guidelines (OMB, 2023), there is less guidance available on measuring SOGI in children, proxy measures (asking one person about another person's identity) and for asking questions in Spanish. We have been investigating how to incorporate SOGI questions into a household survey which requires adding a question about sexual orientation and gender identity for the adult respondent themself, as well as developing questions for the respondent to provide proxy reports about the sexual orientation and gender identity of the other parent (if present in the household) and the respondent's minor child. Finally, we are also testing gender neutral pronouns and family relationships and have developed Spanish translations for each SOGI question. Cognitive testing will be conducted in both English and Spanish to collect and analyze qualitative data from potential respondents to evaluate respondent understanding of revised and new SOGI questions, and willingness and ability to provide accurate response—particularly for proxy reporting of children and other parents. This presentation will discuss the results of this cognitive testing.

Self-Report and Proxy Measurement of Sexual Orientation and Gender Identity: An Examination of 2023 American Housing Survey (AHS) Data

George Carter, *Department of Housing and Urban Development*
Allyson Clark, *U.S. Census Bureau*

The American Housing Survey (AHS), the most comprehensive national housing survey in the United States, is sponsored by the Department of Housing and Urban Development (HUD) and collected by the U.S. Census Bureau. In 2023, sexual orientation and gender identity (SOGI) questions were added to the AHS in response to Executive Order 14075 "Advancing Equality for Lesbian, Gay, Bisexual, Transgender, Queer, and Intersex Individuals." Section 9 of the Executive Order includes a data directive to identify and address barriers to housing faced by LGBTQI+ individuals and families that place them at a high risk of housing instability and homelessness. The SOGI questions enable HUD to analyze these housing experiences, including experiences of HUD-assisted households. All adult respondents in the 2023 AHS were asked self-report questions on their sex recorded at birth, gender identity, and sexual orientation. As in most household surveys, the AHS collects household demographics from one respondent. To examine proxy collection of SOGI data, respondents in half of the AHS sample were asked experimental proxy SOGI questions for adult household members. Both self-report and proxy SOGI questions contained an option for respondents to indicate that they or others in their household use a different identity label than the options provided. Open-ended responses were manually coded into variables with detailed response codes, analyzed for differences between "self" and "proxy" responses, and compared to open-ended responses to SOGI questions in other surveys. Results of the open-ended analysis and ongoing research on the proxy questions will be discussed.

The Current State of Sexual Orientation and Gender Identity (SOGI) Question Research at NCHS: Where We've Been and Where We're Headed

Valerie Ryan, *National Center for Health Statistics*

At NCHS we conducted 350 cognitive interviews on gender identity questions and tested these questions on multiple rounds of our Research and Development Survey (RANDS). We found that the following question provides accurate estimates (reduces false positive and false negative error, compared to two-step gender questions) for both gender minority and non-minority respondents:
Are you (Mark all that apply):

- Female
- Male
- Transgender, non-binary, or another gender

Over a decade ago, our group developed the NHIS sexual identity question:

Which of the following best represents how you think of yourself?

- Lesbian or gay
- Straight, that is not lesbian or gay
- Bisexual
- Something else

This question was written with a close-ended follow-up probe for those who answer 'something else,' based on cognitive interview findings. The probe was tweaked to include the three most popular write-in responses (asexual, pansexual, and queer) and was added to RANDS rounds 6-9. The new probe is:

What do you mean by "something else"?

- I am not straight, but identify with another label such as queer, trisexual, omnisexual, polysexual, or pansexual
- I am asexual or on the asexual spectrum (including, but not limited to, demisexual and greysexual)
- I have not figured out or am in the process of figuring out my sexuality
- I do not use labels to identify myself
- Something else, please explain

Including these new options reduced the number write-ins, however, respondents with complex identities, such as 'demisexual panromantic,' still wrote those in.

Return to Schedule Overview

Thursday, October 24, 2024 8:30 AM

Session H-4: PANEL: Measuring the Health of the Federal Statistical System

Organizer: Steve Pierson, *American Statistical Association*
Chair: Katherine Smith Evans*, Kitty Evans Consulting*
Discussant: Steve Pierson*, American Statistical Association*
Vessey 1

There is a growing recognition that the United States must modernize how data are collected and disseminated. In the past seven years alone, new guidance or legislation has been introduced by the Commission on Evidence-Based Policy Making, the Committee on National Statistics, the White House, and Congress. But how will we know whether the nation is well served by these changes?

This session reviews the results of a year-long investigation into the health of the federal statistical agencies. The investigation—a joint American Statistical Association and George Mason University project with support from the Alfred P. Sloan Foundation—begins an effort to monitor America's statistical infrastructure, much like how civil engineers monitor the physical infrastructure of America's transportation network of roads, bridges, and airports—panelists will discuss how the investigation was conducted, review key findings and recommendations, and outline future work. In the proposed session, the panel would discuss the results of the investigation first described at a panel session at FCSM 2023. That panel was very well attended, and the audience provided invaluable feedback which shaped the course of the investigation. Whereas last year's panel discussed the general considerations for measuring the health of the federal statistical system, the proposed panel for this year's panel focus on the specifics: what was actually measured and future work that could improve the measurements.

We also note that the investigation produced recommendations that were endorsed by the board of the American Statistical Association at its most recent meeting. We have not discussed these recommendations at past FCSM meetings, and we believe they would be of interest to FCSM attendees.

Panelists

- Jonathan Auerbach, *George Mason University*
- Zach Seeskin, *NORC at the University of Chicago*

- May Aydin, *National Center for Science and Engineering Statistics*
- Claire McKay Bowen, *Urban Institute*
- Constance Citro, *Committee on National Statistics*
- Nancy Potok, *NAPx Consulting*

Thursday, October 24, 2024 8:30 AM

Session H-5: Bringing it All Together: Exploring Ways to Mesh Data from Multiple Sources

Organizer: Derrick D. Dennis, *Internal Revenue Service*
Chair: Derrick D. Dennis*, Internal Revenue Service*
Vessey 2

Working 9-5 and 5-9: Demographic Statistics on Entrepreneurs Who Work for a Wage

Nick LaBerge, *U.S. Census Bureau*
Nick LaBerge, *U.S. Census Bureau*
Adela Luque, *U.S. Census Bureau*
Emily Schondelmyer, *U.S. Census Bureau*

Persons who are self-employed and working for a wage are an under-researched group. Exploring the intersection between self-employment and wage work, and how that intersection varies across demographic groups and over time can offer valuable insights for policymakers, researchers, and businesses. As part of Census' Nonemployer Statistics by Demographics program (NES-D), we link confidential 2017-2019 NES-D microdata to W-2 data to estimate the share of nonemployer business owners who also work for a wage by the demographics of the business owners, at the national level and by industry. NES-D is a novel administrative-records-based annual statistical series that provides demographic and other characteristics of businesses with no employees in the U.S. by industry, geography and revenue size.

We will present initial estimates, and discuss observed differentials and patterns as well as employed data and methodology. These new estimates provide enhanced statistical coverage of the intersection between self-employment and wage work, and will become part of NES-D's experimental products series at the Census Bureau. Applications of this data can range from policy analysis and identifying disparities, to gig-economy research and economic forecasting.

A Structural Decomposition Analysis of U.S. Food Expenditures at the State Level

Eliana Zeballos, *Economic Research Service*
Wilson Sinclair, *Economic Research Service*

In 2022, the United States witnessed a total food spending surpassing $2.3 trillion, with inflation-adjusted per capita spending surging over 40% since 1997, as reported by the USDA's Economic Research Service (ERS) Food Expenditure Series (FES). However, changes in food spending patterns have not been consistent over time or geography. For instance, the Great Recession, COVID-19 Pandemic, and the Pandemic recovery periods each coincided with substantial changes in food spending at the national level, but the drivers of these changes have not yet been analyzed at the State level. This study investigates the drivers of changes in State-wide food spending as measured by four macroeconomic indicators—disposable personal income, propensity to spend versus save, propensity to spend on food versus non-food, and substitution between food-at-home (FAH) and away-from-home (FAFH). By employing structural decomposition analysis (SDA) at the State level from 1997 to 2022, the study assesses the drivers of food spending during recessionary and non-recessionary periods. This analysis uses data from various agencies of the Federal Statistical System of the United States, including the ERS State-level FES, which is underpinned by the Quarterly Census of Employment and Wages from the Bureau of Labor Statistics (BLS) and the Economic Census from the Census Bureau (CB). The FES provides estimates which control for inflation using the regional FAH and FAFH CPI from BLS and for population using data from the CB. Income and non-food spending data are obtained from the Bureau of Economic Analysis.

Policymakers and trade associations stand to benefit from insights into regional disparities, especially in recession recovery periods.

Continuous Count Study : Assessing Utility and Availability of Administrative Data

Thomas Mule, *U.S. Census Bureau*
J. David Brown, *U.S. Census Bureau*
Marta Murray-Close, *U.S. Census Bureau*
Jerry Maples, *U.S. Census Bureau*
Timothy Kennel, *U.S. Census Bureau*

The Census Bureau is committed to statistical innovation to meet the nation's twenty-first century challenges for the development and release of official statistics. The Continuous Count Study was one of the Census Bureau's Transformation and Modernization priorities identified to help achieve that goal. This study is researching how information and lower-level geographic and demographic characteristic estimates can be generated throughout the decade blending multiple sources of data. Our study has two areas of focus. First is an annual administrative record population estimates program building off the use of administrative records in the 2020 Census and the Real Time Administrative Record Census evaluation. Second is research on population estimation and small area estimation techniques to generate alternative population estimates that can be used to assess the quality of the administrative record data results. Latest results and research directions for the future will be presented.

Temporary Protected Status (TPS) Recipients in the United States: Lessons Learned from Development of the Office of Homeland Security Statistics' Immigrant Benefits Lifecycle Database

Jason Schachter, *Office of Homeland Security Statistics*
Sarah Miller, *OHSS*
Noah Schofield, *OHSS*
Gregory Su, *OHSS*
Michael Hysong, *OHSS*

The Office of Homeland Security Statistics (OHSS) has expanded its ability to conduct person-centric analysis of migrants as they move through the U.S. immigration system. This included development of the Enforcement Lifecyle database, which enables reporting of individual outcomes for those encountered at the border. Also of interest is development of the Immigrant Benefits Lifecycle database, which would follow individual outcomes for migrants to the United States, as they move through the benefits processes.

One example of an immigration benefit is granting of Temporary Protected Status (TPS), which is a humanitarian program that allows certain nationals to temporarily stay and work in the United States due to unsafe conditions in their home countries. Little is known about how TPS beneficiaries move through the U.S. immigration system, such as length of time from initial arrival to the United States before applying and being granted TPS status, as well as their current immigration status.

Development of the Immigrant Benefits Lifecycle has been hindered by lack of unique identifiers. However, U.S. Customs and Border Protection (CBP)'s Arrival and Departure Information System (ADIS) is a person-centric data system that integrates data across DHS. This paper discusses development of OHSS's Immigrant Benefits Lifecycle database, utilizing ADIS as the "spine" for data linkages from USCIS data on immigrant benefits. Evaluation of the quality of data linkages will be conducted. Linkages and longitudinal analysis will be limited to the TPS population and serve as a proof of concept for future benefit lifecycle database development and analysis.

Some Recent Advances and Open Problems in Post-Linkage Data Analysis

Martin Slawski, *George Mason University*
Brady T. West, *University of Michigan*
Priyanjali Bukke, *George Mason University*
Emanuel Ben-David, *U.S. Census Bureau*

Various contemporary data products such as the Census-Enhanced Health and Retirement Study (CenHRS) involve record linkage of a primary data source like a survey with federal administrative data (e.g., the Census Business Register). Post-linkage analysis performed on the resulting linked files in the face of uncertainty about the correctness of matched records can be a daunting task, specifically in the secondary analysis setting in which only the linked file is available to the data analyst. In this presentation, we will provide an overview of recent advances in this area and outline selected open problems that are expected to play an important role in future developments. We will also discuss "Unlinked Data Analysis" as an alternative paradigm that sidesteps the explicit creation of linked files. Several case studies based on synthetic and real data will be used to illustrate major aspects of the recent developments in this area, including a new contributed R package.

Return to Schedule Overview

Thursday, October 24, 2024 **8:30 AM**

Session H-6: The National Household Travel Survey: Applications and Opportunities

Organizer: Stacey Bricka, *MacroSys*
Chair: Daniel Jenkins*, Federal Highway Administration*
Discussant: Clara Reschovsky*, Bureau of Transportation Statistics*
Room 0105

NHTS Overview and Policy Applications

Daniel Jenkins, *Federal Highway Administration*

The National Household Travel Survey (NHTS) program within FHWA has been collecting travel behavior data to support planning and policy needs for over 50 years. The survey documents household and person level demographics, household vehicle fleet composition, attitudes regarding various aspects of transportation, and all trips made (all modes, all purposes) for a 24-hour period by household members ages 5 and older. For the 2022 iteration of the NHTS a new approach was adopted to provide more timely and relevant transportation statistics to the American public. The impetus for this redesign, referred to as NextGen NHTS, included a host of innovations in transportation which were rapidly emerging but were not being measured during the adoptive phase. To ensure that these innovations are captured more readily, increasing the frequency of data collection was necessary. Now, instead of collecting data every 7-10 years, FHWA is collecting transportation behavior data biannually from US households. While the result is a small sample size with each iteration (n=7,500) over a period of 3 iterations, or 6 years, the combined sample size equals the previous sample size allowing for more robust analyses. This presentation will introduce the key components of the NHTS, discuss how FHWA has responded to changes in survey methods over time, highlight key data metrics, and illustrate how the data is used to address important policy questions, including the Conditions and Performance report to Congress.

Exploring Shifts in Mobility in the U.S.: An Analysis of Travel Trends Using the NHTS

Stacey Bricka, *MacroSys*
Mitchell Fisher, *MacroSys*
Paul Schroeder, *Consultant*

FHWA's NHTS provides insights into travel behavior patterns enhanced by traveler demographic and attitudinal attributes. Key policy topics informed by the 2022 NHTS include use of household vehicles for commercial travel (i.e., rideshare, food and grocery delivery, etc.); use of emerging travel modes such as e-scooters; impact of telecommunications on travel; and whether travel changes related to the COVID-19 pandemic are anticipated to be temporary or permanent. Examples of a few key findings that will be covered in this presentation include:

1. Americans traveled less in 2022 than 2017. The reduced levels of travel reflect changes in daily life.

For example, online errands such as shopping, banking, and telemedicine have reduced the need for short stops to and from work.

2. In 2022, 38% of workers reported a change in work travel, mainly traveling to work less. Of those who reported a change, 70% indicated the change was permanent. Note that the survey was conducted prior to most return-to-office orders.
3. 36% of respondents reported an increase in online shopping frequency, with an overwhelming 83% indicating this was a permanent change. The average number of reported online purchases in the past 30 days doubled across almost all households. Not only does online shopping influence travel patterns, and travel demand models, but the shift from household members making a shopping trip to a commercial vehicle making a home delivery of the purchased goods has significant implications for transportation policy and planning.

An Evaluation of the 2022 National Household Travel Surveys: A Total Survey Error Comparison of the ABS vs. the Probability-Based Panel NextGen NHTS Studies

Paul Lavrakas, *Consultant*
Trent Buskirk, *Old Dominion University*
Alan Pate, *Battelle*
Elizabeth Sloan, *Battelle*
Ta Lui, *Battelle*
Filmon Habtemichael, *Battelle*

This presentation reports findings that address the critical issue of whether probability-panel surveys can provide at least similar findings compared to otherwise equivalent ABS surveys from the perspective of total survey error. The Federal Highway Administration periodically conducts the NHTS to generate official statistics about travel among U.S. residents. In 2022, at the request of OMB, two versions of the NHTS were conducted simultaneously, using the same questionnaire. The version generating the 2022 NHTS official travel statistics used an ABS probability sample, mail recruitment, and online (or mail) data collection. The "experimental" version used an existing probability-panel sample, email recruitment, and online data collection. Both surveys sought to achieve a final sample size of 7,500 households. Data collection for each survey occurred from 1/2022 through 1/2023. The evaluation of the two surveys incorporated sources of sampling and non-sampling errors and was carried out by Battelle. The evaluation revealed very few core attributes that were meaningfully different between the two 2022 NHTS versions among the hundreds of statistical comparisons made. Where differences were noted, they were primarily related to demographic attributes rather than travel-related attributes. Thus, the two surveys, as designed and implemented, basically led to the same conclusions regarding the core attributes in the 2022 NHTS questionnaire. Furthermore, both surveys were judged to be "Fit for Purpose" (cf. Santos, 2014). These are very important findings given the considerable cost differences between these types of surveys as well as current trends in survey research towards considering panel-based solutions.

Unveiling the Roadmap: The Utility and Functionality of NHTS Data Dissemination Tools

C. Ross Wang, *Oak Ridge National Laboratory*
Layla Sun, *MacroSys*

The data that comprise the NHTS are quite complex and contain a lot of detailed information, including each location visited during the travel day. As such the data are stored in four files within in a hierarchical database. The data are made publicly available through a dedicated website hosted and maintained by Oak Ridge National Laboratory (https://nhts.ornl.gov/). The site provides documentation for all 9 data collection cycles conducted since 1969. FHWA has invested in the development of online web-based tools to summarize the data and compiles an annual compendium of uses of the NHTS data. A download section connects users with the desired dataset and a library of publications helps to connect users with reports on various applications of the data. The data dissemination plan is very robust and comprehensive. In addition to the online tools, the primary dissemination approach is through in-person and online workshops, conference sessions, and user group meetings. In addition, analytic reports about the data are issued periodically along with newsletters, briefs, and an annual compendium of uses.

Return to Schedule Overview

Thursday, October 24, 2024 10:30 AM

Session I-1: LLMs Unleashed: Revolutionizing Surveys with AI

Organizer: Erin Boon, *Bureau of Labor Statistics*
Chair: Erin Boon, *Bureau of Labor Statistics*
Chesapeake A

Automated Abstract Tagging: Enhancing Peer-Reviewed Abstract Categorization with MeSH Ontology and Large Language Models

Michael Long, *RTI International*
Kate Burdekin, *RTI International*
Chris Coxen, *RTI International*
Demian Pasquarelli, *RTI International*

Rigorous peer-reviewed review of applications funded by the government requires a large investment from subject matter experts (SMEs), administrative tasks and NIH leadership to equitably fund efforts that push forward science in strategic areas. Using the MeSH (Medical Subject Headings) controlled and hierarchically-organized vocabulary produced by the NLM as a foundation combined with a locally installed, secure, free, open-source Large Language Models (LLMs), RTI was able to automate a process that historically required significant staff time to manually read and categorize the focus of these applications. The novel LLM runs on RTI secure servers and can read the abstract text and identify NIH MeSH ontology terms automatically and without any risk of transmission to the cloud. The identified MeSH terms can then be stored in a database, where SMEs can review, edit, and approve the terms. When comparing results between the LLM, a human tagger, and the NIH's online tool, the LLM had higher similarity with the NIH tool than the human did. By continuing this work, we can improve the ontology label accuracy and general performance of the LLM solution. This approach can extend beyond the abstract tagging if reviewer biographies were available to better align the applications with appropriate reviewers based on their scientific focus reflected in their bibliography. This would allow for a highly efficient solution for matching proposals to proposal reviewers with the goal of focusing the time and effort on reviewing the merit of the applications and removing some of the labor-intensive administrative process.

Detecting LLM-generated Survey Responses

Joshua Lerner, *NORC at the University of Chicago*
Brandon Sepulvado, *NORC at the University of Chicago*
Lilian Huang, *NORC at the University of Chicago*
Ipek Bilgen, *NORC at the University of Chicago*
Leah Christian, *NORC at the University of Chicago*

Detecting potentially falsified responses is vital to ensuring high quality survey data. Although tools to detect some types of falsification exist, the popularity of large language models (LLMs) like ChatGPT and the seeming plausibility of their responses pose a challenge to ensuring that open-ended responses come from verified respondents rather than LLMs. This presentation presents the results of an investigation into whether it is possible to identify when open-ended responses come from LLMs. We generated responses from LLMs to compare to responses from survey respondents to two open-ended questions and generated a machine-learning based classifier to predict whether responses were from the LLM or respondents. We will share results from this analysis, limitations, and potential directions for future research.

Large Language Models do not Respond like Survey Respondents

Brandon Sepulvado, *NORC at the University of Chicago*
Joshua Y. Lerner, *NORC at the University of Chicago*
Lilian Huang, *NORC at the University of Chicago*
Leah Christian, *NORC at the University of Chicago*
Ipek Bilgen, *NORC at the University of Chicago*

Much recent research on large language models (LLMs) has explored their utility for survey research, including having LLM serve as synthetic respondents and for helping refine questionnaire design. However, many of these potential use cases assume that LLMs respond to survey questions in a manner similar to humans. This presentation showcases recent efforts to ascertain the veracity of this assumption.

To understand whether LLM responses to open-ended survey questions resemble those of actual respondents, we fielded multiple experiments using NORC's AmeriSpeak panel. We varied question type, with one question about major issues facing the U.S. designed to elicit a list-type response and another about how respondents understand AI designed to elicit a longer-form and less structured reply, and then used multiple well-known LLMs to respond to these same questions. We compare human and LLM responses in terms of quantitative metrics (i.e., response length, readability, and linguistic diversity) and in terms of content.

Findings indicate that—although human and LLM responses are similar in readability—there are significant differences in response length, linguistic diversity, and content. While humans tend to satisfice with their responses for a given question, LLMs seem to optimize their response to the same prompt. We conclude by discussing the implications for the use of LLMs in survey research and how fine tuning LLMs changes results.

UX Design Principles for Integrating Textbots into Surveys

Elizabeth Dean, *NORC at the University of Chicago*
Soubhik Barari, *NORC at the University of Chicago*
Brandon Sepulvado, *NORC at the University of Chicago*
Leah Christian, *NORC at the University of Chicago*

Federal agencies and research providers are evaluating the potential and challenges of generative AI and Large Language Models (LLMs) to support the collection, reporting, and usage of federal statistics. One compelling and challenging use case is the integration of chat-based AI assistants to self-administered web surveys. We refer to these chat-based assistants as textbots, to distinguish them from customer-service focused chatbot tools. Textbots have the potential to improve survey data quality. Some opportunities include: increasing engagement; probing insufficient, inconsistent, or missing answers; collecting more detail on open-end responses; and restructuring the way we collect complicated data like employment status, industry and occupation, and medication history to be more centered on the respondent's experience. There are many challenges to integrating textbots into surveys, including concerns about bias, human subjects risks, and data security.

This presentation will focus on the user experience (UX) of textbots in surveys. First, we will describe a case study of a textbot integrated into a basic survey. Second, we will summarize fundamental principles of user-centered design (such as context, usability, clarity, and feedback). Finally, we will cite our experience with the textbot case study, existing research on AI chatbots, and user-centered design best practices to develop a set of UX design principles for survey textbots to guide future applications of this technology.

Return to Schedule Overview

Session I-2: New Explorations of Disclosure Avoidance Techniques and Applications

Organizer: Jingchen (Monika) Hu, *Vassar College*
Chair: Saki Kinney, RTI International
Chesapeake B

Exploring the Application of Differential Privacy to a Subset of the Cells of a Table

Habtamu Benecha, *National Agricultural Statistics Service*
Yang Cheng, *National Agricultural Statistics Service*
Michael Jacobsen, *National Agricultural Statistics Service*
John Grant, *National Agricultural Statistics Service*
Lu Chen, *National Agricultural Statistics Service*
Luca Sartore, *National Agricultural Statistics Service*
Valbona Bejleri, *National Agricultural Statistics Service*

USDA's National Agricultural Statistics Service (NASS) conducts the Census of Agriculture (CoA) every five years. The Census is the leading source of statistics on U.S. agriculture, farms and the people who operate them. After census data are collected and summarized, NASS publishes the estimates at national, state, and county levels. In doing so, the agency makes every effort to protect the privacy of individual census respondents. Currently, NASS applies a network flow-based cell suppression system to protect published tables. To improve the protection of respondents, increase utility of published data, and modernize its statistical disclosure limitation techniques, NASS is exploring alternative methods such as Differential Privacy (DP), Random Tabular Adjustment (RTA), and an improved version of the cell suppression approach based on Linear Programming. In recent years, survey researchers are increasingly exploring differential privacy for privacy protection to survey data. In the context of the CoA, DP guarantees that the inclusion or exclusion of a farm operation does not significantly affect census estimates, depending on the specified tuning parameters. While differential privacy applies noise to all cells of a table, not all cells necessarily need protection depending on an agency's policy or some other analytical criteria. In this presentation, the feasibility of combining ideas from cell suppression, RTA, and DP is explored to improve utility of published tables while protecting privacy. Results from real and simulated datasets will be discussed.

The Inaugural Privacy and Public Policy Conference: Advancing Data Privacy and Public Policy Through Collaborative Engagement

Claire McKay Bowen, *Urban Institute*
Jingchen (Monika) Hu, *Vassar College*

In 2023, a collaboration between Vassar College, Penn State University, the Urban Institute, and the National Institute of Statistical Sciences culminated in a two-day workshop focused on the intersection of data privacy, equity, and public policy. The event catalyzed the inception of the inaugural Privacy and Public Policy Conference, held in September 2024. The goal of this conference is to foster and enhance collaboration among privacy experts, researchers, data stewards, data practitioners, and public policymakers. This event serves as a crucial platform to bridge communication gaps across these diverse groups and collectively explore technical and policy solutions for addressing challenges related to data privacy and public policy. This presentation highlights the key discussions and outcomes from the conference, offering valuable insights and lessons learned that can inform the initiation of similar endeavors aiming to engage a diverse spectrum of stakeholders.

The Unintended Effects of Privacy in Decision Tasks

Ferdinando Fioretto, *University of Virginia*

Differential Privacy has become the go-to approach for protecting sensitive information in data releases and learning tasks that are used for critical decision processes. For example, census data is used to allocate funds

and distribute benefits, while several corporations use machine learning systems for financial predictions, hiring decisions, and more. While differential privacy provides strong guarantees, we will show that it may also induce biases and fairness issues in downstream decision processes. In this talk, we delve into the intersection of privacy, fairness, and decision processes, with a focus on understanding and addressing these fairness issues. We first provide an overview of Differential Privacy and its applications in data release and learning tasks. Next, we examine the societal impacts of privacy through a fairness lens and present a framework to illustrate what aspects of the private algorithms and/or data may be responsible for exacerbating unfairness. Finally, we propose a path to partially mitigate the observed fairness issues and discuss challenges that require further exploration.

Generating Synthetic Data for the National Household Food Acquisition and Purchase Survey: A Case Study

Joseph Rodhouse, *Economic Research Service*
Jingchen (Monika) Hu, *Vassar College*

The USDA's National Household Food Acquisition and Purchase Survey (FoodAPS) fills a critical data gap and supports research that informs policymaking on key national priorities, including health and obesity, hunger, and nutrition assistance policy. FoodAPS collects confidential information and accessing and utilizing FoodAPS data for research purposes may be restricted due to privacy regulations and data access protocols. Research outputs using restricted FoodAPS data are heavily scrutinized for disclosure before public release. For researchers and policymakers hoping to conduct studies or make evidence-based policy decisions at levels that require restricted FoodAPS data, this can pose a barrier since traditional statistical disclosure controls can drastically limit data access and reduce analytical validity. However, recent advances in statistics and computing have made synthetic datasets viable alternatives that can increase data access and ameliorate analytical validity issues. As a result, the goal of this project is to explore different approaches to generating synthetic FoodAPS datasets that could increase data access and analytical validity. This paper will evaluate possible modeling strategies and analytical validity metrics and assess potential approaches to quantifying the reduction of risk of disclosure with the generated synthetic data. Finally, a discussion of the results and their implications for making synthetic FoodAPS datasets available for the food research and food policy-making communities will be explored.

Deploying PETs within Integrated Data Systems: Learnings and Challenges

Stephanie Straus, *Georgetown University*
Amy O'Hara, *Georgetown University*

This session explores the policy, technical, and institutional aspects of integrating Privacy Enhancing Technologies (PETs) into government data systems. Building on her landscape analysis of PET use in education data systems, Straus is currently supporting two state agencies in testing PETs: 1. synthetic data in Nebraska's Statewide Workforce & Educational Reporting System and 2. privacy preserving record linkage in Arkansas Office of the CDO. With these demonstrations, she will share the considerations and challenges for government agencies across the lifecycle of a PET implementation, from the legal approval, to institutional sign offs, to safelisting the necessary software and code, as well as preliminary results from her two state partners. Straus will outline how her experiences with state-level jurisdictions can serve as a model for state-federal and federal-federal data sharing efforts, as well as how these techniques and methods can improve upon current data access and linkage practices.

Return to Schedule Overview

Session I-3: Vital Signs and Data: A Healthy Dose of Analytics

Organizer: Jordan Riddle, *Bureau of Transportation Statistics*
Chair: Jordan Riddle*, Bureau of Transportation Statistics*
Chesapeake C

A Machine Learning Approach to Index Creation for Health-related Outcomes

Chandler McClellan, *Agency for Healthcare Research and Quality*
Patricia Keenan, *Agency for Healthcare Research and Quality*
Tom Selden, *Agency for Healthcare Research and Quality*

This work explores the application of machine learning techniques in the development of comprehensive area level indexes for social determinants of health (SDOH). Social determinants play a pivotal role in shaping health outcomes, yet quantifying their impact is a challenging task due to the complex interplay of various factors. Traditional approaches often rely on selection and weighting of variables which may overlook subtle yet influential factors. Additionally, these methods do not explicitly guarantee the associated measure will be related to the underlying factors associated with health outcomes.

In this study, we propose a novel methodology using machine learning to develop indexes for SDOH variables. By leveraging neural network architecture similar to that of autoencoders, neural networks designed for dimensionality reduction, and optimizing for factors associated with health, our methods generate summary indexes of SDOH variables that are predictive of a broad range of health-related outcomes. These machine-learning derived indexes are shown to perform better than current widely used indexes.

The development of machine learning-derived indexes for social determinants of health presents a promising avenue for enhancing our understanding of the complex relationships between socio-economic factors and health outcomes. By providing more accurate and comprehensive measures, these indexes hold significant potential for informing evidence-based policymaking, resource allocation, and targeted interventions aimed at addressing health disparities and promoting health equity within communities

Developing a Cost-effective, Real-Time Surveillance Tool to Monitor Workers' PPE Concerns During Respiratory Pandemics

Nora Y. Payne, *National Institute for Occupational Safety and Health*
Emily J. Haas, *National Institute for Occupational Safety and Health*

Personal protective equipment (PPE) plays a critical role in protecting workers from infectious airborne viruses during a respiratory pandemic. However, workers across many industries may not be protected during a pandemic when PPE is unavailable due to shortages, when PPE is not worn or worn incorrectly by workers using PPE that is new to them, or when organizational practices to support proper PPE use are not followed. Monitoring these PPE challenges can provide critical information for federal agencies, industry groups, and employers, particularly about worker populations whose PPE challenges are less studied. We present a proof-of-concept for a cost-effective, rapidly-deployable surveillance framework that uses machine learning techniques to automatically detect PPE concerns present in complaints submitted to the Occupational Safety and Health Administration (OSHA). We detail our methodology and challenges as well as findings from a retrospective application of our machine learning framework that demonstrate the potential utility of our approach.

Enhancing Data Services: Insights from NCHS' Health Data User Survey

Ryne Paulose, *National Center for Health Statistics*
Catherine Lesesne, *Deloitte Consulting*
Dagny Olivares, *National Center for Health Statistics*
Amy Branum, *National Center for Health Statistics*
Kiana Morris, *National Center for Health Statistics*

In today's data-driven world, statistical agencies within the Federal Government, like the National Center for Health Statistics (NCHS), play a critical role in providing timely, accurate, objective, and relevant statistical data essential for evidence-based decision-making across various domains. In 2023, NCHS conducted the Health Data User Survey to gain insights into the preferences and needs of health data users, especially of NCHS products and services.

A questionnaire was distributed for 4-weeks to federal and non-federal individuals via national and professional organizations and ListServs, resulting in 2,617 respondents. Questions covered health data use, health topics of interest, data access, and preferences.

The survey findings revealed a significant reliance on NCHS data and publications, with 64.2% of respondents using these resources across various sectors including state/local government, federal government, and academic/research institutions. Users preferred accessing data files for statistical analyses, while non-users favored in-depth reports and simple graphics. Additionally, 52% of users of NCHS resources conducted their own statistical analyses, indicating a high level of engagement with the data. These findings underscore the importance of tailoring products and services to meet diverse user needs.

Routinely engaging stakeholders and conducting surveys like the Health Data User Survey are important for statistical agencies like NCHS to ensure relevance of products and services. By understanding and meeting user needs effectively, agencies can enhance their capacity to facilitate evidence-based decision-making.

Identifying Stimulant and Opioid abuse in Clinical Notes

Rihem Badwe, *National Center for Health Statistics*
Nikki Adams, *National Center for Health Statistics*

Substance abuse remains a significant national public health concern focused on the increase in stimulant use and co-use of stimulants and opioids. The National Center for Health Statistics' National Hospital Care Survey (NHCS) provides national estimates on patient care related to drug use in the United States. To better identify patients with stimulant use in the NHCS collected electronic health records (EHR), an enhanced algorithm was developed utilizing natural language processing (NLP) methods. To verify the NLP methods, an annotated gold standard dataset with high quality data was developed.

Two clinicians annotated 1,000 records in the 2020 NHCS dataset related to stimulant or opioid use, and/or key words indicating stimulant/opioid use in the clinical notes. The sample was divided into 3 groups to target for sampling encounter notes that contain (1) stimulant terms only; (2) stimulant and opioid terms; and (3) neither stimulant nor opioid terms.

It is challenging to identify stimulant use in hospital data because diagnoses codes don't always capture specific types of substance use and there are no standardized methods to capture the information in the clinical notes. In the annotation process, output variables were developed to ensure the annotation questionnaire would capture relevant information needed for the NLP algorithm. To measure the algorithm performance, "ground truth" annotated data for each variable was compared to the algorithm output for that variable. This presentation will provide an overview of the development annotation questionnaire, data collection, and process of developing output variables for the final enhanced stimulant algorithm.

Return to Schedule Overview

Thursday, October 24, 2024 **10:30 AM**

Session I-4: Leading the Federal Statistical System: Today's Lessons for Tomorrow's Success

Organizer: Alyssa Holdren, *Bureau of Economic Analysis*
Chair: Vipin Arora*, Bureau of Economic Analysis*
Vessey 1

Census Bureau Leadership Perspective
Ron Jarmin, *U.S. Census Bureau*

Provides a glimpse into the day-to-day leadership activities and longer-looking strategies, plans, and priorities at the Census Bureau

National Center for Health Statistics Leadership Perspective
Brian Moyer, *National Center for Health Statistics*

Provides a glimpse into the day-to-day leadership activities and longer-looking strategies, plans, and priorities at the National Center for Health Statistics

National Center for Science and Engineering Statistics Leadership Perspective
Emilda Rivers, *National Center for Science and Engineering Statistics*

Provides a glimpse into the day-to-day leadership activities and longer-looking strategies, plans, and priorities at the National Center for Science and Engineering Statistics

Bureau of Labor Statistics Leadership Perspective
Bill Wiatrowski, *Bureau of Labor Statistics*

Provides a glimpse into the day-to-day leadership activities and longer-looking strategies, plans, and priorities at the Bureau of Labor Statistics

Return to Schedule Overview

Thursday, October 24, 2024 **10:30 AM**

Session I-5: Data Linkages Support Expanded Analyses on Social and Market Factors in the U.S. Health System

Organizer: Alice Zawacki, *U.S. Census Bureau*
Chair: Alice Zawacki*, U.S. Census Bureau*
Vessey 2

Linkage of the National Ambulatory Medical Care Survey Health Center Component to U.S. Housing and Urban Development Administrative Data
Jessie L. Parker, *National Center for Health Statistics*
Cordell Golden, *National Center for Health Statistics*
Jill J. Ashman, *National Center for Health Statistics*
Sonja N. Williams, *National Center for Health Statistics*
Brian W. Ward, *National Center for Health Statistics*

In 2021, the National Ambulatory Medical Care Survey (NAMCS) Health Center Component modernized its data collection and began collecting clinical visit data from health centers using electronic health record (EHR) submission. With this shift from manual, in-person, data abstraction to data collection via EHR submission, for the first time in its survey history the NAMCS Health Center Component is now collecting personally identifiable information (PII). The collection of these PII data has allowed for the National Center for Health Statistics to link NAMCS Health Center Component data with external data sources at the patient level. The resulting linked data will support a wide array of patient outcomes studies, including the opportunity to study complex relationships between housing and health. This presentation describes the process used to link these data to the U.S. Housing and Urban Development (HUD) administrative data. Processes used to perform an assessment of the NAMCS PII data are presented, along with results of this quality assessment. In addition, the methodology behind the matching techniques used to link the NAMCS Health Center Component and HUD administrative data is also presented. Finally, plans for linkage of NAMCS to additional external data sources, and how these data are made available for performing patient-centered outcomes research are detailed.

AHRQ SDOH Database: Opportunities for Geographic Data Linkages

Patricia S. Keenan, *Agency for Healthcare Research and Quality*
James Kirby, *Agency for Healthcare Research and Quality*

Social Determinants of Health (SDOH) are estimated to account for over 50% of county-level variations in health outcomes and are significant drivers of health disparities. As such, small-area SDOH data – data at the community or sub-county level – are key for understanding and addressing community-level variations and improving health equity. This presentation will provide an overview of contents, applications, and upcoming enhancements to the Agency for Healthcare Research and Quality (AHRQ) Social Determinants of Health Database, a consolidated set of national standardized datasets on social determinants of health factors. The Database enhancements are funded by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) project, "Small-Area Community SDOH Data: Enhancements and Linkages to Inform Action." The SDOH Database synthesizes variables from multiple publicly available data sources across five SDOH domains and is linkable at multiple geographic levels. The presentation will include examples of SDOH Database linkages in research.

The Long-Term Decline in Small Firms Offering Health Insurance: Drivers, Dynamics, and Mitigating Policies

Mark K. Meiselbach, *Johns Hopkins Bloomberg School of Public Health*
Jean Abraham, *University of Minnesota School of Public Health*

There has been a substantial decline in the percentage of small firms that offer health insurance over the last two decades. The drivers and consequences of this trend are not well-understood. We use 2002-2020 Medical Expenditure Panel Survey – Insurance/Employer Component data to measure small firm (<50 employees) health insurance offer rates at the state-year level and link to enrollment from the Current Population Survey (CPS) and state expenditures from the National Association of State Budget Officers (NASBO). We estimate difference-in-differences models of within-state changes in small firm health insurance offer rates and state policies. We separately regress CPS enrollment among individuals at small firms and NASBO state expenditures on state-year level offer rates. All analyses use include state and year fixed effects with robust standard errors. 30 states experienced a decrease of over 10 percentage points (p.p.) from 2002 to 2020 in offer rates. State minimum wage increases were consistently associated with a decline in the percentage of small firm offer rate. Each 10 p.p. decrease in state-year level small firm offer rate was associated with a 1.1 p.p. decline in employer-sponsored enrollment, a 0.8 p.p. increase in Medicaid enrollment ($50 increase in state Medicaid expenditures per capita), a 0.7 increase in individual market enrollment, but no significant changes to uninsurance, except in non-Medicaid expansion state. Our findings suggest that state policies contribute to the decline of small firm health insurance offers and that the decline may translate to greater Medicaid expenditures. However, Medicaid expansion likely mitigates resulting coverage losses.

A Comparison of Individual and Small Area Level Social Measures and Heterogeneity in Correlation with Health Outcomes and Mortality

Victoria Udalova, *U.S. Census Bureau*
Nicole Gladish, *Stanford University*
Aubrey Limburg, *U.S. Census Bureau*
Bob Phillips, *American Board of Family Medicine*
David Rehkopf, *Stanford University*

Small-area level social deprivation measures are being implemented in the U.S. to adjust provider payments to address social risks faced by patients above and beyond their underlying medical conditions. Currently, there is no empirical analysis of how strongly any of these measures are related to health outcomes. It is also still unknown which small-area level measures best capture social risk related to morbidity and mortality, and whether they capture this equally well for a variety of individuals. The purpose of this project is to 1) examine how closely individual measures of deprivation are associated with area-level deprivation measures, 2) compare the relative strength with which area-based social measures and individual social measures are correlated with health outcomes and mortality, and 3) the extent to which these associations vary by age, gender, race, ethnicity, and rurality. This research is a collaboration between the American Board of Family Medicine (ABFM), Stanford University, and the Enhancing Health Data (EHealth) Program at the Census Bureau. We rely on several data sources: area-based measures from publicly available American Community Survey (ACS) 5-year estimates (2016-2020); individual socioeconomic measures from restricted ACS microdata (2005-2022); health outcomes from electronic health records (EHRs) as part of the American Family Cohort (AFC); and mortality data from the Census Numident. We aim to identify the best (most predictive) area-based measures that are equally predictive across key domains of inequity (age, gender, race and rurality), and most closely match prediction from individual-level social indicators (i.e. household poverty and level of education).

Measurement Matters: Corroborating Hospital Surveys with Census Bureau Records

Alice Zawacki, *U.S. Census Bureau*
Atul Gupta, *Wharton School of Business*
Jackson Reimer, *Wharton School of Business*

The hospital industry accounts for the largest share of spending in the U.S. healthcare sector. According to our analysis, 50% of papers on hospitals published in top journals between 2020-2022 utilized data from the American Hospital Association's Annual Survey (AHA). However, little evidence has been brought to bear on whether the AHA survey faithfully characterizes the industry at large and if there are aspects where researchers should be more cautious. We fill this gap by name-address matching the 2002-2019 AHA to the administrative and survey data maintained in the U.S. Census Bureau's Longitudinal Business Database (LBD) and undertook four main benchmarking exercises. First, we evaluated AHA coverage of hospital establishments by characterizing facilities present in the LBD but not present in the AHA and the implications that this has for measuring aggregate entry and exit over time. Second, we determined the AHA share of hospitals being involved in a merger, a frequent AHA application, and compared this to LBD measurement of mergers. Third, we aggregated AHA records to the system-level and compare spending levels with firm-level responses to the Census Bureau's Annual Capital Expenditures Survey. Fourth, we measured concordance between AHA and LBD reports of payroll expenditures and employment. LBD firm-level revenue was compared to the Centers for Medicare and Medicaid Services cost reports, perhaps the most widely cited data on the subject. Taken together, we determined how well AHA survey data comports with administrative and survey data on many features of the hospital industry.

Return to Schedule Overview

Session I-6: Piecing Together the Puzzle: Novel Methods for Missing Data

Organizer: Yan Liu, *Internal Revenue Service*
Chair: Yan Liu, *Internal Revenue Service*
Room 0105

A Modular Approach to Survey Editing and Imputation for Agriculture Statistics

Gunnar Ingle, *Summit Consulting LLC*
Albert Lee, *Summit Consulting LLC*

Each year, the U.S. Department of Agriculture's National Agricultural Statistics Service (NASS) conducts more than 100 surveys to quantify every aspect of agriculture in the United States. The quality and consistency of the survey responses vary by survey and question. NASS is working to automate its survey editing and imputation process to improve data quality and consistency. In recent years, Summit has partnered with NASS to develop a modularized and generalizable solution that can encompass all surveys and effectively manage rules and data. This process includes (1) creating and developing deterministic edit and questionnaire routing rules, (2) imputing item nonresponses using an Expectation-Maximization-like imputation method, (3) automating error resolution based on the Fellegi-Holt algorithm, and (4) executing validation checks to determine if analyst intervention is needed. Summit executed a series of R scripts to extract survey responses, send them through the edit and imputation algorithms, and display post-edit results to users via a user interface. This paper describes the processes and methodologies used in each step, and provides the highlights and challenges of developing such a system.

A Sequential Imputation Approach for Sensitive Survey Question

Andres Mira, *U.S. Census Bureau*

Sensitive survey questions tend to fail the missing at random assumption at the center of the common imputation procedures used to deal with item nonresponse. This project proposes and compares the quality of an enhanced imputation procedure that incorporates para-data information as an alternative to the current hot-deck imputation method used by the Census Bureau. Contact mode and contact attempts are used proxies for sensitivity an individual has to the survey. This allows item nonrespondents to be matched to similarly sensitive respondents when assigning imputed values. The focus of this project will be on the citizenship question asked in the American Community Survey as it is a sensitive question and used to produce official statistics on the size of the Foreign-born and citizen population of the US. Imputed values from the traditional imputation method and the enhanced method will be compared to linked citizenship status from administrative records.

Enhancing Administrative Tax Record Imputations through Machine Learning: Utilizing Workgroups to Research Machine Learning to Strengthen Production Processes

Katherine Ann Willyard, *U.S. Census Bureau*
Mark Frame, *U.S. Census Bureau*
Ming Ray Lia, *U.S. Census Bureau*
Angelica Phillips, *U.S. Census Bureau*
Amelia Ingram, *U.S. Census Bureau*
James HoShek, *U.S. Census Bureau*
Albert Nedelman, *U.S. Census Bureau*
Sam Shirazi, *U.S. Census Bureau*

Many government agencies are seeking to better understand how Artificial Intelligence (AI) can enhance production processes. Within the Census Bureau's Social, Economic and Housing Statistics Branch, a workgroup was developed to understand how machine learning could further enhance Small Area Income and

Poverty Estimates (SAIPE), which are estimates of the poverty status of school age children in families in poverty used in Title I allocations. The purpose of this work is two-fold. First, how the workgroup was organized to research enhancements to production processes, what research questions were considered, and how a final research question was chosen will be described for the purpose of informing others how workgroups can be utilized to develop machine learning enhancements to production processes. Second, we will compare machine learning and non-machine learning methods and results for imputing block geocodes on individual tax returns for the purpose of demonstrating the extent to which machine learning methods enhances administrative tax record imputations for SAIPE 2022. This work will be of interest to both small area estimation and machine learning researchers, as well as those unsure of how to begin understanding how AI can be used in their work.

Exploring the feasibility of imputation techniques for the Commodity Flow Survey (CFS)

Gritiya Tanner, *U.S. Census Bureau*
Elizabeth Jeninga, *U.S. Census Bureau*

The Commodity Flow Survey (CFS) is a joint effort by the Bureau of Transportation Statistics (BTS and the U.S. Census Bureau to collect information on how U.S. establishments transport raw materials and finished goods. Data was collected on the types of commodities shipped, mode of transportation, shipment value, shipment weight, origin, and destinations of the shipments (including exports). For 2022 survey cycle, the collected data has increased 16 times from previous cycle and mode of transportation was frequently missing which required a shift in imputation strategy. The data currently is edited and/or imputed using rule-based approaches. This presentation will describe the preliminary research that we explore to expand the imputation methodology for missing mode of transportation in the Commodity Flow Survey (CFS) using supervised machine learning algorithm.

Integrating Transformer-Based Architecture and Transfer Learning for Ethnicity/Gender Imputations on Census and Voter File Data

Chris Esposo, *Murmuration.org and Georgia Institute of Technology*
Nancy Zhao, *Murmuration.org*
Francisco Martinez, *Murmuration.org*

This paper introduces a transformer-based architecture designed to predict ethnicity and gender from names using the natural language text of those names as well as the phonetic representations of those names. We contextualize the model and the results within a comprehensive literature review on name-only ethnicity models, both in classical machine learning and deep neural networks. Notable instances of these include Chintalapati, Laohaprapanon, and Sood (2023), who authored the well-known 'ethncolr' package for Python and R, and Jain, Enamorado, and Rudin (2022), who leveraged non-neural network techniques in favor of greater interpretability. Our model, trained on first, middle, and last names and ethnicity labels collected through the Voting Rights Act of 1965, achieves state-of-the-art results both for name-only imputation tasks, as well as a modeling task that includes tabular attributes about those individuals from the L2 voter files. We validate these results using supplementary survey data. The model's novel architecture leverages spatial entropies and additional tabular attributes from Census to reconstruct community profiles effectively.

In a second portion of the study, we also address the challenges of integrating many disjointed Census flat files together to achieve a superior prediction model via the technique known as transfer learning. This approach provides avenues of approach in building performant models that don't require explicit join-keys between the disparate datasets, enabling robust supervised learning under limited data conditions. It also opens the discussion on what is reconstructable about communities from federal data even without these keys.

Using Historical Cell Means to Generate JOLTS Imputed Data

Mark Crankshaw, *Bureau of Labor Statistics*

Almost a decade ago, JOLTS replaced its original nearest neighbor hot-deck item imputation algorithm with a model-based system. The nearest neighbor algorithm that was employed by JOLTS had a number of theoretical

and empirical shortcomings. The nearest neighbor approach relied solely on the level of reported employment, which carried little JOLTS informative value, and completely disregards employment dynamics, which carried the bulk of JOLTS informative value. While the model-based system adopted to replace the original nearest neighbor hot-deck imputation approach has performed well and stood the test of time, it may be unnecessarily complex.

Return to Schedule Overview

Thursday, October 24, 2024 **1:45 PM**

Session J-1: Keeping AI out of Trouble: Guardrails and Applications for Federal data

Organizer: Brandon Kopp, *Bureau of Labor Statistics*
Chair: Brandon Kopp*, Bureau of Labor Statistics*
Chesapeake A

Chatbot Evaluation: Methods and Challenges

Ruhan Circi, *American Institutes for Research*
Bhashithe Abeysinghe, *American Institutes for Research*

Chatbot development is rapidly evolving due to advances in Large Language Model (LLM) APIs, with federal agencies launching LLM-based applications, as exemplified in Executive Order 13960. Creating reliable LLM-powered applications requires careful consideration of performance and ethical standards, as highlighted by authors like Srivastava et al. (2023). Evaluating these applications is crucial, incorporating both technical assessment and trust-oriented frameworks.

LLMs, fundamental to many applications, pose unique challenges in chatbot development, such as hallucination and tone issues. Given the widespread adoption of LLM-based generative applications like chatbots, robust evaluation systems are therefore essential. Two main evaluation approaches, automated metrics, and human evaluation, are commonly used in tandem. Considering time and cost constraints, balancing evaluation procedures without sacrificing accuracy is vital.

Human evaluation, a crucial aspect that often needs to be well-detailed in research papers, remains integral in the context of chatbot development. Additionally, diverse metrics assess different aspects of chatbot responses, yet the specificity of testing data is often overlooked. Incorporating various question types informed by cognitive psychology frameworks could enhance systematic evaluation.

In our research, we propose improvements in evaluation techniques, informed by a cognitive psychology framework, to enhance chatbot response assessment. We will present our findings on an experimental chatbot utilizing Retrieval Augmented Generation (RAG) and Vector Databases, which contribute to the ongoing advancements in chatbot development.

Impact of Artificial Intelligence in the US Federal Government Agencies

Ujjayini Das, *University of Maryland*
Srijeeta Mitra, *University of Maryland*

Artificial intelligence (AI) can enhance responsive and efficient public services, with many federal agencies exploring and funding AI solutions for administrative tasks. Despite rapid advances, academic research on how these agencies acquire and regulate AI is sparse. The AI Impact Project seeks to address this by surveying U.S. federal employees about their attitudes toward AI, assessing AI's effects on government operations, civil service, and citizens. The survey, administered via Qualtrics, targets Federal employees drawn from the University of Maryland Alumni pool and researchers' LinkedIn connections. Though non-probabilistic, it aligns with population distribution via demographic linking variables from the Federal Employee Viewpoint Survey.

The AI Impact Survey assesses perceptions of trustworthy AI, covering fair use, confidentiality, system reliability, cybersecurity, and transparency. It also measures views on AI's public administration benefits and concerns, like safeguarding individual confidentiality through synthetic data generation while preserving data quality.

Rising concerns about potential biases from AI implementations prompt a critical question: How can we mitigate these biases with AI in use? While Federal agencies can greatly benefit from a knowledgeable adoption of AI for timely data release, the ever-evolving nature of AI necessitates questioning whether current training is sufficient for employees to fully leverage AI while maintaining data quality. Providing an in-depth look at AI adoption in the Federal govt. and it's employees' fluency in using AI at their workplace, the survey motivates policy discussions to enhance AI training and implement effective development programs in governmental agencies to promote higher quality data generation.

Considerations for Defining a Framework for Reproducibility in Survey Data Processing

Kiegan Rice, *NORC at the University of Chicago*
Stas Kolenikov, *NORC at the University of Chicago*
Amy Ihde, *NORC at the University of Chicago*
Quentin Brummet, *NORC at the University of Chicago*
Karen Grigorian, *NORC at the University of Chicago*
Lauren Seward, *NORC at the University of Chicago*

Carefully considering the transparency, clarity, and reproducibility of data processing tasks during preparation of survey data files is a key component of ensuring the quality of produced data files. The importance of considering these aspects of analytic work has been emphasized by multiple recent NASEM reports ("Reproducibility and Replicability in Science"; "Transparency in Statistical Information"). Clearly documenting and structuring data processing tasks allows multiple analysts to complete a data processing task, re-run it at a later date, or recomplete the process on a different machine or in a different computing environment. This is even more important in multi-year surveys, where rigorous and reproducible data processing ensures process consistency and reuse across years. Reproducibility improves credibility of data products, also mapping directly to coherence and scientific integrity of the FCSM Data Quality Framework. This presentation describes the implementation of reproducible data processing practices in the Entrepreneurship in the Population (EPOP) Survey, a five-year survey project focused on measuring entrepreneurship activity in the U.S. We first discuss a systematic review during the second year of the survey which identified areas of strength and areas for improvement on reproducibility and process documentation. This systematic review led to the creation of four main principles of reproducibility on analytic tasks and the design of a software-agnostic framework for ensuring reproducibility of work across teams and years. We conclude by discussing lessons learned from this work and broader considerations for defining and implementing a reproducibility framework in survey data processing.

Algorithmic Bias: Developing a Multidimensional Framework

Ruhan Circi, *American Institutes for Research*
Juanita Hicks, *American Institutes for Research*

In our rapidly advancing technological era, algorithms are used to extract meaning from data and play an increasingly important role in decision-making processes across federal agencies. The potential for the use of data is undeniable in any industry with a growing number of innovations in methods and technology. Although machine learning (ML)/Artificial intelligence (AI) has the promise to improve the outputs and reduce inequity (e.g., in health, Feehan et al., 2021), in practice, they are also prone to the risk of propagating existing gaps or having unintended harm by creating a feedback loop between data, algorithms, and users (e.g., Hooker, 2021).

Researchers approach algorithmic bias from two perspectives: a) technical and b) non-technical. Technical perspectives focus on data and model choices, architecture, and loss functions (e.g., Lalor et al., 2024). Non-technical perspectives cover transparency, interpretability, accountability, and ethical aspects (e.g., Cecere et al., 2024). Algorithmic bias is a multidimensional issue requiring an effective framework that adopts a diverse

and multidimensional perspective.
Our framework synthesizes insights from the growing literature on algorithmic bias by developing an index that takes both technical and non-technical approaches. We examine its implications and identify strategies to mitigate associated risks in the context of survey practices. We discuss our framework, which recognizes the complex nature of algorithmic bias and ensures the promotion of fairness throughout ML/AI systems' lifecycles. We illustrate the application of our framework using a real federal dataset on decision-making algorithms.

Official Statistics for Responsible AI: The Role of the Federal Statistical System in Enabling a More Accountable AI/ML Ecosystem

Tomo Lazovich, *U.S. Census Bureau*
Michael Walton, *U.S. Census Bureau*
Atul Rawal, *U.S. Census Bureau*
Anna Vasylytsya, *U.S. Census Bureau*
Curtis Mitchell, *U.S. Census Bureau*
Diamond Nwankwo, *U.S. Census Bureau*

With the popularity of generative AI and the spread of the use of algorithmic tools in critical settings such as finance and healthcare, the U.S. government has increasingly focused on ways to mitigate the potential harms of data-driven systems. The Blueprint for an AI Bill of Rights, NIST AI Risk Management Framework, and recent Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI have all emphasized the need for government to better understand the societal impacts of AI and algorithmic decision systems. While there has been a great focus on enabling development of AI and machine learning systems within the U.S. Federal Statistical System (FSS), there has been comparatively little thought about FSS's opportunity to facilitate broader adoption of responsible AI (RAI) practices. As the country's premier source of statistical information, the FSS is in a unique position to enable understanding of the social impacts of AI systems, particularly in sensitive and regulated areas like housing, credit, healthcare, and employment. In this policy paper, we propose three ways that FSS agencies can enhance the RAI ecosystem: 1) data collection on societal impacts of AI through existing or novel survey products; 2) creation of computational tools and educational materials to encourage use of federal open data, including demographic data, for algorithm auditing and responsible AI practices; and 3) research across agencies on AI decision systems as drivers of social inequality, with a focus on employment, credit, housing, and healthcare.

Return to Schedule Overview

Thursday, October 24, 2024 1:45 PM

Session J-2: Communicating Quality in an Evolving Federal Data Ecosystem

Organizer: John Finamore, *National Center for Science and Engineering Statistics*
Chair: ICSP Member (TBD)*, ICSP Member (TBD)*
Chesapeake B

An Agency's Holistic Quality Communication Strategy (Covering Both Internal and External Communication Efforts)

, *National Center for Science and Engineering Statistics*

Communicating Quality Associated with an Experimental Data Product

, *U.S. Census Bureau*

Communicating Quality Associated with an Access Tiered Approach to Data

Dissemination (Synthetic Data; Validation Server)
IRS-SOI,

Communicating Quality Associated with Alternative Data Sources
, *TBD*

Return to Schedule Overview

Thursday, October 24, 2024 1:45 PM

Session J-3: Advancing Equity in Survey Research: Innovations in Survey Research and Assessment

Organizer: Robin Bachman, *U.S. Census Bureau*
Chair: Robin Bachman*, U.S. Census Bureau*
Chesapeake C

CDC's Youth Risk Behavior Survey to DOE's Office for Civil Rights Civil Rights Data Collection: Using Innovative Data Linkage to Address Health Equity
Kathleen Krause, *Centers for Disease Control and Prevention*
Jonetta J. Mpofu, *Centers for Disease Control and Prevention*

Background. For at least the past 30 years, the U.S. Department of Education's Office for Civil Rights' Civil Rights Data Collection (CRDC) has found that school expulsion practices have disproportionately targeted Black youth K-12 while the U.S. Centers for Disease Control and Prevention's (CDC) Youth Risk Behavior Survey (YRBS) has found that Black or African American students grades 9-12 (as compared to White) are more likely to skip school due to feeling unsafe. We linked these two datasets to uncover school-level structural factors influencing student health.

Objective. To describe the challenges and resolutions to data linkage, processing, survey weighting, and dissemination encountered with analysis of YRBS and CRDC data.

Methods. Nationally representative data of U.S. high school students from the 2019 YRBS were merged with the national data of school expulsion practices from the 2017–2018 CRDC (N=125 schools and N=12,779 students). Weighted random-intercept multilevel logistic regression models were created to examine the association between school expulsion practices and a student skipping school with an interaction term for expulsion*race/ethnicity.

Results. Data were linked with minimal processing issues. Data were weighted to maintain the national representativeness with some limitations. Dissemination of findings have been successful with some challenges. The analysis found a greater association between school expulsion practices and skipping school due to feeling unsafe among Black students (as compared to White).

Conclusions. Innovative linkage of these federal datasets accelerated using data for action to address systemic inequities in educational policies, practices, and environments to support positive health outcomes.

Considerations of Data Diversity and Equity in Probability Sampling and Survey Research
David Dutwin, *NORC at the University of Chicago*
Ashani Johnson-Turbes, *NORC at the University of Chicago*
J. Michael Dennis, *NORC at the University of Chicago*

In the past decade survey research has experienced an increase in focus on the need for data equity and fairness as well as robust consideration of diversity and equity. Such focus spans both diversity and equity in the

workplace as well as in the data itself. Many calls for diversity in data come from Federal sources, including the NIH, Departments of Justice and Commerce, and even the White House itself. How do we explore whether our data is equitable, and more importantly, how do we ensure that our data is diverse and equitable? This presentation provides a framework for both processes to analyze the degree to which data is rich in representation of all groups and subgroups of a given target population as well as explores methods and approaches survey researchers have utilized to increase representation and diversity across, in particular, hard to reach groups, principally across racial, ethnic boundaries and sexual identity. We focus particularly on strategies to build data equity in probability-based panels. We present a metric of sample diversity for Federal researchers to consider and provide a range of case studies to increase diverse representation in survey respondents.

Reducing Deficit Interpretations of Large-Scale Assessment Results Displays: Testing Score Reporting Approaches to Support Equity

Emily Kerzabi, *ETS*
Katherine E. Castellano, *ETS*
Mari Kevelson, *ETS*
Vinetha Belur, *ETS*
Eli Holder, *3iap*
Leslie Villegas, *New America*
Renee Savoie, *Connecticut Department of Education*
Jennifer Cain, *Minnesota Department of Education*
Molly Prower, *ETS*

This presentation will provide insights into the current state of knowledge on best practices in anti-deficit score reporting for national assessments to improve large-scale educational assessment data dissemination. It will highlight approaches to score reporting that may help to counter deficit narratives often inferred from achievement gaps. One example is presenting results in ways that emphasize structural rather than individual explanations for differences in learning outcomes, to help uncover potential levers to improve educational equity and thus better inform educational policy.

We will share findings from interviews with 20 national assessment results users, including state education agency staff, policy and advocacy group leaders, and policy researchers and experts. The interview results will address ways to make large scale assessment results displays more useful and actionable for policymakers, and on possible approaches to shift from a deficit-based to an asset-based reporting approach. We also synthesized diverse literature, including research on score reporting approaches, deficit reporting of achievement gaps, and best practices in data visualization, particularly for equity. Our presentation will focus on findings from the interviews and literature review, in addition to showcasing initial efforts to translate these findings into reporting displays co-designed with three of our interviewees, a UX/UI researcher, and a data visualization expert. The team will study the extent to which these new displays with a more asset-based focus may support the prioritization of equity-focused policies, emphasize structural vs individual reasons for differences in performance, and overall comprehensibility of assessment results displays.

What Are We Missing? Defining Historically Undercounted Populations in Economic Surveys at the Census Bureau

Melissa Cidade, *Bureau of Justice Statistics*
Hillary Steinberg, *U.S. Census Bureau*

The Census Bureau has long had interest in Historically Undercounted Populations (HUP), or groups that are underestimated in its surveys and statistical products. The current Census Bureau Strategic Plan highlights data equity as an integral part of moving the Bureau forward; HUPs are a strategic priority for the Bureau. However, these populations and the mechanisms that lead to undercounting have not been well defined in economic surveys – surveys of businesses and other institutions.

The purpose of this presentation is to propose a framework for defining HUPs in economic surveys. The proposed framework is the result of collaboration with the Census Bureau's HUP Working Group, as well as a subgroup specific to the Economic Directorate. Through shared experiences and an open forum, we have

garnered feedback on the component definitional parts of HUPs in economic surveys.

We first situate this conversation in the broader Census Bureau's efforts. Then, we enumerate how the differing context for economic surveys, (including records-based reporting, proxy reporting, and other response processes specific to establishment surveys) may present barriers to accurate counts. From there, we offer three categories of HUPs in economic surveys: coverage, sampling, and nonresponse; validity, measurement, and processing; and economic activity issues. For each activity, we give examples from existing surveys and offer methodologies to potentially address these concerns. Finally, we outline next steps for an approach to studying HUPS in economic surveys, including innovative research and calls for integrating community-based methods, where we engage demographic populations who have historically been undercounted.

Return to Schedule Overview

Thursday, October 24, 2024 1:45 PM

Session J-4: The U.S. Federal Statistical System: Why We're Here, What We Do, Who We Are, and Career Opportunities

Organizer: Bob Sivinski, *Office of Management and Budget*
Chair: Karin Orvis, *Office of Management and Budget*
Discussant: Karin Orvis, *Office of Management and Budget*
Vessey 1

Total Survey Error as a Framework for Data Equity

Sarah Forrest, *National Center for Health Statistics*

Data inequities can hinder our ability to accurately understand and address health inequities. The Total Survey Error (TSE) provides a systematic approach for identifying various sources of error throughout the survey process. This TSE paradigm provides a useful framework for understanding data inequities prior to producing population estimates. Inequities in the survey design, measurement, representation, and dissemination of data can potentially be compounded, resulting in statistics that may be less accurate, precise, relevant, or available for some groups. We propose using the TSE framework viewed through an equity lens to facilitate understanding

of potential data inequities throughout the survey process. We introduce two additional components to the TSE paradigm: design bias and dissemination bias. These biases can undermine the credibility and relevance of survey statistics for underrepresented groups. Credibility is compromised when the survey design team excludes diverse perspectives, perpetuating unchallenged assumptions of dominant groups and failing to capture varied experiences accurately. Relevance is at risk when statistics are not equitably

disseminated or accessible for all groups, especially those that disproportionately suffer from inequities in outcomes.

The equity-driven perspective on the TSE framework underscores inclusivity and representation as foundational principles for building trust in survey research. This framework accounts for survey design and dissemination biases and serves as a decision aid for producing more equitably accurate, credible, and relevant statistics. This approach ensures that survey research is not only methodologically sound but fosters more equitable data practices that honor the diversity in our communities.

Government Data of the People, By the People, For the People: Navigating Citizen Privacy Concerns

Claire McKay Bowen, *Urban Institute*

In a data-driven era, data users and researchers frequently leverage personal or confidential data to help

policymakers make evidence-based, data-informed decisions, such as improving economic recovery or creating a more efficient COVID-19 vaccine distribution. However, access to confidential data comes with several privacy concerns, especially for underrepresented groups. Striking the right balance is crucial to avoid real disclosure risks that may not be obvious, like stalkers utilizing excessive location data or malicious parties learning sensitive medical information through linkable health or genetic data. This talk provides an overview of the various technical solutions that attempt to address the intricate data privacy challenges faced by the U.S. government and private sector in data collection and dissemination.

How the U.S. Statistical System is Expanding Secure Access for Evidence Building

Emilda Rivers, *National Center for Science and Engineering Statistics*

As the challenges facing the Nation continue to evolve and become more complex, so does the information required to inform decisions and policymaking. Addressing the information needs of the Nation efficiently and effectively require coordination and collaboration within the Federal statistical system and across a broad set of partners in the data and evidence ecosystem. To inform this need for coordination and collaboration, the federal statistical agencies are collectively leading work to increase the use and availability of government data to inform policy, while protecting privacy and confidentiality of those data. This presentation will share significant developments that have occurred in recent years, such as the establishment of a Standard Application Process for requesting secure access and the conducting of innovative data linkage and privacy protection research through a National Secure Data Service demonstration). Current challenges and future opportunities will also be discussed, as well as areas where greater engagement, input, and perspectives from stakeholders are needed.

Unlocking New Insights through Restricted Microdata: 30 Years of Innovative Research from Federal Statistical Research Data Centers

Nate Ramsey, *U.S. Census Bureau*

The Census Bureau opened its first Research Data Center in 1994 with the goal of delivering programmatic and scientific benefits to the agency and the public by allowing qualified researchers access to internal microdata to conduct approved statistical analyses. Thirty years later, the program – rebranded in 2015 as the Federal Statistical Research Data Center (FSRDC) network – hosts projects and provisions data for 10 federal agencies, operates 33 secure centers nationwide, partners with hundreds of local research institutions, and manages a portfolio of over 550 projects and 1,200 researchers. This presentation will provide an overview of the FSRDC program and how it implements a highly secure network that upholds its legal and ethical responsibilities to protect respondent privacy and confidentiality. It will review how the process works for researchers, from submitting project proposals via the Standard Application Process to clearing statistical output. And it will highlight successes this collaborative effort has delivered over the years and future plans involving additional physical locations, virtual access, new agency partnerships, and increasing equitable access to data.

Return to Schedule Overview

Session J-5: Let's Link Up: Data Integration Insights, Innovations, and Impacts

Organizer: Alyssa Holdren, *Bureau of Economic Analysis*
Chair: Alyssa Holdren*, Bureau of Economic Analysis*
Vessey 2

A Large Scale, High Quality U.S. Occupational Database: Results from Merged IRS and ACS Write-Ins

Carl Sanders, *U.S. Census Bureau*
Victoria Bryant, *Internal Revenue Service*
David Grusky, *Stanford University*
Michael Hout, *New York University*
Lynda Laughlin, *U.S. Census Bureau*
Ananda Martin-Caughey, *Brown University*
Javier Miranda, *University of Jena*
Kevin Pierce, *Internal Revenue Service*
Christin Landivar, *U.S. Census Bureau*
Bryce VanderBerg, *U.S. Census Bureau*

Measuring worker occupations is crucial for understanding long-term U.S. socio-economic changes. Yet, challenges like incomplete data coverage and noisy survey responses persist, hindering effective measurement. In this project, we develop a comprehensive occupation database by merging Census Bureau and IRS data, overcoming these obstacles to create a large, high-quality, and linkable resource of individual worker occupations. We align Federal Tax Information (FTI) data from all electronically filed tax returns with ACS survey data from 2019. This comparison of individual occupation responses between the two sources aims to grasp their respective strengths and weaknesses. Utilizing Token Set Ratios (TSRs) for approximate string matching, we assess write-in responses in relation to age, gender, earnings, and other factors. Subsequently, we investigate the sensitivity of each write-in to year-over-year occupational transitions. We find that the overall quality of matches across the two write-in data sets is bimodal, with over 50% of the sample having a high-quality match, and another 25% of the sample with what seems to be a no-match. The average write-in match score varies significantly over occupation and age. Results from a regression model indicate that quality of matches is positively associated with higher socio-economic

status and the ability of the U.S. Census's probabilistic identity matching system to identify the worker. Finally, after creating a two-year panel, we find that the IRS and ACS report significantly different occupational mobility rates, with IRS more conservative and ACS more liberal in assigning changes.

A Preliminary Study to Evaluate Variance of Blended Index from Establishment Survey and Administrative Data

Daniel Yang, *Bureau of Labor Statistics*

The International Price Program (IPP) at the Bureau of Labor Statistics (BLS) produces Import and Export Price Indexes (MXPI) which consists of two components: import and export merchandise. The MXPI is implementing a transition to replace a portion of the directly collected establishment survey data with Census Trade Data (CTD). CTD is administrative trade transaction data that includes specific shipment records, such as international Harmonized System (HS) product classification codes for the U.S., trade values in dollar, foreign country of import or export, shipment quantity, etc. CTD has the advantage of covering almost all U.S. trade within a month. However, the price that is collected by CTD is an average price of shipment which is different from the directly collected survey. MXPI established a Bootstrap resampling process on the directly collected survey data to evaluate the sampling error, meanwhile, the types of errors in CTD have not been considered in the past. In this study, we explored resampling methods to estimate the variance of index blended from two

data sources: an establishment survey and a census.

Non-random PIK Assignment and Selection into Linked Data: Implications for Research Using Linked Data

Kyle Raze, *U.S. Census Bureau*
Nicole Perales, *U.S. Census Bureau*

The U.S. Census Bureau's Person Identification Validation System facilitates anonymous linkages between survey and administrative records by assigning Protected Identification Keys (PIKs) to individuals. However, previous research demonstrates that the likelihood of being assigned a PIK varies by several socioeconomic and demographic characteristics. While the potentially non-random nature of PIK assignment is generally acknowledged by users of PIK-linked data, less is known about the systematic drivers of selection into PIK assignment, the magnitude of the bias induced by this selection, and the efficacy of methods to mitigate the bias. Using the American Community Survey and IRS 1040 filings, we comprehensively document differences in PIK assignment over nearly two decades and test for imbalances in the composition of linked samples. We then quantify the magnitude of bias in income measurement for linked samples and evaluate the performance of several methods in mitigating the bias, including inverse probability weighting, entropy balancing, empirical likelihood weighting, and Manski bounding. We evaluate these methods in both rich and sparce variable spaces, as well as in the presence of high and low PIK rates, to provide practical guidance for researchers and improve inference from linked data.

Salvaging Data from an Incomplete Sample Through Statistical Data Integration

Wendy Van de Kerckhove, *Westat*
Tom Krenzke, *Westat*
Benjamin Schneider, *Westat*

Incomplete survey data can arise when there are unexpected disruptions to data collection, such as the recent pandemic. The result is a sample that is a product of the probability-based sample design, the non-probabilistic mechanism that determined which sampled cases were worked prior to the disruption, and nonresponse. We describe a method used in the U.S. Program for the International Assessment of Adult Competencies (PIAAC) for combining incomplete survey data with complete survey data. The sample design consisted of a nationally representative core sample and a supplemental state-based sample, where the goal of the supplemental sample was to improve small area estimates. Given the uncertainty in response rates and potential pandemic-related disruptions, the design allowed for the flexibility to shift fieldwork effort from the supplemental sample to the core sample as needed. In fact, data collection for the state supplement was halted less than halfway into the data collection period, before interviewers had visited all areas. To salvage the collected data from the incomplete supplemental sample, we combined it with the core sample by using a composite weighting technique. We describe the weighting methods and an evaluation of the combined data.

Tax Education is a Form of Wealth Insurance: Linking FFIEC data, IRS Gig-Ecosystem Data and Digital Divide Data to Identify Tax Education Outreach Needs

Barbara J. Robles, PhD, *Federal Reserve Board (Retired)*
Caroline Bruckner, *American University*
Marjorie E. Kornhauser, *Tulane University*

As a combined policy and research focus, we link publicly available administrative data that identify the potential need for tax education and enhanced outreach by employing FFIEC data that identifies under-served counties linked to IRS (Internal Revenue Service) county-level data that captures tax status. Gig or free-lance workers have grown exponentially since the Great Recession and the Covid pandemic, (MBO Partners, 2024; Abraham, et al, 2023; Bruckner and Coil, 2024). We identify counties that may be experiencing an increase in 'opportunity jobs and income supplementing' while also examining the combined complexity of their tax obligations. In addition, our linked-data analysis is a novel approach that provides a non-survey, evidence-based understanding of the growth in hybrid income generation focused on place-based locales that may need more concentrated tax outreach services:

(i) FFIEC distressed and under-served county data and CRA (Community Reinvestment Act) business loan originations by county (publicly available, 2021)

(ii) IRS county level data for 2021 that capture industry sectors and cohort filing data (filers using: paid tax preparers vs. free tax-filing volunteer prepared sites, pensions, taxable social security, self-employment, number of returns with business income, number of returns filed electronically, number of returns elderly, number of returns computer prepared, etc.)

(iii) Digital Divide County Level data – Census county level data where fixed broadband service delivering advertised speeds of at least 100 Mbps download and 20 Mbps upload is not available.

Our investigation provides insight into policy and spatial information that may enhance tax compliance by identifying tax literacy/education gaps

Return to Schedule Overview

Thursday, October 24, 2024 1:45 PM

Session J-6: Asking Good Questions: Survey Instrument Design Considerations

Organizer: Elina T. Page, *Economic Research Service*
Chair: Elina T. Page*, Economic Research Service*
Room 0105

CDC's Youth Risk Behavior Surveillance System and Improved Usefulness of School-based Surveillance Data

Jennifer C. Smith-Grant, *Centers for Disease Control and Prevention*
Shari Shanklin, *Centers for Disease Control and Prevention*
Kathleen H. Krause, *Centers for Disease Control and Prevention*

Background. For the past 30 years, the Centers for Disease Control and Prevention (CDC) has administered the school-based Youth Risk Behavior Survey (YRBS) biennially in high schools across the nation to assess behaviors and experiences among America's youth. To ensure that the YRBS is able to monitor issues of emerging importance, CDC revises its questionnaires each administration cycle by adding, deleting, and modifying questions.

Objective. To describe the process involved in revising and adding questions to the standard and national YRBS questionnaires, contributing to the usefulness of YRBS data.

Methods. During the two-year span between YRBS administration cycles, CDC accepts new question and question revision submissions from health agencies, educational organizations, funding recipients, and a host of other partners and colleagues. External subject matter experts are sometimes called upon for additional input on appropriate wording. Approved changes are then voted on through a formal ballot process by representatives from health and education agencies funded to administer YRBS. For new or substantially revised questions, CDC conducts cognitive testing.

Results. For the 2023 YRBS national questionnaire, there were 8 modifications, 14 new questions, and 6 deleted questions, compared to the 2021 questionnaire. The standard high school questionnaire had 4 deleted questions, 4 new questions, and 8 modified questions. Additionally, 10 questions were cognitively tested.

Conclusions. Keeping YRBS data up to date by including items of emerging interest serves to improve data usefulness. Revising questionnaires also promotes using data for action.

Developing More Effective Interview Materials for a World with Increasing Telephone Surveys

Ariana Welsh, *Bureau of Labor Statistics*
Erica Yu, *Bureau of Labor Statistics*

The Consumer Expenditure Interview Survey is collected by the Census Bureau on behalf of the Bureau of Labor Statistics. It was originally designed as an in-person interview but has increasingly moved to telephone interviewing in recent years. The information booklet is an informational guide, similar to showcards, designed to aid respondents in the survey by providing examples of the expenditure items targeted for each question. Originally, the information booklet (infobook) was designed for use as a paper booklet to be viewed by the respondent during in-person interviews; however, with the increase in telephone-conducted interviews, Census interviewers can point respondents to a PDF version of the infobook that is available on the BLS website. Evidence suggests that the PDF infobook suffers from low use, with responses to a field staff survey in 2021 reporting that 80 percent of interviewers do not use it in their interviews. In 2021, CE proposed a redesign of the online PDF infobook to make it more conducive for use in telephone-conducted interviews. Across two pre-testing studies, feedback was collected from 45 participants on various experimental design features of the infobook. These features were tested in tandem with different styles of interview approaches. In this presentation, the redesign and results of these studies regarding the new online infobook's features will be discussed.

Do Micro-surveys get Respondents in the Door?

Elise Christopher, *National Center for Education Statistics*
Sean Simone, *National Center for Education Statistics*

For its recent field test for the First Follow-up of the High School and Beyond Longitudinal Study of 2022 (HS&B:22), NCES experimented with different data collection strategies for parent surveys. In the Base Year (2022-23 school year), parent response rates were the lowest among the 5 different respondent groups for the study. That is, they were lower than those of high school students, their math teachers, school counselors and school principals. Because parents provide important demographic data about their high schooler sample members, eliminating the parent questionnaire entirely would leave data gaps in key variables. A frequent reason for survey refusal reported to interviewers is that parents are "too busy" or "don't have the time" to participate. While an abbreviated questionnaire is often offered toward the end of the data collection window for NCES sample surveys as a refusal conversion technique, few studies have tried drastically shortening the questionnaire at the outset of data collection. Considering this, NCES explored options for dramatically minimizing the time burden requested to test whether reducing the burden might encourage more parents to respond.

To test procedures to boost response rates and/or minimize missingness for key variables, NCES experimented with data collection utilizing micro-questionnaires in paper-and-pencil interview (PAPI), computer-assisted telephone interview (CATI) and web-based interview modes. Results are discussed in terms of response rates, data quality and potential for bias. Implications for further implementation of micro-surveying sum up the discussion.

What Can Prompt Engineering Learn From Questionnaire Design?

Lilian Huang, *NORC at the University of Chicago*
Brandon Sepulvado, *NORC at the University of Chicago*
Elizabeth Dean, *NORC at the University of Chicago*
Joshua Lerner, *NORC at the University of Chicago*
Ipek Bilgen, *NORC at the University of Chicago*
Leah Christian, *NORC at the University of Chicago*

There is an extensive body of literature on questionnaire design, and the recent field of prompt engineering for Large Language Models (LLMs) presents a parallel, as users make alterations to prompt wording to refine and obtain desired output from these generative AI models. While prompt engineering is a rapidly growing field, its literature and practices are still nascent. As such, we look to survey literature for insights that may guide us

in developing a systematic approach for interacting with LLMs to derive optimal output.

Through exploring a series of real-world examples in LLM research, we demonstrate how prompt engineering methods can overlap with the more firmly codified best practices of questionnaire design, while noting key differences in their overall aims and specific strategies. For example, common prompt engineering practices include: providing detailed background context; giving specifications for output format and length; including examples of desired results; and instructing the model to adopt a specific persona.

Meanwhile, questionnaire designers typically aim for a 5th to 9th grade reading level, and consider recency effects when listing response options, to prevent respondents from being biased towards the first or last option. However, they generally avoid providing respondents with specific personas or examples, and questions are kept as short as possible to minimize cognitive burden – in contrast with the technique of supplying LLMs with richer context.

By bringing these different approaches into conversation, we show that survey design can fruitfully inform how researchers carry out prompt engineering.

Return to Schedule Overview

Thursday, October 24, 2024 3:30 PM

Session K-1: Using AI and ML to Improve Data Analytics 2

Organizer: Drake L. Gibson, *Department of Homeland Security*
Chair: Drake L. Gibson, *Department of Homeland Security*
Chesapeake A

Advanced Machine Learning Techniques for Tobacco Cessation Prediction in Complex Survey Data: A Case Study Using the PATH Study

Hanxia Li, *Univerisity of Oklahoma Health Science Center*
Sixia Chen, PhD, *University of Oklahoma Health Sciences Center*

Background: Addressing the analytical challenges posed by complex survey data, this study applies advanced machine learning (ML) techniques to the Population Assessment of Tobacco and Health (PATH) Study data, aiming to predict tobacco cessation behaviors.

Data: Utilizing the PATH Study's rich, longitudinal data, which encompasses detailed information on tobacco use and cessation, the study navigates through the complexities of stratified sampling and weighting inherent in federal statistical analysis.

Methods: A range of ML algorithms suited for complex survey structures, including weighted random forests and adaptive boosting, are employed. Key steps include preprocessing for survey design considerations and robust feature engineering to capture the nuanced, longitudinal data aspects.

Results: Our ML models identify critical predictors of tobacco cessation, enhancing prediction accuracy and ensuring representativeness for the broader population.

Implications: This research demonstrates the value of ML in deriving insights from complex survey data, providing federal agencies and public health policymakers with actionable intelligence for designing targeted tobacco cessation interventions. It underscores the efficacy of ML in complex survey data analysis, offering a model for similar research endeavors.

Application of Machine Learning Approaches to Predict Membership in ARMS NOL Frame

Bayazid H. Sarkar, *National Agricultural Statistics Service*
Peter Quan, *National Agricultural Statistics Service*

The Agriculture Resource Management Survey (ARMS) is administered annually by the United States Department of Agriculture's National Agricultural Statistics Services to ascertain U.S. farm and ranch production practices, resource use, and economic information. The survey is administered in three phases: Phase 1 is a screener for Phases II and III; Phase II collects production practices and cost data; and Phase III collects farm economic and operator characteristics data. ARMS Phase III utilizes a dual frame design: List and Area frames. For the Area Frame component, the June Area Survey (JAS) is used to identify operators that are not on the ARMS List Frame. These specific operations comprise the ARMS Area Frame Not on List (ARMS NOL) sampling frame.

Using the JAS to compile the ARMS NOL sampling frame is an important step to maintain sampling frame integrity and to obtain accurate and complete farm population survey indicators. Machine learning approaches will be used in this project to conduct screening level evaluations of the ARMS NOL sampling frame.

Five supervised classification methods will be employed: Random Forest, Support Vector Machine, Extreme Gradient Boosting, Neural Network, and Logistic Regression in this research study. The model building process will provide insight into the relationship between farm characteristics and ARMS NOL sampling frame. In this analysis, the performance of five different classification models will be evaluated in terms of accuracy, recall, precision, specificity and F1 scores using JAS data for Texas.

Approaches to Digital Cleaning: Blending Traditional Methods with Crowdsourcing and AI to Clean a Public Sector Establishment Survey Population Frame

Jake Soffronoff, *Institute of Museum and Library Services*
Matt Sweeney, *American Institutes for Research*
Ai Rene Ong, *American Institutes for Research*

The Institute of Museum and Library Services (IMLS) piloted a first-of-its-kind establishment survey of museums – the National Museum Survey (NMS) – 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 museum 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 ChatGPT for uncovering business URLs, using web scraping to obtain email addresses, and using the Amazon MTurk 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. Note: Neither Amazon, nor OpenAI, nor any other entity is a party to, nor a sponsor of, this presentation.

Assessing Subjective Probabilistic Expectations in Household Surveys with Audio Records

Nicolás Forteza, *Bank of Spain*
Javier J. Alonso, *Bank of Spain*
Laura Crespo, *Bank of Spain*

In survey methodology, general compliance with protocols and individual interviewer performance has been analyzed with audio recordings. This is a resource intensive task since audios listening must be performed. On the other hand, little work has been done in analyzing subjective probabilistic expectations questions. In economics, agents form expectations for unknown quantities to take decisions, and very often the research problem is to infer the subjective probability distributions that express such expectations. In this paper, we develop a state-of-the-art audio transcription and speaker diarization machine learning pipeline and apply it to audio recordings of a subjective probabilistic expectations question from the Spanish Survey of Household Finances. We first compare the variables from the pipeline with a question evaluation sheet completed by the survey team. Then, we evaluate the interviewer question reading behavior using novel natural language processing techniques. We find that the extracted audio features are useful for assessing compliance,

interviewer performance and for detecting biased responses from interviewer-induced household probabilistic expectations.

Machine Learning Classification of Product Categories in Scanner Data used in the United States Consumer Price Index

Brendan Williams, *Bureau of Labor Statistics*
Zach Whitford, *Bureau of Labor Statistics*

The Bureau of Labor Statistics (BLS) has been using a machine learning classifier for sales data used in the Consumer Price Index since publication of the March 2019 index. The initial modeling process used a "bag-of-words" approach with a logit regression to classify hundreds of thousands of products. Individual products were mapped into "Entry Level Item" categories and used to calculate lower-level indexes, which are aggregated as the constituent components of the overall CPI. The BLS was able to incorporate a large volume of corporate data into the CPI, which would not have been possible given the resource demands of manual classification. We established a production process to regularly evaluate classifications and flag low-probability predictions for review. The BLS continues to improve and expand on this process. We have identified problem points and expanded the technique to classify grocery scanner data. We continue to build on our text frequency approach to natural language processing and evaluate alternative classifiers including neural networks, support vector machines, and eXtreme Gradient Boosting (XGBoost). We explore nested classification steps as a means to address problem areas.

Return to Schedule Overview

Thursday, October 24, 2024 3:30 PM

Session K-2: Safe Data Technologies: Safely Expanding Access to Administrative Tax Data

Organizer: Claire McKay Bowen, *Urban Institute*
Chair: Claire McKay Bowen*, Urban Institute*
Chesapeake B

Synthesizing the Supplemental Synthetic Public Use File

Victoria Bryant, *Statistics of Income*
Chris Rexrode, *Statistics of Income*
Derek Gutierrez, *Statistics of Income*

US government agencies possess data that are invaluable for evaluating public policy, but often cannot not be released publicly due to disclosure concerns. For instance, the Statistics of Income division (SOI) of the Internal Revenue Service releases an annual public use file of individual income tax returns that is critical to tax analysts in government agencies, nonprofit research organizations, and the private sector. However, safeguarding this file through traditional statistical data privacy methods significantly and adversely impacts data quality. And, as the vast amount of personal information available in public and private domain combined with easy access to enormous computational power, these methods continue to chip away at the file's ultimate utility.

In this talk, we describe our approach to generating a fully synthetic representation of the individual tax data by using sequential Classification and Regression Trees and kernel density smoothing. This synthetic data file represents previously unreleased information useful for tax policy modeling. We also tested and evaluated the tradeoffs between data utility and disclosure risks of different parameterizations using a variety of validation metrics. The resulting synthetic data set has high utility, particularly for summary statistics and microsimulation, and low disclosure risk.

Developing a Synthetic IRS SOI Public Use File with Tidysynthesis and Syntheval

Gabriel Morrison, *Urban Institute*
Aaron R. Williams, *Urban Institute*

Access to high-quality data is essential for building knowledge and evidence-based decision making across many domains including tax policy research. The Internal Revenue Service (IRS) supports this research by releasing individual income return data through its Public Use File (PUF). The IRS has protected taxpayer confidentiality by statistically altering records in the PUF. Data synthesis for statistical disclosure control is an alternative method for safely expanding access to administrative microdata while reducing disclosure risks. The IRS has adopted synthetic data generation and is moving to release more recent years of data in synthetic versions of the PUF.

This talk describes two R packages being used by the IRS to generate and evaluate their synthetic data. The tidysynthesis R package sequentially synthesizes data with parametric and nonparametric models. The talk highlights tidysynthesis's ability to apply mid-synthesis variable constraints and infuse noise into estimates, two cutting-edge features. The talk also presents the syntheval R package, which contains tools for evaluating synthetic data. The talk then moves to describe how IRS and Urban Institute data scientists are collaborating to generate and evaluate a synthetic PUF with tidysynthesis and syntheval. The talk concludes with future directions for development for the packages and for the IRS's synthetic PUF.

Building an Automated Validation Server Prototype

Graham MacDonald, *Urban Institute*
Erika Tyagi, *Urban Institute*
Josh Miller, *Urban Institute*
Silke Taylor, *Urban Institute*

The Urban Institute built a prototype of the first automated validation server in 2021. "An automated validation server is a digital tool that creates an intermediate layer between a researcher and the original, confidential data. With this intermediate layer, a researcher can analyze the confidential data without actually seeing them. [An approved] researcher would submit their analysis code to the validation server, which would return a table of results, either showing the tabulations or regression results from the analysis with noise added to protect privacy." (MacDonald 2024) The goal is to create a new tool for data analysts that can significantly expand and streamline access to high quality information necessary to inform public policy.
After extensive testing and user feedback, the team made substantive improvements and launched a second prototype of the validation server in early 2024. The upgrades enable researchers to use a more familiar programming language (R), conduct a wide range of common analyses that weren't previously possible under strict privacy protections, and improve the usability of the privacy budget mechanism. Built in the Amazon Web Services cloud to meet the highest federal government security standards, the team believes this innovative new technology is ready for user testing that could inform a future National Secure Data Service. The team hopes to continue to improve the algorithm – addressing issues such as speed, noise calculations, and sensitive error messages – in the future, and plans to open source the code and provide a full technical paper later in 2024.

Privacy Considerations and Challenges of the Validation Server

Joshua Snoke, *RAND*
Claire McKay Bowen, *Urban Institute*
Aaron R. Williams, *Urban Institute*
Andres Felipe Barrientos, *Florida State University*

This talk describes the complex set of privacy considerations and challenges that arise with creating a validation server. Validation servers allow users to query statistics from underlying confidential data, and we utilize formal privacy concepts to reduce the need for manual disclosure review, which can be costly and time consuming. We address the various aspects of the query system which produces privacy concerns that must be addressed, such as defining a privacy mechanism, allocating a privacy budget, handling errors, and helping users understand the impact of privacy noise on their statistics. The findings of this talk are based on an ongoing

four-year project to develop a prototype validation server for IRS administrative tax data with an interdisciplinary team of social scientists, software engineers, and statisticians.

Return to Schedule Overview

Thursday, October 24, 2024 3:30 PM

Session K-3: Recent Methodological Findings and Advances in Probability-based Panel Research

Organizer: Katherine Irimata, *National Center for Health Statistics*
Chair: Katherine Irimata, *National Center for Health Statistics*
Chesapeake C

Comparison of Data Quality from Two Online Surveys from the Rapid Surveys System

Bill Cai, *National Center for Health Statistics*
Katherine Irimata, *National Center for Health Statistics*
Yulei He, *National Center for Health Statistics*
Jim Dahlhamer, *National Center for Health Statistics*
Van Parsons, *National Center for Health Statistics*
Rong Wei, *National Center for Health Statistics*

Obtaining timely and critical health-related data have become increasingly important. Existing population health surveys, which provide the highest level of data quality, are not able to provide this information as fast as commercial survey panels because of longer planning, data collection, and processing times. However, there is a lack of comparison between different panels in the literature, especially in situations where surveys from both panels collect the same information on the same target population. In 2023, two rounds of the National Center for Health Statistics Rapid Surveys System were conducted using two panels - AmeriSpeak (conducted by NORC) and KnowledgePanel (conducted by Ipsos Public Affairs) – which fielded identical questionnaires in the same time period. This research investigates how the estimates from the two panels differ from each other and from the 2023 National Health Interview Survey (NHIS) estimates, which is used as a reference survey. A strategy is also proposed to best utilize survey data from two panels to produce estimates close to population estimates from the NHIS.

Comparison of Two Data Collection Methodologies from a Web Panel Survey

Rong Wei, *National Center for Health Statistics*
Katherine Irimata, *National Center for Health Statistics*
Jim Dahlhamer, *National Center for Health Statistics*
Van Parsons, *National Center for Health Statistics*

Commercial probability panels have developed data collection strategies to mitigate potential nonresponse and coverage biases, including conducting nonresponse follow up at the panelist recruitment stage, sending survey reminders to panelists during the field period, and in some cases offering a tablet or telephone administration options to complete the survey. Panel survey literature have reported on the impact of certain aspects of data collection strategies although panel vendors may utilize different approaches and so the incorporation of enhanced methodologies across various panels has not been straightforward to evaluate. In the Rapid Surveys System Rounds 1 and 2, two panel vendors (NORC and Ipsos) applied two different data collection methods to a split sample to study the impact of enhanced data collection methodologies. More specifically, NORC used two methods: Method 1 selected a sample from the AmeriSpeak Panel using their standard data collection protocol and Method 2 selected a sample from the AmeriSpeak Federal Panel using an adjusted protocol with higher nonresponse follow up during the recruitment stage and additional survey reminders. Ipsos' Method 1 used their standard web interviewing procedure while Method 2 used telephone prompting and telephone interviewing for nonrespondents. In this study, the two collection methods are compared by analyzing health

and demographic estimates for differences and by comparing survey completion rates for demographical domains.

Testing Proximity of Panel Survey Estimates to those Produced from a Reference Survey

Van Parsons, *National Center for Health Statistics*
Yulei He, *National Center for Health Statistics*

Commercial web-based panel surveys can be used to complement existing large scale government surveys to provide health information about U.S. populations. While the panel surveys are considered of adequate quality for many purposes, their use as part of government official statistics requires some assessment of the panel-produced estimates with comparable estimates produced from a high-quality government reference (benchmark) survey. These assessments are typically based on several benchmark variables that are common to both panel surveys and the benchmark survey. By assumption, the reference survey estimators will be unbiased for a targeted population, but the panel estimates are allowed some acceptable level of bias. In such a circumstance, panel estimates, individually or combined, can be released with some assurance of being "representative". We consider some composite null hypothesis testing methods for assessing acceptable bias-levels of the panel benchmark estimates. We apply this method to the National Center for Health Statistics (NCHS) Rapid Surveys System using the NCHS National Health Interview Survey as the reference survey.

Bias Evaluation for Web Health Surveys, A Sensitivity Analysis Approach

Yulei He, *National Center for Health Statistics*
Yan Li, *University of Maryland*
Katherine Irimata, *National Center for Health Statistics*
Guangyu Zhang, *National Center for Health Statistics*

To improve the timeliness of data products, survey researchers and practitioners have increasingly used web surveys to collect information for population health research and dissemination. From the statistical inferential perspective, these data are often referred to as nonprobability samples due to the lack of a well-defined probability sampling structure, or they come from probability panel surveys yet are subject to high nonresponse. Certain statistical adjustments are therefore needed to make proper inferences using web surveys. With a high-quality reference probability survey available, one popular adjustment approach is to create pseudo weights that properly "weight" the web survey samples back to the target population behind the reference survey. Li et al. (2022) showed that it is crucial to include variables that are both related to the outcome of interest and the sample selection into web surveys ("confounders" in epidemiology) in the weighting adjustment. In practice, however, it can be challenging to ensure that all important confounders are included in the data collection and subsequent weighting adjustment. Therefore, a plausible strategy for evaluation is to conduct a sensitivity analysis based on the web survey estimates when certain confounders are excluded. We illustrate this idea using some theoretical arguments, a simulation study, and an application to real data, utilizing public-use data from the National Center for Health Statistics' Research and Development Survey and National Health Interview Survey.

Return to Schedule Overview

Thursday, October 24, 2024 3:30 PM

Session K-4: The Future is Now: How the U.S. Statistical System is Enabling Evidence Building to Inform Policy

Organizer: John Finamore, *National Center for Science and Engineering Statistics*
Chair: Dr. Karin Orvis*, Office of Management and Budget*
Vessey 1

Fundamental Responsibilities of Recognized Statistical Agencies and Units
Kerrie Leslie,

TBD

Expanding Secure Access to CIPSEA Data Assets
Spiro Stefanou, *Economic Research Service*

TBD

The National Secure Data Service
Emilda Rivers,

TBD

Presumption of Data Access for Statistical Agencies and Units
Katy Rother,

TBD

Return to Schedule Overview

Thursday, October 24, 2024 3:30 PM

Session K-5: Optimizing Survey Methodologies for Enhanced Data Quality

Organizer: Barbara Downs, *U.S. Census Bureau*
Chair: Barbara Downs*, U.S. Census Bureau*
Vessey 2

Analyzing the Impact of Survey Cost Reduction Measures on Data Quality in the Consumer Expenditure Interview Survey
Graham Jones, *Bureau of Labor Statistics*
Tucker Miller, *Bureau of Labor Statistics*

The Consumer Expenditure (CE) Survey is a nationwide household survey program sponsored by the U.S. Bureau of Labor Statistics (BLS) and fielded by the U.S. Census Bureau. To facilitate survey fielding, the BLS contracts with the Census Bureau to collect CE Survey data and deliver it back to the BLS for processing and publication. In August and September of 2022, the BLS and Census implemented survey cost-reduction measures that restricted interviewers from returning to households for Wave 4 interviews that were previously incomplete (Type-A) interviews. As a result of these measures, there was a rise in Interview Survey non-contact and non-response rates during the third quarter of 2022. While these measures clearly impacted survey response rates, it is less clear whether they affected survey reporting quality during this period. The

purpose of this analysis will be to identify whether differences in sample composition or survey reporting quality exist between the cost-reduction period and those months immediately before and after. This will be accomplished by comparing components of Wave 4 reporting quality (in-person interviews, information booklet use, record use, length of interview, etc.) between periods, as well as the demographic composition of the samples. Identifying the impacts of these policies is important for informing future CE Survey fielding decisions made by the BLS.

Call Me Later, Maybe: Collecting Contact Information and Permission to E-mail, Mail, Call, or Text Survey Respondents

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

For longitudinal studies, collecting and maintaining contact information from respondents is critical for administering follow-up surveys. When these details may change before subsequent communication, for example, a sample of postsecondary students is likely to change their mailing address or school e-mail address when they graduate, and employees are likely to change their work contact information if they leave their job, it is particularly critical to maintain additional types of contact information.

The U.S. Department of Education's National Teacher and Principal Survey (NTPS) samples schools, and then principals and teachers within those schools. While we first contact staff through the school at which they work, we ask for their personal contact information (personal e-mail address, cellphone number, home mailing address) in order to be able to re-contact principals and teachers who may leave their job.

In our 2020-21 collection, we asked school staff to affirmatively check a box and opt in to receive text messages for follow-up purposes, while in our 2023-24 collection we asked respondents to check a box to opt out. We analyze respondent demographic characteristics, questionnaire layout, and the mode in which they completed a survey (web or paper) to determine whether these impacted the types of contact information they provided and whether they agreed to communicate through text messaging.

Evaluating the Accuracy and Efficacy of the Income and Program Screeners in the 2022 Survey of Income and Program Participation

Katy Giefer, *U.S. Census Bureau*
Adrianne R Brown, *U.S. Census Bureau*

The Survey of Income and Program Participation (SIPP) is noted for its lengthy and detailed data collection. While some questions ask about the interview month or the reference year, many of the survey questions are designed to produce monthly-level information. The result is a nationally-representative, longitudinal data set that provides comprehensive information about the income and programs people use to cover expenses, as well as other topics including household composition and employment. To reduce respondent burden, the SIPP relies on income and program screener questions that ask about annual income, monthly income, and program participation during the reference year. High-income respondents are "screened" out of questions about programs that provide support to low-income households. Evaluation of the income and program screeners is approached from two perspectives. First, we evaluate how accurate responses to the income screeners are compared to edited income. Second, we examine how many low-income people were asked the means-tested program questions based on their responses to the income and program screener questions. Using data from the 2022 SIPP, our results offer a descriptive overview of how well the income and program screener questions are performing. These findings provide the groundwork for evidence-based decision making for improvements to the screener questions.

Measuring Uncertainty: Bunching and Interviewer Effect in Probabilistic Expectations

Luis Guirola, *Bank of Spain*
Laura Crespo, *Bank of Spain*
Carlos Gento, *Bank of Spain*
Ernesto Villanueva, *Bank of Spain*

This paper considers the measurement of uncertainty via probabilistic expectations items, where respondents

assign probabilities to alternative scenarios. We look at three different surveys studying the same population, but which vary in their implementatio and mode of interview. In all three surveys, a significant number of respondents allocate all probability points to a single scenario, a phenomenon referred to as "bunching." We then explore whether this behavior stems from genuine certainty about the scenario or is a result of response effects, such as the task's complexity or respondents' reluctance to engage fully.

The analysis shows that the prevalence of bunching differs based on the mode of the survey—whether it is conducted online or through in-person interviews—and is inversely related to the respondents' financial literacy. Furthermore, for in-person interviews, we find evidence suggesting that interviewers can significantly influence bunching: we document substantial interviewer effects (ICC = 17%), and that its incidence was markedly reduced following an intervention that instructed them how to properly ask the question.

These findings highlight the impact of response effects on the accurate measurement of uncertainty, suggesting that both the mode of data collection and the approach of the interviewer play critical roles in influencing respondents' behavior and the resulting data quality. This underscores the need for careful consideration in survey design and administration to mitigate response effects and improve the measurement of probabilistic expectations.

Strategies to Maximize Batch Locating Service Data in Longitudinal Studies using Survey of Doctorate Recipients

Christopher Wong, *NORC at the University of Chicago*
Flora Lan, *National Center for Science and Engineering Statistics*

Batch locating services searches can be a powerful tool for obtaining up-to-date contact information to use for recruitment of individuals in a list sample. An advantage of these searches is they provide immediate results and can be highly cost effective when compared to more manual locating techniques. They can be particularly well-suited for longitudinal studies that have robust identifiers and historical information to use for matching. However, a challenge for data collection managers can be developing standards for determining the record-by-record usability of the returned information, as the precision of the vendor-provided returns may not always be fit for efficient recruiting. This presentation reviews techniques used on the 2023 Survey of Doctorate Recipients (SDR) for assessing the quality of mailing addresses, phone numbers, and email addresses provided from batch searches. The SDR collects information from individuals with a U.S. research doctoral degree in a science, engineering, or health field. It is sponsored by the National Center for Science and Engineering Statistics within the U.S. National Science Foundation and by the National Institutes of Health. These review techniques fit within a two-dimensional framework for assessing quality where one dimension is based on accuracy (i.e., is the returned contact information associated with the targeted sample person) and the other is based on recency (is the returned contact information the most up to date). The presentation includes strategies for developing record-level quality indicators that can be used to drive data collection management decisions based on the batch search results.

Return to Schedule Overview

Session K-6: Shifting into Safer Gear: Data-Driven Strategies in Transportation

Organizer: Stuart Gluck, *Federal Railroad Administration*
Chair: Stuart Gluck, *Federal Railroad Administration*
Discussant: Janine McFadden, *Federal Railroad Administration*
Room 0105

How to Prioritize Locations for Railroad Safety Inspection Programs

Young-Jun Kweon, *Bureau of Transportation Statistics*
Jianqiang Ye, *Federal Railroad Administration*
Ruby Li, *Federal Railroad Administration*

The Federal Railroad Administration (FRA) is responsible for ensuring safety compliance of railroads in the US and performs ongoing safety inspections across the country. FRA has established a system of prioritizing locations for safety inspection utilizing a risk-based approach including incident predictive models, severity scoring, and stakeholder engagement. There are six technical disciplines established for safety inspection programs at FRA and each discipline is designed to address unique aspects of railroad safety compliance and enforcement such as Operating Practices and Motive Power & Equipment (MP&E). Each discipline involves a different group of stakeholders and different data elements. Thus, risk modeling is conducted separately for each discipline. FRA is developing the second generation of risk models with enhanced data. For the second generation, a 7-step process was established for ensuring consistency in model development across the six disciplines and decision charts are used to guide selection of an appropriate model. The 7-step process has been proven to be effective. For MP&E, a zero-inflated negative binomial was selected as the final model with the assistance of the decision charts. Unlike the first version, live data feeds updated on a regular basis enter the second generation of the risk models for prioritizing inspection locations. The presentation will provide a brief overview of the history and background of risk modeling at FRA, the explanation of the 7-step process and the decision charts, the results of the MP&E risk model and risk score, and prioritization based on the developed model.

Improvements in Data Collection and Linkage for Risk Modeling in Railroad Safety Inspection Programs

Andrew LaBounty, *Federal Railroad Administration*
Patrick Johnson, *Federal Railroad Administration*
Ruby Li, *Federal Railroad Administration*
Tony Ye, *Federal Railroad Administration*

Federal Railroad Administration (FRA) promotes and regulates safety throughout the nation's railroad industry. In keeping with FRA's mission of enhancing safety, the agency utilizes risk models based on an inventory of geographic inspection points (locations) to allow the agency to identify and address areas of higher risks. Custom risk models were developed for the major industry disciplines including, Track, Motive Power and Equipment, Operating Practices, and etc. Each discipline's specific needs are addressed by focusing on the relevant data for the discipline and analysis of relevant criteria. In the development of the second generation of risk models, the data collection and linkage between disparate datasets were enhanced and improved by leveraging geospatial analysis to create spatial joins and determine areas of influence using Thiessen polygons. Unlike first-generation risk models, which utilized static scoring methodologies, the latest second-generation risk models rely on live data connection feeds which regularly update the resulting risk scores. Live data connections are facilitated through the use a data analytics pipeline, developed by the FRA, which leverages a data lake to routinely ingest and transform data that is subsequently fed into the second-generation risk models to produce up-to-date risk scores for each discipline. The visualization of up-to-date risk scores in geospatial web applications is complimented by individual inspector priorities which are entered directly into the mapping applications. The presentation will provide an overview of FRA's risk modeling practices, data

pipeline development, and discuss the improvements in data collection and linkage in the second generation of risk modeling.

Linking Laboratory Vehicle Test Data and Police Crash Report Data for Evaluating a Vehicle Safety Regulation

Young-Jun Kweon, *Bureau of Transportation Statistics*

According to 2014-2018 national statistics, only 2% of vehicles in all traffic crashes were involved in rollover crashes, but 24% of all fatalities resulted from rollovers. At a rollover, a stronger vehicle roof would prevent occupants from being severely injured. The need for a stronger roof led to creation of the vehicle safety standard for roof crush resistance. This study was to provide a basis for assessing effectiveness of the most recent standard. While developing analysis data, the study encountered several issues. First, roof strength data came from laboratory tests and injury data came from police reports and there was no information linking them. To link the disparate data, the study proposed two approaches based on vehicle identification numbers. Second, traffic crash reporting practices vary across the states. Since each state has its own unique crash report form, definitions of data elements are not consistent across states. To establish consistent definitions, the study devised and tested various combinations of data elements. Based on the data linkage and the consistent definition, the study successfully formed the analysis data by marrying 358 cases of laboratory tests and about 800,000 cases of rollover traffic crashes. The study developed three binary logit models including overdispersed and Bayesian models and calculated probabilities of being severely injured varying by roof strength. This presentation explains how the study created reliable data linkage and consistent rollover definitions and discusses the final model results and the safety effect of the most recent vehicle safety regulation.

Risk Modeling for Railroad Incidents: Variable Screening, Model Type Selection, Functional Form Suggestion, and Model Development

Young-Jun Kweon, *Bureau of Transportation Statistics*
Jianqiang Ye, *Federal Railroad Administration*
Ruby Li, *Federal Railroad Administration*

Risk modeling is a process of assessing and quantifying potential risks under given circumstances and it finds application in various fields. It typically involves constructing a mathematical representation of a relationship between a variable reflecting the risk and variables characterizing the circumstances. In the area of railroad safety inspection, risk modeling is found to be useful in prioritizing locations so that limited human resources can be utilized effectively. The Federal Railroad Administration (FRA) audits, inspects, and assesses conditions of railroads to determine if a railroad complies with federal safety regulations. FRA has incorporated risk modeling into its safety inspection programs. This study was to develop a risk model for the safety inspection discipline of Motive Power and Equipment (MP&E) of railroads relating the number of MP&E-related train incidents to potential predictor variables coming from nine cross-agency data sources. The study developed 7-step process for developing models consistently across six inspection disciplines and decision charts for selecting appropriate models for analysis. The process involves screening variables, determining an appropriate model type, selecting variables and suggesting functional forms, and selecting a specific model. The process led to two model types being appropriate, zero-inflated and zero-hurdle models accounting for extra variability due to preponderous zeros. This presentation will illustrate the 7-step process using MP&E data from variable screening to final model development and discuss the estimated final model and its implementation in FRA's safety inspection program. The developed process and decision charts are applicable in any field where regression analysis is appropriate.

Return to Schedule Overview