FedCASIC 2018

April 17-18, 2018
Suitland, MD

Federal Computer Assisted Survey Information Collection Workshops

Preliminary Program (as of 04/02/18)

Sponsored by
The U.S. Census Bureau and
The Bureau of Labor Statistics

Tuesday, April 17, 2018

Arrival and Registration: 8:00 am – 9:00 am
Pre-Function Area

Plenary Session: 9:00 am – 10:20 am
Keynote address -- Start at the End: Why dissemination should be a forethought
Room: Auditorium
Keynote Speaker: Zach Whitman, U.S. Census Bureau Chief Data Officer

Chief Data Officer Zach Whitman, Ph.D. joined the Census Bureau in July 2016 to promote use of Census data in public and commercial sectors. His prior experience at the Departments of Housing and Urban Development, Homeland Security, and Transportation included civic data engagement events for federal agencies and public-private collaborations. Whitman has spearheaded hackathon events with organizations like Zillow, Uber, and the Amazon Alexa team. He was also a founding team member of the Datausa.io project with MIT, an online experience that showcases open data from the government through visualizations and storytelling. Through his interactions with the public, he has discovered the need to release data in more efficient ways and in broader consumable methods

Session 1A: 10:45 am – 11:45 am
Designing for Users
Room: Auditorium 1 & 2
Chair: Sarah Lessem, *National Center for Health Statistics*

Conducting Usability Testing on Complex Systems
Jean E. Fox, *Bureau of Labor Statistics*
Robin Kaplan, *Bureau of Labor Statistics*

A usability test administrator always tries to learn as much possible about the domain, to better understand the users and their tasks. Projects at statistical agencies pose a particular challenge because the systems are often designed for highly educated, highly trained, and highly experienced users, who are performing fairly complex tasks. In some cases, the person conducting the usability test is involved throughout development, with the time to learn about the users and their tasks, and involvement in the design of the system's features. However, in many cases, the request for assistance comes at the end of development, with little time to learn much about the users, their tasks, or the system. It may also be difficult to obtain sufficient training in the topic to fully appreciate the users' knowledge. Given the complexity of these systems, it is especially important to incorporate usability testing into the development process. However, it may be necessary to adapt traditional usability testing methods to accommodate these conditions. The co-authors of this presentation recently conducted several usability tests of complex systems used by highly skilled economists, statisticians, and researchers. Although we were already experienced in conducting usability tests, we found that some of the common strategies wouldn't work in these situations. We tried different approaches, some of which worked better than others. In this presentation, we will share our experiences and lessons learned with these usability tests.

Session 1A: 10:45 am – 11:45 am
Designing for Users
Room: Auditorium 1 & 2

We will address a number of topics, including:

- Learning the system - How much do you really need to know about how the system works? What are good ways to learn what you need to know quickly?
- Picking tasks - What should you consider as you select tasks for the test? What are some strategies for identifying appropriate tasks?
- Training participants - How much training should you provide to the participants, who might receive several weeks of training before using the final product? What are good approaches to providing the training?
- Staffing the tests - Who should observe the tests? How should they be involved in providing help to the participants?
- Logistics - Since these systems often depend on a complex database, what should you consider regarding the setup and management of data for each participant?
- Recording data - How do you know when the participants are experiencing problems? How can you identify solutions to these problems? How do you know if the participant completes the task successfully?

Creating More Meaningful and Explanatory Visualizations
Christopher Griggs, *RTI International*
Roger Jesrani, *RTI International*
Jorgen Waldermo, *RTI International*
Rebecca Watkins, *RTI International*
Bharathi Jayanthi Golla *RTI International*

Creating meaningful visualizations that accurately represent study data has vast benefits for project teams and managers. For SPARS, the new online data entry, reporting, and training system for SAMHSA, user experience (UX) design goals were created to directly address the end user's behavior when interacting with visualizations. This yielded a suite of high performing, interactive, and intuitive visualizations. In this presentation, we will detail the UX paradigms and strategies used to encourage user interaction, reduce friction, and lead users to meaningful actions within the application. Learn what technology infrastructure is needed and the programming stacks used to develop these visual interactions.

Session 1A: 10:45 am – 11:45 am
Designing for Users
Room: Auditorium 1 & 2

Data Visualization Testing
Bharathi Jayanthi Golla, *RTI International*
Imrul Hasan, *RTI International*
Sangeetha Immani, *RTI International*
Al-Nisa Berry, *RTI International*
Anwar Mohammed, *RTI International*

Data Visualization is anything that converts data sources into a visual representation (like charts, graphs, maps, sometimes even just tables). Data visualization tools are a great way to create impactful reports. A well-designed report can give users an understanding of their data quickly and easily. However, data visualization testing needs specific testing approaches and strategies when compared to standard reports testing or testing web based applications. This presentation will address suggested processes on the type of testing and skills required for data visualization testing by answering the following questions:

- How is testing data visualization different from traditional web-based applications?
- Does data visualization testing require specialized SQA skills and techniques?
- What are some of the challenges?

We answer these questions by analyzing the current trends in data visualization reports development and testing, and by proposing our views on the topic based on lessons learnt testing data visualization.

Session 1B: 10:45 am – 11:45 am
Data Security
Room: Auditorium 3
Chair: Tamara Adams, *U.S. Census Bureau*

The GDPR is Coming and it is Coming Fast...
Peter Milla, *Peter Milla Consulting*

Compliance with the EU General Data Protection Regulation (GDPR) applies to any organization that collects personal data in the EU. The GDPR significantly reshapes the data protection landscape for organizations that collect and process the data of European citizens. The Regulation grants extended rights to data subjects, allows data subjects to bring legal action against organizations in case of data breach and also imposes fines up to 4% of annual global turnover or €20 million (whichever is higher). With the EU-US Privacy Shield program, enforcement will also be carried out by the US Federal Trade Commission. Many US organizations are uncertain about what the Regulation's requirements are and what their obligations might be. The presentation will address a number of essential areas, including:

- The requirements of the regulation
- How it applies to organizations that collect personal data from EU citizens
- Essential requirements for compliance
- Specific guidance for companies based or operating in the US.

Session 1B: 10:45 am – 11:45 am
Data Security
Room: Auditorium 3

The presentation will also provide easy to understand guidance on:

- The rights of data subjects
- The data transfers outside the European Union
- "Data Controllers" and "Data Processors"
- The operational, strategic, regulatory, statutory and contractual aspects of risk management -- including subcontracting
- Data privacy audits and impact assessments
- The importance of Information Security compliance
- Policies and procedures
- GDPR awareness and training
- Data Protection Officers
- Data breach reporting

Implications of Data Integrity & Security Standards for Managing Web Survey Fieldwork
Carl Ramirez, *U.S. Government Accountability Office*

Self-administered web survey systems hold and transmit sensitive sample and response data, and require some level of security and access control to questionnaires, to not only avoid disclosure of individual respondent PII and business proprietary data, but also to ensure data integrity and quality throughout the survey lifecycle. But the spectrum of access control methods - with varying restrictiveness on respondent access to questionnaires, and on access-facilitating management practices - has implications for respondent burden and cooperation with fieldwork methods. For example, successful fieldwork activity often involves two-way communication (especially of informants at key organizations with their survey "case managers") between sample members and help-desk or other survey management staff. This presentation starts with this example: how do access restrictions, namely restrictions imposed on survey staff in having access to and communicating passwords to original or proxy respondents, impact support calls or telephone followup efforts. The presentation also briefly describes information system access policies and practices, and examples of the various ways they have been implemented in different survey systems in today's statistical system. A case is also made for reexamining traditional approaches to survey database access security in the context of the "protective envelope" that includes authorized researcher activity and access to data.

Exploring the Feasibility of Blockchain Technology for Medicare Beneficiaries
Laxminarayana Ganapathi, *RTI International*

Privacy as well as integrity of the health records have been a concern for a long time. Recent advances in Blockchain technology and distributed ledger frameworks have pointed to new possibilities and solutions to maintain the confidentiality as well as integrity of the transactions involving health records. As promising as they seem, the challenges of these new technologies are significant both in terms of resource requirement and implementation difficulties considering the legacy systems being used across

Session 1B: 10:45 am – 11:45 am
Data Security
Room: Auditorium 3

multiple stake holders. Centers for Medicare and Medicaid services is a major player housing the claims data of more than 50 million that includes nearly all senior population and those with chronic conditions in the USA. Continuous progress over the years has made this vast data very structured and available for research. We intend to review and explore the feasibility of using Blockchain technology to authenticate the various transactions. There are several stake holders in the multiple transactions around a beneficiary. Even so, there could be opportunities for a centralized ledger to do most of the authentications. An efficient implementation could certainly lead to drastic reduction in fraudulent claims besides other benefits of the blockchain technology.

Session 1C: 10:45 am – 11:45 am
Sample Development
Room: Conference Rooms 3 & 4
Chair: John Boyle, *ICF International*

Comparing Two RDS Approaches to Extend the Reach of a Probability-Based Panel
Becky Reimer, *NORC at the University of Chicago*
Vicki Pineau, *NORC at the University of Chicago*
Stuart Michaels, *NORC at the University of Chicago*
Rosalind Koff, *NORC at the University of Chicago*
Stephanie Jwo, *NORC at the University of Chicago*
J. Michael Dennis, *NORC at the University of Chicago*

Surveying rare populations such as LGBT individuals is often of great interest for policymakers, but can be both costly and logistically difficult to execute using a probability sampling approach due to low incidence. Given these challenges, studies of such populations are often forced to use less rigorous samples methods like opt-in web panel samples. This paper presents additional pilot research conducted by NORC to explore a cost-effective alternative method for reaching rare populations that begins with a probability-based panel sample and incorporates respondent driven sampling (RDS) to augment the sample with individuals outside of the panel. The originating sample source for this pilot work is NORC's AmeriSpeak Panel®, which uses the probability-based NORC National Frame to construct an address-based nationally representative sample panel with sample coverage of approximately 97% of US households. The subpopulation of interest for this pilot is the lesbian, gay, bisexual, and transgender communities. In the pilot, we surveyed panelists who had previously self-identified as either belonging to one of these groups, or as non-LGBT, and then asked these cooperative respondents to enlist LGBT friends and family members to complete the same survey. This work builds on a pilot study conducted by NORC in early 2017 with altered survey engagement language/materials and also with an experiment to compare the effectiveness of two different RDS methods: 1) asking panelists to refer others directly by providing their contact information to researchers at the end of the survey, vs. 2) asking panelists to recruit others by providing them with survey links to distribute when and how they saw fit after completing the survey. Best practices for addressing concerns such as respondent privacy and

Session 1C: 10:45 am – 11:45 am
Sample Development
Room: Conference Rooms 3 & 4

maintaining the control to send survey reminders will be discussed, along with findings regarding the effectiveness of these two approaches for recruiting multiple rounds of survey referrals.

Rebuilding the Frame Post-Disaster – Assessing the Impact of Infrastructure Damage to Landlines in Houston
Thomas Brassell, *ICF International*

The construction of representative sample frames for telephone surveys in post-disaster areas is a known challenge given infrastructure damage that can impact landline accessibility (Kessler et al., 2008). Phone number validation services may be able to provide researchers studying post-disaster zones with an opportunity to more efficiently assess areas where damage exists to the landline infrastructure. Our study explores the application of phone number validation services as a means of assessing landline infrastructure damage. Hurricane Harvey interrupted the Health of Houston Survey, a telephone survey conducted via a dual-frame RDD sample. Prior to interruption, 65% of the landline sample selected from a 1+ working bank RDD frame were prescreened as nonworking. When the survey resumes in February, we will conduct a number of assessments. First, we will rerun numbers previously determined as working through validation services and dialer tests to determine whether working status has changed post-disaster. Second, we will randomly sample new landline numbers in the area, and determine working status through real-time validation services and dialer calls. We will analyze whether there are geographic clusters of non-working numbers, as based on the geographic area associated with the telephone number, these clusters may suggest areas still in recovery from the disaster. We will compare to damage maps of the area and discuss the utility of such maps in survey research of post-disaster areas. Geographic differences in the pre-screened nonworking rates may indicate areas still in recovery from the disaster, but may also reflect the choice to not renew landline service. In the cell sample prior to the hurricane, 72% of respondents stated they were cell-only. We will compare this rate to the post-hurricane survey. The results will help to better inform future research of potential alternative approaches to sample design post-disaster.

Using Westat's Virtual Listing System to Replace Listing in the Field
Jane Shepherd, *Westat*
Michelle Amsbary, *Westat*
Mike Giangrande, *Westat*
Brad Edwards, *Westat*

The 2018 Commercial Buildings Energy Consumption Survey (CBECS) is a national sample survey conducted by the U.S. Energy Information Administration (EIA) to collect information on the stock of U.S. commercial buildings. It provides the only national source of statistical information about the energy-related characteristics of commercial buildings. In the past, constructing the sampling frame has been a timely and costly task involving a full field listing operation. For the 2018 CBECS, Westat is exploring the implementation of virtual listing (without sending data collection staff to the field) with processes

Session 1C: 10:45 am – 11:45 am
Sample Development
Room: Conference Rooms 3 & 4

developed by Westat's Geospatial Services Group. The Virtual Listing System (VLS) is a fully integrated, custom, web-based system of observing, cataloging, and documenting CBECS eligible commercial buildings in the sampled segments.

Posters and Demonstrations: 12:45 pm – 1:15 pm
Pre-Function Area

The Advantage of Using Auto-Dialers with a Cell-phone Sample
Patricia E. Vanderwolf, *ICF International*
Deirdre Middleton, *ICF International*
Randy Zuwallack, *ICF International*

The Telephone Consumer Protection Act (TCPA) restricts telephone solicitations and limits the use of automated telephone equipment. The TCPA applies to autodialed, prerecorded or artificial voice telephone calls, including text messages, made by the government and government contractors, as well as non-government organizations/businesses. In 2016, the Federal Communications Commission (FCC) ruled that the TCPA did not apply to telephone calls made by or on behalf of the federal government in the conduct of official government business, except when a call made by a contractor does not comply with the government's instructions. Prior to this ruling, telephone calls to potential respondents for government surveys were required to be manually dialed. In 2017 ICF conducted a monthly RDD CATI survey of working adults for the U.S. Bureau of Labor Statistics. As part of this survey we conducted a randomized experiment over three months testing the use of an autodialer with the cell-phone sample, the cell TCPA (non-dialer, required prior to the new ruling), and a landline dialer sample. The test demonstrated that using the cell-phone autodialer with settings customized for the project improves the production rate and dial rate, and decreases the interview length without decreasing the response rate.

R Shiny Data Collection Dashboard
Nathan Lotze, *Bureau of Labor Statistics*
Bryan Beverly, *Bureau of Labor Statistics*

We created a user-friendly tool for the BLS Current Employment Statistics (CES) program to measure real-time survey data collection performance. We used R Shiny software to build this interactive dashboard that queries the Oracle databases used to store our CATI, web, and Electronic Data Interchange (EDI) survey data. In order to generate easy-to-understand visualizations, this dashboard gives users the ability to customize: (1) the selection of the data being queried and (2) how the results are displayed. We will demonstrate how survey managers can visualize collection performance of their survey in real-time.

Posters and Demonstrations: 12:45 pm – 1:15 pm

Pre-Function Area

Efficiencies in Multimode and Multidevice Management with Blaise 5

Kathleen O'Reagan, *Westat*
Nikki Brown, *Westat*
Rakesh Kudupudi, *Westat*
Richard Frey, *Westat*

Most survey research projects today involve multimode data collection options including respondent and interviewer mediated modes. Blaise 5 enables the developer to program a survey instrument once and then easily deploy across multiple modes for efficient data collection operations. The demonstration will highlight the uses of Blaise 5 for in-person and self-administered data collection modes using various devices, and show how these efficiencies enable researchers to easily adapt the study design throughout the data collection period. In addition we will demonstrate the integration of the Blaise 5 instrument with Westat's multi-mode management system, M3. M3 is a platform for establishing a robust integrated systems environment for conducting multi-mode surveys with flexibility and efficiency.

Agile testing and the role of the Agile tester

Bharathi Jayanthi Golla, *RTI International*
Sangeetha Immani, *RTI International*
Imrul Hassan, *RTI International*
Al-Nisa Berry, *RTI International*
Anwar Mohammed, *RTI International*

Agile methodology is a practice that promotes continuous iteration of development and testing through the software development lifecycle of a given project. Both development and testing activities are concurrent, unlike other methodologies which often ends up being expensive.
Agile testing involves testing from the customer perspective as early as possible, and testing early and often as code becomes available and stable enough from module/unit level testing. With other traditional testing models, the amount of testing grows exponentially as the product grows and Quality Assurance invariably struggles to keep up. This presentation provides an overview of the role of testers on agile teams. How testers can add more value on agile teams by contributing earlier and moving from defect detection to defect prevention.

Leveraging a 185 million record master location data source on mobile devices for household surveys

Greg Clary, *Mi-Corporation*

A hardware-independent, mobile survey software platform is integrated with a master location data source for US household surveys. The mobile survey platform enables data collection across iOS, Android, and Windows. Data replication software allows interviewers to access existing master location data such as household locations and survey meta-data while offline. The data replication software applies filters so that only the relevant part of the master location data is provided to an interviewer's mobile device. All data collection and data lookups can occur while the user is disconnected, and survey data is uploaded

Posters and Demonstrations: 12:45 pm – 1:15 pm

Pre-Function Area

upon network availability. The master location data source has more than 185 million property records. It is optimized to provide the address and geolocation of every household in the United States based on multiple sources. Each household is given a unique identifier. The master location data source stores meta-data information pertaining to a survey, indexed by unique identifier. Together, the mobile survey platform and the master location data source provide interviewers immediate, offline confirmation of survey design parameters based on location, information about when to offer incentives, and verification of interviewer location.

Cleaning, Processing, and Archiving Administrative Records

Juan Salazar, *MarkLogic*

The use of administrative records and third-party data is one of the four Key Innovation Areas that the 2020 Census has designed, in order to reduce the overall cost of the program, by leveraging information that the population has already provided the government, or that can be mined from other sources (e.g. Fannie/Freddie, Zillow, etc.)

One of the biggest challenges to executing on this promise, is that each of these data providers (e.g. other Federal Agencies, State and Local Governments, real estate companies/websites) will likely offer their own data in whatever format it already is in – so not only is the data in silos, but when integrating it, it will be a heavy-manual process to clean the data before it can even be processed.

The main reason why it's so labor-intensive and time-consuming to integrate data in different formats is because most of the world is still using forty-year old technology – relational database management systems (RDBMS or "relational database".) With relational database technology it becomes necessary to define the full schema (i.e. tables, rows/columns) up-front before the data can be imported into a single repository – so each and every "column" needs to be matched up with the target schema – typically through an Extract, Transform, and Load (ETL) process.

But most of this work is not required to fulfill the mission of this Key Innovation Area. For example, if the Bureau gathers an entire tax return for a household; what information is really needed? Possibly, out of dozens or hundreds of data fields (e.g. first name, last name, address, dependents, deductions, etc.) enumeration probably only requires a few – so why spend so much time normalizing every single data element when only the needed ones can be harmonized, as you go, in an agile manner?

With modern databases – document / multi-model databases, to be exact – the data can be ingested as-is and only the fields needed for the functions required can be quickly harmonized and exposed via APIs. This demo will show this process of data harmonization, live, in action.

Moreover, administrative records typically include PII and information protected by Title 13 and Title 26. We can also entertain questions as to how this data can be protected based on roles, down to the row-level, even to the point that a specific element can be redacted so that a user without rights won't even know that it's part of the dataset. These security capabilities are important during processing, consumption, and archiving – the full lifecycle of using administrative records. As a note, this same approach can be utilized for other surveys.

Session 2A: 1:15pm – 2:30pm
Perceptions of Confidentiality
Room Auditorium 1 & 2
Chair: Alisu Schoua-Glusberg

Understanding of Electronic Security Statements and Its Effects on Respondent Cooperation
Alisu Schoua-Glusberg, *Research Support Services*
Patricia L. Goerman, *U.S. Census Bureau*

The Census Bureau has conducted research designed to investigate respondent perceptions of privacy and confidentiality messages for years. One type of such messages are those that respondents often see when first accessing a government computer system to complete a survey online. Such messages are intended to provide the respondent assurances of confidentiality and privacy, as well as any limitations. However, because of complex legal language these messages sometimes include, they may not be fully clear or reassuring to respondents.
This presentation will examine respondent reactions to a security statement tested in focus groups in 2015 with English speakers as well as with limited English proficient (LEP) immigrants speaking Spanish, Russian, Chinese, Vietnamese, and Korean. These focus groups were part of a comprehensive Census Bureau study that included cognitive and usability testing of paper and electronic Census forms in multiple languages.
Among other focus group activities, participants were shown a printout of the first screen they would see after reaching a webpage to access the census form. The purpose of the message was to let respondents know that they had reached a government computer system, that they should only use this system to complete the census form, that any information they entered could be used for statistical purposes, that their information would be protected under privacy laws, that by using that computer system they would be authorizing the Census Bureau to use their information, and that using that computer system for unauthorized purposes would be against the law.
Focus group participants were asked to read it and to mention anything they particularly liked or disliked in its content. In the discussion that ensued, participants' understanding of the message was elicited. In the presentation we will focus on interpretation of the purpose of the message and on participants' reactions. Potential implications for participation in electronic data collection efforts will be discussed.

Filling in the Blanks: Public opinion about the use of administrative records
Casey S. Eggleston, *U.S. Census Bureau*
Jennifer Childs, *U.S. Census Bureau*
Gerson Morales, *U.S. Census Bureau*

If your household did not return a census form in 2020, do you feel it would or would not be an invasion of your privacy to use information from other government agencies to find out the names and ages of the people living at your address? An interdepartmental Privacy and Confidentiality team at the U.S. Census Bureau is considering questions like this one and engaging in ongoing monitoring of public perception of the use of administrative records in surveys and censuses. Since Fall 2012, the Census Bureau has fielded a variety of survey items in an ongoing nightly national telephone survey asking about sharing of government records between agencies and the use of such records in federal surveys. In this presentation, we summarize our findings about privacy and confidentiality concerns related to the Census

Session 2A: 1:15pm – 2:30pm
Perceptions of Confidentiality
Room Auditorium 1 & 2

Bureau's use of administrative records and discuss the practical and ethical considerations that arise when considering the use of administrative records in survey research.

Informed Consent in the Age of Cybersecurity: Testing messages and exploring perceptions
Alfred Tuttle, *U.S. Census Bureau*

The U.S. Census Bureau's mission is to serve as the leading source of quality data about our country's people and economy. The success of our programs depends on the public's perception of the Census Bureau as a credible and reliable protector of individuals' personal information. Our survey and census programs follow informed consent procedures as required by federal laws and regulations and our own policies. An important part of our mission is to communicate clearly to respondents how their data will be used and protected. Researchers from the Census Bureau conducted research to identify messages that clearly describe security procedures without alarming respondents unnecessarily. We conducted qualitative research via in-person interviews and online surveys to get feedback from members of the public on various messages related to privacy and confidentiality and identify strategies for crafting messages that are informative and reassuring. This presentation will summarize the results of the message testing. In addition, we will present relevant findings from focus groups we conducted previously with respondents and nonrespondents about their perceptions of privacy and security in the context of government statistical collections, and from pretesting research with questions about perceptions of information security risks.

Using Card-Sorting Strategies to Examine Persuasive and Confidentiality Messaging in Surveys
Aryn Hernandez, *U.S. Census Bureau*
Krysten Mesner, *U.S. Census Bureau*
Diane K. Willimack, *U.S. Census Bureau*

The U.S. Census Bureau's economic surveys rely on sample businesses' participation in order to produce accurate and reliable statistics about the status of the U.S. economy. While many of the Census Bureau's business surveys are mandatory, we prefer to persuade response rather than threaten legal consequences of nonresponse. This requires effective persuasive communication with businesses and their communities. This need for effective survey communication with businesses is heightened in an all-electronic collection environment. Without a paper questionnaire or interviewer intervention, letters become the primary means of making contact, both initially and for nonresponse follow-up. Research has demonstrated that business respondents pay limited attention to much of the content of the letters they receive from statistical organizations, usually focusing on login instructions, mandatory notices, and due dates. Nevertheless, U.S. Federal statistical agencies are legally required to provide respondents with specific information about response burden, confidentiality, privacy, and cybersecurity associated with survey participation. In addition, research indicates that survey participation among business respondents may be motivated by information about how their data are used. Census Bureau researchers recently conducted testing of various messages with business respondents. In this paper, we describe strategies

Session 2A: 1:15pm – 2:30pm
Perceptions of Confidentiality
Room Auditorium 1 & 2

and methodologies we used to test statements informing respondents about privacy policies and data use. Of particular interest is our use of a card-sorting strategy to gauge business respondents' knowledge of our use of their reported data, identify persuasive appeals that resonate, and, in the end, sequence proposed letter content to be most compelling from their perspective. Paired with traditional cognitive methods and statistical analysis of message rankings, the card-sort revealed useful, sometimes surprising, results.

Session 2B: 1:15 pm – 2:30 pm
Roundtable: Management challenges associated with new and emerging survey challenges and approaches
Room: Auditorium 3
Moderator: Karen M. Davis, *RTI International*
Panelists: Nicholas Johnson, *Bureau of Labor Statistics*
Eloise Parker, *U.S. Census Bureau*
Kyle Fennell, *NORC at the University of Chicago*
Gregg Peterson, *Survey Research Center at the University of Michigan*

This panel will discuss current challenges related to the changing nature of data collection, security, and/or use of administrative or extant data sources in surveys. Many projects today are primarily multi-mode and may have many variable components that require specialized programming customization and the integration of different devices and technologies. Many surveys utilize mobile devices, or mobile applications, or web-based surveys – this panel will cover ways that organizations are dealing with security and other challenges around using innovative approaches to obtaining survey data. The acquisition, testing and integration of devices and technologies, deployment considerations, accessibility, and other implementation challenges will be discussed. Panelists will discuss approaches used by their organizations and examples of how they have addressed these challenges.

Session 2C: 1:15 pm – 2:30 pm
Paradata
Room: Conference Rooms 3 & 4
Chair: Chris Stringer, *U.S. Census Bureau*

Standardizing Web Paradata Tools at the U.S. Census Bureau
Renee Ellis, *U.S. Census Bureau*
Joanna Fane Lineback, *U.S. Census Bureau*

The collection and analysis of paradata provides a powerful tool for monitoring and improving surveys. Paradata can be used to monitor the collection of data, as a tool for adaptive survey designs, or as a way to identify survey or question issues in order to improve them. Paradata are currently being used all

Session 2C: 1:15 pm – 2:30 pm
Paradata
Room: Conference Rooms 3 & 4

across the Census Bureau for these purposes. However, with respect to web paradata, the collection and processing has been customized by surveyor program area. This has resulted in varying methods and some duplication of effort to produce similar results. For example, web paradata are unstructured by nature. Across the bureau, many separate tools have been developed to create structured datasets for analysis. The Census Bureau Web Paradata Team was formed to address bureau-wide paradata issues, from policy issues that govern web paradata use, to defining common web paradata terms, to standardized programs for formatting the data. To allow for increased collaboration and facilitate knowledge sharing across the Census Bureau, a current priority is to develop a common set of metrics for analyzing the paradata to be used as a starting point for all managers and researchers. The purpose of this presentation is to share current work in progress on efforts to standardize web paradata tools where applicable across the Census Bureau, and obtain feedback on standardizing general definitions, metrics, and formulas.

Using Audit Trail Files for Missing Data Extraction in Wave 2 of the 2014 Survey of Income and Program Participation

Patrick C. Campanello, *U.S. Census Bureau*
Shelley K. Irving, *U.S. Census Bureau*

The reengineered Survey of Income and Program Participation (SIPP) instrument debuted in 2014 and included a redesign of the instrument from the ground up. New to the 2014 panel of SIPP were a move to annual interviewing and the introduction of an event history calendar (EHC). The EHC helps respondents recall information in a more natural 'autobiographical' manner by using life events as triggers to help respondents recall other economic events. Data are entered as spells with distinct beginning and ending months. Once a spell is reported, respondents are asked a set of detailed questions related to the topic at hand. We collect data for the reference period (i.e., the preceding calendar year) and for any months in the current year up to the interview month. While processing the Wave 1 data, we discovered that there were fields in the EHC with reported data whose values are missing in the output dataset. These values were missing despite having been entered into the CAPI instrument by the Field Representatives (FRs). Because these variables contained reported data, we did not want to lose this information and have to impute the missing values. To help fill in the gaps, we turned to Python programming to extract the reported data from audit trail files. Audit trail files contain a complete log of FRs' interactions with the SIPP instrument, including all responses entered into the instrument. The Python program allowed us to extract the reported data and attach it to the data file, rather than having to rely on imputation methods to fill in the missing values. As we began processing data from Wave 2 (collected in 2015, covering calendar year 2014), we originally followed the same approach for extracting missing data that we had used with Wave 1. While this approach was partially successful, we quickly encountered new and unexpected developments, particularly related to the presence of provisional spells. Provisional spells are spells that were ongoing at the time of interview in the previous wave's interview. We feed them back from the previous wave into the current wave, so that they show up in the current year's interview; this both helps with respondent recall and shortens the interview. FRs are expected to verify the dates that the provisional spell started and ended and enter (or verify) the detailed questions about the topic. For

Session 2C: 1:15 pm – 2:30 pm
Paradata
Room: Conference Rooms 3 & 4

variables with missing values, however, these provisional spells proved problematic, as the Wave 1 code was unable to handle complex FR activity related to adding or updating spells. These developments caused us to reassess how we approached audit trail data extraction. The initial step of this task required updating a portion of code used to determine which missing answers should be replaced with audit trail data. The new code also utilizes more data from each spell to improve its accuracy, as well as leveraging data from spells of the same or related topics. This not only helps to ensure the correct assignment of each spell's data, but also provides a means of directing changes to other related topics' spells. This presentation describes our Wave 2 missing data extraction process, including how the Wave 2 processing differs from Wave 1's. We recap the problems we encountered and how we approached them. We close by looking ahead to Waves 3 and 4, and how we can apply the lessons learned from the previous waves to make future waves and panels more efficient.

Paradata from the Canadian 2016 Census of Population Electronic Questionnaires
Anthony Bremner, Statistics Canada

The Canadian Census of Population has used online questionnaires since 2006. In 2011, 53.8% of households submitted their census form online. In 2016, that number increased to 68.3%. The goal for 2021 is to have even higher online submission rates. Statistics Canada analyzed paradata from the 2016 online Census forms to study: respondent behaviour, data quality and data collection improvements. Compared with paper questionnaires, the 2016 online forms improved data quality and eliminated the need for some field follow-up. This presentation will provide highlights of the paradata analysis with a particular focus on findings by type of device by respondent profiles, respondent behaviour and respondent burden. The presentation will provide an analysis of online users by type of device including their demographic profile. Respondent behaviour will contain information regarding patterns of behaviour (e.g., use of a 'save' button, abandoning the form, etc.); switching of devices (e.g., from smartphone to other device); last question completed; item nonresponse and edit failure rates. The length of time to complete sections of the questionnaire will illustrate respondent burden. The audience will learn the importance of paradata analysis as a tool to understand the impact of respondent's behaviour on questionnaire design. What devices are being used? Who are using smartphones? Is the user experience on these devices the same as using desktop computers? Is the online application more difficult to navigate on a smaller device? Is the font on a smartphone readable? How much time was needed to complete a question, or submit a questionnaire? What effect does this have on data quality? The answers to these questions could have an influence on the design of future online forms.

Statistics Canada's Experiences to Manage Responsive Collection Design for Household Surveys
Francois Laflamme, Statistics Canada

Paradata research has focused on identifying opportunities for strategic improvement in data collection that could be operationally viable and lead to enhancements in quality or cost efficiency. To that end, Statistics Canada has developed and implemented a responsive collection design (RCD) strategy for computer-assisted telephone interview (CATI) household surveys to maximize quality and efficiency and to potentially reduce costs. RCD is an adaptive approach to survey data collection that uses information

Session 2C: 1:15 pm – 2:30 pm
Paradata
Room: Conference Rooms 3 & 4

available prior to and during data collection to adjust the collection strategy for the remaining in-progress cases. In practice, the survey managers monitor and analyze collection progress against a predetermined set of indicators for two purposes: to identify critical data-collection milestones that require significant changes to the collection approach and to adjust collection strategies to make the most efficient use of remaining available resources. In the RCD context, numerous considerations come into play when determining which aspects of data collection to adjust and how to adjust them. Paradata sources play a key role in the planning, development and implementation of active management for RCD surveys. Since 2009, Statistics Canada has conducted several RCD surveys. This paper describes Statistics Canada's experiences in implementing and monitoring this type of surveys.

Session 3A: 2:45 pm – 4:00 pm
Using Incentives to Encourage Survey Participation
Room: Auditorium 1 & 2
Chair: Catherine Haggerty, *NOROC at the University of Chicago*

The Evolution of the "Early Bird" Offer on the NLS
Kymn Kochanek*, NORC at the University of Chicago*
Lauren Seward*, NORC at the University of Chicago*
Vicki Wilmer*, NORC at the University of Chicago*

The National Longitudinal Survey of Youth – 1979 was on the forefront of offering an "early bird" incentive to respondents to call in, make and keep appointments before additional outreach was made to convince them to complete their interview. The first test was conducted in Round 20 and now after Round 27 we are able to provide a retrospective on how the biennial offer of additional money has impacted the study in terms of maintaining the panel at a high response rate, hastening completion to keep to the timeline, and getting the easiest cases cheaply in order to invest in the toughest cases at the end. The Round 27 also provided a new opportunity to offer the "early bird" to the Young Adult sample, children of the female youth respondents, aged 12-35. This paper examines this generation's uptake of this incentive strategy compared to their parents'.

Using the Data You Have to Get the Data You Need: Propensity modeling, incentive escalation, and responsive design for the Survey of Consumer Finances
Kate Bachtell, *NORC at the University of Chicago*
Catherine Haggerty, *NORC at the University of Chicago*
Shannon Nelson, *NORC at the University of Chicago*
Kevin Moore, *Board of Governors of the Federal Reserve System*

As for other national surveys that collect complex financial data, the use of monetary incentives to encourage participation for the Survey of Consumer Finances (SCF) has expanded over the past decade. In 2016 researchers at the Board of Governors of the Federal Reserve System (FRB) and NORC at the

Session 3A: 2:45 pm – 4:00 pm
Using Incentives to Encourage Survey Participation
Room: Auditorium 1 & 2

University of Chicago (NORC) implemented an incentive escalation strategy involving the use of propensity modeling to identify households posing the highest risk of nonresponse early in the field period. In this paper we first examine the efficacy of the 2016 incentive escalation strategy by comparing the final 2016 results to a simulated test using the 2013 SCF data. We then describe plans for the refinement and expansion of the 2016 escalation strategy within a responsive design for the 2019 SCF.

Optimal Offer Strategies in a Paperless World – Incentive experiment results from a multi-wave student survey
Jill Connelly, *NORC at the University of Chicago*
Karen Grigorian, *NORC at the University of Chicago*
Scott Sederstrom, *NORC at the University of Chicago*

Although there is general consensus about optimal ways to offer incentives in phone and mail surveys, researchers are still discovering how incentives can best be used to increase participation in web surveys particularly when contact with the survey sample is limited to electronic modes like email and texts. To learn more about incentives offered under these conditions, NORC implemented an experiment during the 1st data collection wave for the CollegePoint Virtual Advising Program Evaluation Surveys when 8,000 HS seniors who had been offered the opportunity for CollegePoint advising were recruited to complete a 10 minute web survey. The incentive was an Amazon gift code, but the timing, amount, and associated messaging was uncertain. To obtain reliable incentive information, we conducted an experiment that was designed to test 3 conditions: 1) Timing of incentive delivery (i.e., pre-paid vs. post-paid); 2) Incentive amount (i.e., $10 or $15); and 3) Foreshadowing (i.e., foreshadowing additional incentives for future participation and a bonus for perfect participation or not foreshadowing). We presents the results of this incentive experiment and its impact on subsequent waves.

Unpacking the Use of Incentives in Probability-Based Web Surveys: Evidence across recent experimental studies
Michael J. Stern, *NORC at the University of Chicago*
Erin Fordyce, *NORC at the University of Chicago*

There have been major advancements in the development of best practices for the administration of surveys that offer a web option. However, there are still significant gaps in our knowledge. One area of significant concern is the use of incentives to drive people to the web. Although we can effectively deliver pre-incentives through mailed letters for web-based surveys, it is also common to include a URL in the cover letter and ask the respondent to type it into their computer's web browser. Therefore, there is a burdensome step added to the engagement process, which can discourage participation. Here, we attempt to unpack some these issues by examining what incentives motivate respondents compared to those that perform well in other modes through examining five experimental studies, all of which are different in some an important way. Importantly, the studies include using the web as part of multimode, sequential design as well as experimental studies using all three modes fielded concurrently. The results

Session 3A: 2:45 pm – 4:00 pm
Using Incentives to Encourage Survey Participation
Room: Auditorium 1 & 2

from this work paint a broad picture of how incentives function within and across studies using a web based survey component.

Session 3B: 2:45 pm – 4:00 pm
Roundtable: Management challenges associated with Employee Recruitment / Retention / Development in Survey Research
Room: Auditorium 3
Moderator: Jane Shepherd, *Westat*
Panelists: Kyle Fennell, *NORC at the University of Chicago*
Eloise Parker, *U.S. Census Bureau*
Karen Davis, *RTI International*
Jennifer Edgar, *Bureau of Labor Statistics*

This panel will discuss challenges of recruiting, retention, and staff development in survey research and programming. Organizations face many challenges in the current recruiting environment and the panel will explore approaches to staff development and retention. The panelists will discuss best practices and examples of how their organizations have addressed these ongoing challenges and how they develop strategic approaches to plan for future needs including staff training and development, and succession planning.

Session 3C: 2:45 pm – 4:00 pm
Survey Management
Room: Conference Rooms 3 & 4
Chair: Jason Fields, *U.S. Census Bureau*

Active Management Framework to Monitor and Manage Data Collection
Francois Laflamme, *Statistics Canada*

Data collection survey managers need efficient and practical tools to assess and actively manage data collection survey processes and performance. The trend towards decreasing response rates in household surveys, the increasing complexity of collection strategies, the increasing requirements in terms of providing timely and factual information and the recent changes in the collection vision at Statistics Canada have clearly identified the need to develop more adapted tools to actively monitor and manage data collection throughout the collection cycle. This paper essentially describes the Active Management Framework developed and implemented at Statistics Canada to monitor and manage data collection for all household and agricultural surveys.

Session 3C: 2:45 pm – 4:00 pm
Survey Management
Room: Conference Rooms 3 & 4

Developing a Quick Business Survey for an Existing Online Platform
Sharon S. Stang, *Bureau of Labor Statistics*
Emily Thomas*, Bureau of Labor Statistics*

The Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW) is a federal/state cooperative program that publishes a count of monthly employment and total quarterly wages. QCEW covers 98 percent of U.S. jobs. Estimates are available at the county, MSA, state and national levels by industry. To maintain accurate industry and geographic data for businesses, the QCEW program conducts the Annual Refiling Survey (ARS). This survey asks businesses to review and either verify or update their industry and geographic information on a three-year basis. Approximately 1.2 million establishments are solicited annually for the ARS. As a short survey that reaches a large audience annually, the ARS offers the opportunity to append additional surveys after respondents complete the ARS. QCEW developed a pilot Business Research Survey (BRS) and tested this process in 2018. This paper will outline the BRS development, test, and results.

Creating a Redesigned Questionnaire for the Consumer Expenditure Survey using Colectica
Parvati Krishnamurty, *Bureau of Labor Statistics*

Changes to questionnaire items are routine in surveys. In addition, as surveys attempt to be more flexible to lower respondent burden, skip patterns become more complex. The documentation of changes to survey items and the interdependencies between the item, and how easily these documentation can be accessed, significantly affect the efficiency with which survey metadata can be queried. The quality and accessibility of these documentation also significantly affect the efficiency of updates to data processing systems needed to accurately process these survey changes. One of the key elements of the Gemini redesign of the Consumer Expenditure (CE) Survey involves creating an updated and streamlined version of the current CAPI questionnaire. This instrument redesign also provides an opportunity for the CE program to move towards more efficient and comprehensive documentation of survey metadata. Colectica Questionnaires software is being used to create specifications for the new CAPI instrument which will later be programmed into Blaise. Colectica Questionnaires, is part of the Colectica suite of software based on the Data Documentation Initiative (DDI) international standard for describing surveys, and generates various outputs including source code for computer assisted information systems. We will discuss the important features of the CAPI instrument redesign, such as aggregation, question order, use of screeners, record use, as well as the CE program's experience with using Colectica to generate specifications for the redesigned instrument.

Session 3C: 2:45 pm – 4:00 pm
Survey Management
Room: Conference Rooms 3 & 4

Using Design Thinking to Build Innovative Systems
Jean E. Fox, *Bureau of Labor Statistics*

Organizations of all kinds are using the Design Thinking process to (1) build innovative systems that better meet users' needs, (2) facilitate collaboration, and (3) reduce risk. Design Thinking is a structured process that relies on several tenets, including:

- an early and continual focus on the users throughout the development lifecycle
- a focus on proposing hypotheses to best tested, rather than generating solutions to be implemented without testing
- the establishment of teams with diverse membership Design thinking is an established process, with a framework guiding the activities at all stages of development.

As reported by Liedtka (2017), the approach is fairly consistent across all types of organizations, including government, non-profit, and private. It also crosses domains, illustrating that design thinking can be adapted to a variety of projects. For example, design thinking could help statistical agencies develop innovative solutions for data collection, analysis, and dissemination that better meet the needs of the respondents, survey managers, analysists, and any other stakeholders. It could also be used for other projects involving a product or service people would use, including administrative systems, customer services, and training programs. In this presentation, I will define Design Thinking. I'll describe the stages in the structured process and the methods used at each stage. I will explain how it is used in a wide variety of organizations (including government agencies), and share the benefits of this approach to system development. Our agency is just starting a few projects with this approach, so if possible, I will share experiences using the method. Liedtka, J. (2017). Evaluating the impact of design thinking in action. Academy of Management Annual Meeting Proceedings, 2017(1), pp 1-6.

Wednesday, April 18, 2018

Arrival and Registration: 8:00 am – 9:00 am

Pre-Function Area

Session 4A: 9:00 am – 10:15 am

Introduction to Online Testing for Research at Federal Agencies

Room: Auditorium 1 & 2
Chair: Erica C. Yu, *Bureau of Labor Statistics*

Small-scale research can provide an important service to survey programs by providing evidence to support changes to questionnaires and data collection procedures before they are implemented in the field. An emerging technology in this field is the use of online testing platforms to increase the efficiency of this research – it can reduce time and costs, improve data quality, and more. This session will introduce the audience to online testing, provide guidance on how to supplement and complement in-lab research, and detail case studies from our experiences using several platforms. The session will also provide practical considerations for using these platforms at your own agency.

Integrating Online Platforms into Survey Research

Erica C. Yu, *Bureau of Labor Statistics*
This presentation will describe online testing and how it fits into the research process. We will describe the types of platforms available to use and the advantages and disadvantages of different methods.

Choosing a Sample and Platform: Research design considerations for online testing

Jessica Holzberg, *U.S. Census Bureau*
This presentation will elaborate on the online testing platforms described in the previous presentation. We will describe different types of sample providers and testing platforms available to researchers interested in conducting online testing, and make recommendations on how to select the most appropriate tools based on the research questions.

Approaches to Using Online Samples to Confirm, Inform, and Augment Lab Studies

Robin L. Kaplan, *Bureau of Labor Statistics*
We will describe three case studies in which online testing was used: confirming previously conducted qualitative results collected in an exploratory lab study, informing the research questions and protocol for an in-lab study, and augmenting ongoing lab research.

Policy Considerations for Online Testing

Jennifer Childs, *U.S. Census Bureau*
We will discuss the logistics of introducing the use of these platforms at your own agency. As was the case when surveys started moving from PAPI to CASIC, there are many policy considerations when moving toward online pre-testing. Jennifer Childs will cover some of these considerations by walking through the Census Bureau's path from initial consideration to a formalized plan, including IT security, data protection, and confidentiality protections.

Session 4B: 9:00 am – 10:15 am
Field Management

Room: Auditorium 3
Chair: Joan Hill, *U.S. Census Bureau*

The Effects of Survey Enhancements on Reporting in the Medical Expenditure Panel Survey

Samuel H. Zuvekas, *Agency for Healthcare Research and Quality*
Adam I. Biener, *Agency for Healthcare Research and Quality*
Wendy D. Hicks, *Westat*

It is well established that survey respondents imperfectly recall health care use in surveys. However, careful attention to both survey design and fielding procedures can enhance recall. We seek to examine the effects of a comprehensive, multi-pronged approach to changing field procedures in the Medical Expenditure Panel Survey (MEPS) to improve quality of health care use reporting. Conducted annually since 1996, the MEPS is the leading large-scale nationally representative health survey with detailed individual and household information on health care use and expenditures. These survey enhancements were undertaken in 2013-2014 because of concerns over a sudden level drop in the quality of reporting in the MEPS in 2010 that persisted into 2011-2012. The approach combined focused retraining of field supervisors and interviewers, developing quality metrics and reports for ongoing monitoring of interviewers, and revising advanced letters and materials sent to respondents. We use longitudinal MEPS data from 2008 through 2015, combining detailed information reported by household respondents with sociodemographic and health status characteristics for each person, and paradata on the characteristics of the interviews and interviewers. We take advantage of the longitudinal data and timings of major trainings and changes in field procedures in regression models, separating out the effects of the trainings and other fielding changes to the extent possible. We find the 2013-2014 data quality improvement activities substantially improved reporting quality. Positive interviewer behaviors increased substantially to above pre-2010 levels, and utilization reporting has recovered to above pre-2010 levels, returning MEPS to trend. Importantly, these substantial gains occurred in 2013 prior to extensive in-person training for most of the field force. Lessons learned from the 2013-2014 data quality initiative, especially the importance of respondents' record keeping, are central to a major redesign of the MEPS survey instrument scheduled to be fielded in January of 2018.

Project Management for Field Managers

Kyle Fennell, *NORC at the University of Chicago*

In 2017, staff from across NORC participated in a review of the role field managers play is guiding data collection efforts to successful conclusions. This review highlighted the degree to which field management is split between process driven tasks (assigning cases, holding report calls,...) and activities which fit the PMI definition of a project in that they are temporary endeavors undertaken to create a unique product or service. This presentation details the types of "projects" field managers run while managing data collection efforts. For each type of project, the benefits of applying formal project management techniques and insights is explored. NORC has included formal project management training as part of a new training curriculum rolled out in 2018. This presentation concludes by outlining this curriculum.

Session 4B: 9:00 am – 10:15 am
Field Management
Room: Auditorium 3

Using Machine Learning Methods to Improve Responsive Designs in Face-to-Face Surveys
Jane Shepherd, *Westat*
Gonzalo Rivero, *Westat*
Brad Edwards, *Westat*
Roger Tourangeau, *Westat*
Tammy Cook, *Westat*

Responsive designs for data collection are one of the most promising theoretical developments that tackle the falling response rates of probability surveys. Adapting the sample composition during the fielding period to incorporate available information about the observed cooperation from the selected sample is expected to increase efficiency and reduce the operational cost of data collection in relation to a fixed design. However, while appealing in theory, responsive designs pose a number of challenges for their practical implementation in face-to-face surveys. In this paper we discuss the design, implementation, and performance of a fully integrated responsive design model that Westat built for the PATH Reliability study. We focus on the development of an array of machine learning models that use the latest available paradata and contextual information for each case to produce and sequentially update response propensities for each of the interview tasks. We also discuss the deployment of our models to a fully automatized system that integrates response propensities with a planning and job scheduling engine. Finally, we present a set of experiments that we deployed to measure interviewer compliance. Short summary of topics included:

- Describe the tools and processes used to develop the statistical models.
- Describe the infrastructure that hosts the models, feeds the data, and exposes the predictions to the final user.
- Discuss the methodological design to estimate the business value of the tool
- Discuss future plans for the tool

Using Paradata to Develop and Implement an Interviewer Performance Profile
Heidi Guyer, *Survey Research Center at the University of Michigan*
Wen Chang, *Survey Research Center at the University of Michigan*
Brady West, *Survey Research Center at the University of Michigan*

A large national field study has utilized paradata to monitor interviewer performance for over ten years. New quality indicators have recently been developed and organized into a performance profile using heat maps to highlight interviewers' performance. This presentation will describe the key statistics identified for monitoring purposes, the paradata utilized, and the development and display of the quality indicators. First, principal component analysis was used to assess keystroke data from the interview to identify predictors of poor performance or potentially poor data quality. Three main areas were identified: 1) reading question text too quickly, 2) frequent error checks and 3) a high proportion of refused or don't know responses. Z-scores were calculated for interviewers to determine if they were above or below the mean values of each component. Interviewer scores were highlighted in green to indicate those with the

Session 4B: 9:00 am – 10:15 am
Field Management
Room: Auditorium 3

best performance on any of the three characteristics, or in red to indicate those with the worst performance. Next, the estimated eligibility propensity, response propensity and contact propensity were estimated for each of the cases in an interviewer's assigned sample using logistic regression. The estimated likelihoods are compared to the actual outcomes for each case on an on-going basis utilizing a propensity-adjusted interviewer performance (PAIP) indicator, which is designed to evaluate the performance after accounting for the difficulty of the assigned sample. The PAIP indicator is displayed for each component and for each interviewer and highlighted in the same way as the quality scores (green=good, red=bad). Additional proxy indicators for nonresponse bias are also displayed in the profile. The information is organized in easy-to-read Excel spreadsheets with one row per interviewer. This allows for a quick visual assessment of high performers versus low performers. Low performers are counseled in areas of concern (i.e. lower than expected eligibility rate, going too fast, etc). On-going monitoring for improvement is possible due to the continuous updating of information as the data are collected. The interviewer performance profile is useful for managers to monitor and guide interviewers as well as for survey methodologists to determine specific areas where measurement error may be occurring.

Session 4C: 9:00 am – 10:15 am
Quality Issues in the Integration of Multiple Data Sources
Room: Conference Rooms 3 & 4
Moderator: John L. Eltinge, *U.S. Census Bureau*
Panelists: Alexandra Brown, *Joint Program in Survey Methodology*
Chris Chapman, *National Center for Education Statistics*
Joseph L. Schafer, *U.S. Census Bureau*
Linda J. Young, *National Agriculture Statistics Service*

In recent years, statistical agencies in the United States and other countries have developed increasing interest in the integration of multiple data sources for the production of statistical information. These sources often include traditional sample surveys, but may also include forms of "non-designed data" (sometimes described as "organic data" or "big data"), e.g., administrative or commercial records; extracts from websites or images; social media traces; and passive sensors.

Decisions regarding the prospective use of such sources depend on a wide range of issues that require rigorous assessment of multiple dimensions of quality, risk and cost. Over the past several months, a working group of the Federal Committee on Statistical Methodology (FCSM) has explored some of these issues through a series of three workshops and several related stakeholder meetings. This panel session synthesizes some results of that exploration, with emphasis on:

(1) criteria that may be important in developing transparent reports on some predominant dimensions of input data quality, processing quality and output data quality; and
(2) high-priority areas of methodological and empirical research that may lead to a more refined understanding of (1), and to improved design and practice in the integration of multiple data sources.

The session will close with discussion a set of questions for audience members intended to shed additional light on the issues identified in (1) and (2).

Session 5A: 10:30 am – 11:45 am
Online Testing to Improve Surveys

Room: Auditorium 1 & 2
Chair: Erica C. Yu, *Bureau of Labor Statistics*

Using Mechanical Turk to Collect Longitudinal Data

Jesse Chandler, *Mathematica Policy Research*
Leib Litman, *Lander College*
Jonathan Robin, *Lander College*

Longitudinal studies have many advantages over cross-sectional studies, but with the tradeoff of having to retain workers across waves of data collection. For this reason, designing an appropriately powered longitudinal study requires an accurate estimate of retention rates. Unfortunately, estimates of retention rates for online research studies are not readily available. We estimate retention rates for longitudinal studies conducted on Amazon Mechanical Turk (MTurk) through a meta-analytic review of 157 longitudinal studies conducted on this platform. Average retention rates are low but highly variable across studies. We also demonstrate that a large portion of the variability in retention rates is explained by factors that are controlled by researchers, including whether workers are notified about the study, whether the sample is restricted to reputable workers and how much money workers are paid. Retention rates can be increased to above 75% when best practices are followed.

Using Online Crowdsourcing Platforms to Obtain Information from Special Populations

Rebecca L. Morrison, *National Center for Science and Engineering Statistics*
Flora Lan, *National Center for Science and Engineering Statistics*
Karen S. Hamrick, *National Center for Science and Engineering Statistics*

In mid-December 2017, the National Center for Science and Engineering Statistics (NCSES), a federal statistical agency within the National Science Foundation, conducted a brief survey with participants recruited from an online crowdsourcing platform (Amazon's Mechanical Turk, also known as MTurk) as a pilot project on quick turnaround pre-testing. The survey attempted to collect information from the college-educated population, a group of great interest to the agency. The study also included two experiments: one where each participant received one of three confidentiality pledges, and another where all participants answered multiple questions about income. This presentation will shed light on four research questions: 1. How successful were we at recruiting and collecting data from our target population, college graduates? 2. How closely did the MTurk-recruited population match the college-educated population on key demographic characteristics? 3. How did participants' comprehension of the three confidentiality pledges vary? Did the confidentiality pledges affect response? 4. How consistently did participants report their income within the survey? Finally, we will provide some insight into practical and operational matters, such as the materials provided to obtain OMB approval and MTurk's utility for our purpose of conducting rapidly-deployed brief surveys.

Session 5A: 10:30 am – 11:45 am
Online Testing to Improve Surveys
Room: Auditorium 1 & 2

Using Online Survey Panels to Evaluate Response Error in Health Insurance Questions
Bridget M. Reynolds, *National Center for Health Statistics*
Paul Scanlon, *National Center for Health Statistics*

Policy changes affecting the health insurance landscape in the United States have renewed interest in, and required modifications to, survey instruments that measure health insurance coverage in the population. Given these marketplace changes, as well as longstanding complexities surrounding coverage types and options, many Americans experience uncertainty and confusion when asked about their health insurance plans. To the extent that this uncertainty impacts the accuracy of survey responses, it can result in error and bias in estimates of health insurance coverage. Cognitive testing studies conducted by federal agencies have consistently uncovered potential for response error in relation to survey questions about health insurance. To examine whether and how these qualitative findings generalize to a larger population, a series of close-ended web probes were embedded alongside National Health Interview Survey (NHIS) health insurance questions on a questionnaire administered to an online panel of over 2,000 survey respondents. These web probes asked about the respondent's confidence in his or her answers to the health insurance questions, how knowledgeable they felt about health insurance, and how they obtained their insurance. Additionally, the web panel vendor provided supplementary data on a sub-sample of respondents who had previously completed one of their in-house health surveys. We will use responses to the embedded web probes, along with the vendor-supplied auxiliary data, to assess the likelihood of response error in the health insurance questions. This presentation will provide preliminary findings from this web survey project [the NCHS Research and Development Survey (RANDS)] and will use the findings as context for a broader discussion on some of the promises and potential pitfalls of using commercially available web panels for question evaluation and design projects.

Willingness of the Public to Share Geolocation Information in a U.S. Census Bureau Survey
Elizabeth M. Nichols, *U.S. Census Bureau*
Erica Olmsted-Hawala, *U.S. Census Bureau*

In 2016 the U.S. Census Bureau conducted a split-panel experiment to explore the public's willingness to share geolocation information within a survey. A sample of 2,000 participants from the Census Bureau's nonprobability panel were invited to take part in the study. They were sent an email asking them to complete an online survey on a mobile device while they were at home. Within the survey, one question asked for their address and then the survey requested permission to access their geolocation information. Depending on the study condition, the survey varied how the geolocation request was made using a two-by-two design. One factor experimentally manipulated the number of geolocation permission requests: either by only using the device's default location permission request, or by an explicit question within the survey asking permission followed by the device's default location permission request. The other factor experimentally manipulated the place in the survey when the address and geolocation requests appeared: either towards the beginning of the survey or towards the end of the survey. Results showed that the treatment that explicitly asked for permission in addition to the device's default permission request marginally increased the sharing of the geolocation data. Results also showed that

Session 5A: 10:30 am – 11:45 am
Online Testing to Improve Surveys
Room: Auditorium 1 & 2

placing the address and geolocation request towards the end of the survey significantly increased the sharing of that data. Results indicated that respondents with more education and nonminority respondents shared their location data more often, but allowing location data to be shared did not depend on age or sex of the respondent. Assuming that the respondents reported truthfully that they were at home while taking the survey and entered their home address, we found the geolocation data to be accurate to the correct block a little more than 50 percent of the time.

Session 5B: 10:30 am – 11:45 am
CAPI Tools
Room: Auditorium 3
Chair: John Baker, *U.S. Census Bureau*

Improving Data Collection for Prescribed Medicines through an Enhanced Lookup Tool
Jennifer Vanicek, *NORC at the University of Chicago*
Emma Kaufman, *NORC at the University of Chicago*
Andrea Mayfield, *NORC at the University of Chicago*
Becky Reimer, *NORC at the University of Chicago*
Sarah Lehan, *NORC at the University of Chicago*
Y. Michael Yang, *NORC at the University of Chicago*
Valeri Cooke, *NORC at the University of Chicago*
B. Peebles, *NORC at the University of Chicago*

How can a survey instrument be optimized to improve the efficiency and accuracy of survey data collected about prescribed medicines? The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. Among many other important measures, MCBS collects detailed information about respondents' prescribed medicines, including the name, strength, and form. NORC and CMS recently designed and implemented an enhanced prescribed medicine lookup tool within the CAPI instrument that streamlines both data collection and data processing for prescribed medicines. The revised design is intended to increase the rate at which interviewers use the lookup to enter medicines, minimize manual entry of medicine data, and provide flexibility to accommodate situations in which the respondent does not have complete information about reported medicines. Furthermore, the new lookup allows interviewers to select medicines in a way that closely mirrors the data files used by CMS for post-processing and claims matching, reducing the need for manual data review and matching procedures. This presentation will outline the motivation for developing the new tool and describe the process of integrating the lookup into a complex questionnaire design. In addition, the process of developing a new data structure to allow for easier data cleaning and matching to post-processing data files will be described. A demonstration of the lookup functionality will be provided, highlighting the new interface and flexibility of the tool for field interviewers. Finally, the presentation will feature an early assessment of the results of the lookup on

Session 5B: 10:30 am – 11:45 am
CAPI Tools
Room: Auditorium 3

some paradata and data quality measures, such as interview timings, frequency of manual medicine data entry as compared to the prior version, and impacts on the medicine name cleaning process.

Complex Use of Voxco, Commercial-Off-the-Shelf Software, for Data Collection
Nathan Sikes, *RTI International*
Jean Robinson, *RTI International*
Ying Qin, *RTI International*
Michelle Krzyzanowski, *RTI International*

This presentation will address the use of Voxco, a commercial-off-the-shelf (COTS) software, for complex data collection studies. Study designers may select COTS software for a variety of reasons, including but not limited to financial considerations and client requirements. "COTS" typically implies a one-size-fits-all approach that meets most of the needs it was designed to address. However, complex studies using COTS software require thoughtful and ingenious approaches to be successful. We will describe innovative solutions our programming staff has developed using Voxco and in-house enhancements to meet study requirements that Voxco alone could not fully address. These solutions are not modifications of Voxco itself, but programming solutions applied on top of the software. Programming staffs in other organizations can apply these solutions as well. We will share how we implemented these features, among others, to enhance the use of Voxco in our organization:

- Customized grid-type questions with checkbox, radio button, dropdown list, and free text all in one question
- The use of JavaScript code to validate data input by the user to overcome limitations of Voxco validation
- Implementation of HTML styling to customized questions
- Automated initialization of surveys, eliminating the need for manual intervention
- Preloading multiple instruments at once, even if we weren't sure an instrument would be used by the respondent
- Automation of preloads from previously answered surveys to new surveys

Study designers can overcome the limitations of COTS software with clever programming and a will to make things work. We expect survey managers and developers to benefit from this presentation.

Session 5B: 10:30 am – 11:45 am
CAPI Tools
Room: Auditorium 3

Moving from Blaise 4.8 to Blaise 5: Preparing a prototype survey based on the NSDUH
Gilbert Rodriguez, *RTI International*
Martin Meyer, *RTI International*
Vorapranee Wickelgren, *RTI International*
Emily Geisen, *RTI International*
Hilary Zelko, *RTI International*
Ramasubramanian Suresh, *RTI International*

Some surveys that were once programmed using Blaise 4 are now being migrated to Blaise 5 to take advantage of the new features and capabilities that Blaise 5 offers. However, that migration is not without challenges. For example, functionality that was previously performed in alien routers and with layout instructions must now be implemented differently. SAMHSA and RTI are currently in the process of evaluating Blaise 5 for potential future programming of the National Survey on Drug Use and Health (NSDUH) questionnaire. NSDUH is sponsored by the Substance Abuse and Mental Health Services Administration and provides national, state and substate data on substance use and mental health in the civilian, noninstitutionalized population ages 12 and older. Approximately 67,500 NSDUH interviews are completed annually and approximately 600 field interviewers are staffed on the study. Through some initial experimentation and prototyping, the NSDUH software development team has figured out how to replicate the majority of the NSDUH survey's Blaise 4.8 functionality using Blaise 5. This presentation will share some observations, and some lessons the NSDUH team learned while preparing a prototype Blaise 5 demonstration survey, using Blaise 5, and targeting touchscreen capabilities of Windows 10 laptops.

MEPS: Using technology to reduce burden now and improve quality later
Wendy D. Hicks, *Westat*
Adam I. Biener, *Agency for Healthcare Research and Quality*
Marie Stagnitti, *Agency for Healthcare Research and Quality*

The Medical Expenditure Panel Survey (MEPS) is a set of large-scale surveys of families and individuals, their medical providers (doctors, hospitals, pharmacies, etc) and employers across the United States. The Household Component of MEPS collects data from individual households and then supplements those data with a follow-up operation contacting the household members' medical providers. The follow-up operation is called the Medical Provider Component. The Medical Provider data are used solely for editing and imputation of the HC data, and improve the quality of the estimates of health care costs. This January 2018, MEPS launched a new CAPI instrument for the Household Component. Given the importance of the medical provider component of MEPS in producing accurate estimates of health care costs, the new CAPI instrument included a new tool that helps interviewers and respondents identify the medical providers reported by household members. The medical provider information captured with the tool will serve as input for the Medical Provider follow-up operation. The new tool uses a search feature similar to that used by Google and other common search engines, and intends to: - ease the burden during an interview in identifying medical providers, - reduce reporting of independent providers covered by the same parent organization, reducing HC respondent consent requirements as well as streamlining MPC follow-up

Session 5B: 10:30 am – 11:45 am
CAPI Tools
Room: Auditorium 3

operations - increase the accuracy of the provider information for the follow-up operation, This paper discusses the technology used for the new search tool, and the use of historical panel data to help guide the development of the requirements for the tool. We will provide a short demonstration of the new tool and discuss the planned approach for evaluating the tool to guide future enhancements.

Session 5C: 10:30 am – 11:45 am
Leveraging Alternative Data and Methods
Room: Conference Rooms 3 & 4
Chair: Lisa Blumerman, *U.S. Census Bureau*

Assessment of Bias in Estimates from a Sample of Self-Reported Web Users
Meena Khare, *National Center for Health Statistics*

Nonresponse and increasing survey costs continue to influence traditional methods of data collection. In recent years, researchers are exploring alternative methods of data collection to reduce bias and survey cost. According to the Pew Research Center telephone surveys, the Internet use among the U.S. adults has increased from 14% in 1996 to 88% in 2016. Therefore, Web surveys and multimode data collection from households and establishments have become a common, alternative cost-saving approach. The National Health Interview Survey (NHIS) has been collecting information on the Internet and email usage among adults since 2012. Among 2014-2015 NHIS adult respondents, 74.1% report using the Internet or emails. This paper presents an assessment of bias in selected estimates using data from a random sample of self-reported Web users (defined as an NHIS respondent who uses the Internet or email) and compares with the overall NHIS estimates.

Consumer Price Index Outlet Frame Survey Redesign
Madeleine Saxton, *Bureau of Labor Statistics*
Anya Stockburger, *Bureau of Labor Statistics*

As part of its sampling framework, the U.S. Consumer Price Index (CPI) conducts the Telephone Point of Purchase Survey (TPOPS) as an independent survey of households to obtain an unbiased and representative frame of stores and service providers (outlets) from which to select the Commodities and Services (C&S) Pricing Survey sample. As a random digit dial (RDD) telephone survey, TPOPS is no longer a cost effective frame source and the Bureau of Labor Statistics anticipates replacing this survey with the collection of new establishment data in the Consumer Expenditure Survey (CE). In addition to eliminating the reliance on outdated RDD methodology, the consolidation of the two household surveys used to construct the CPI reduces aggregate respondent burden by eliminating redundant questions (e.g. expenditure data and demographic information) and simplifies CPI weighting formulas. The coupling of where consumers shop with the host of other demographic, income, and expenditure information already collected in the CE Surveys also creates opportunities for future research to improve CPI

Session 5C: 10:30 am – 11:45 am
Leveraging Alternative Data and Methods
Room: Conference Rooms 3 & 4

compilation methodologies. This paper will describe the benefits, challenges, and opportunities for the US CPI's sampling methodology as part of the outlet frame survey redesign.

Challenges of Using Administrative Records to Supplement Response Data in the ACS
Sandra L. Clark, *U.S. Census Bureau*
R. Chase Sawyer, *U.S. Census Bureau*

As part of the "big data revolution", administrative records have become a hot topic in survey research and are increasingly used to replace or supplement survey data. The U.S. Census Bureau is researching ways to use administrative records in the American Community Survey (ACS). Research has found impressive match rates for some survey items when linking administrative records to households in sample for the ACS; suggests there is potential for reduction in respondent burden by asking fewer survey questions of ACS responders; and, an opportunity to significantly improve item missing data rates for some survey items. However, there are numerous challenges the ACS program will need to overcome in order for the survey to make use of administrative records. The Census Bureau recently conducted a test to simulate the use of publically-available administrative records to replace responses to the 2015 ACS for the following four survey items: property value, property tax, year structure built, and acreage. This research discusses the challenges observed during the test, such as obtaining and securing administrative records, implementing the use of the data into the survey operations, and the impact the use of the administrative records would have on published ACS estimates.

Applying Machine Learning Techniques to Transportation Surveys
Jane Shepherd, *Westat*
Marcelo Simas, *Westat*
Anthony Fucci, *Westat*
Alexander Cates, *Westat*

Over the past few years a group of Westat data scientists and survey practitioners has worked together to integrate several text processing and supervised machine learning techniques into transportation surveys, primarily for household travel surveys. This presentation will showcase the application of machine learning and other data science techniques to recently concluded or still active travel surveys: Â· Use of text processing and random forest models to assist in up-coding open-end text responses Â· Use of text processing and vector space model connected to online instruments and trained using the NAICS category descriptions to assist participants in selecting an occupation from Standard Occupational Classification (SOC) system in a household travel survey pilot; and Â· Use of spatial point clustering, random forests and decision trees to assist in the processing of travel data derived from GPS traces captured using smartphone apps. These methods have been mainly developed and applied using the R statistical computing tool, which was integrated with other components using a common relational database and the OpenCPU platform. The presentation will conclude by identifying lessons learned in the

Session 5C: 10:30 am – 11:45 am
Leveraging Alternative Data and Methods
Room: Conference Rooms 3 & 4

application of these methods and suggestion areas for future research into the application of machine learning methods to survey research.

Session 6A: 12:45 pm – 1:45 pm
Using Online Samples
Room: Auditorium 1 & 2
Chair: Howard Hogan, *U.S. Census Bureau*

Using Digital and Social Media in Research in the 21st Century
Jane Shepherd, *Westat*
Amelia Burke-Garcia, *Westat*

Researchers today are struggling with a complex combination of technological and social changes, e.g. the abandonment of traditional telephone lines and a more migratory population, which make participation in research a difficult proposition. Yet, incorporating digital and social media into traditional social science research methods is a challenge as researchers must strike a balance between participant privacy and confidentiality with the openness and transparency that characterize digital media. Westat has been testing the promise and limitations of these non-traditional strategies for recruiting and retaining participants in Federal government-sponsored research studies for a number of years. During this time, we have accrued numerous learnings about the different channels, audiences and cost implications – and how these can be leveraged to drive more qualified participants earlier in the process, thereby conserving valuable resources. As well, the use of online influencer strategies in communication is a growing area and may hold promise for research. Westat's expertise and experience in these emerging areas will be explored. Numerous demonstration examples will be provided. Short summary of topics included:

- Overview of paid Google, Twitter, Facebook & Instagram advertisements
- Overview of influencer strategies
- Iterative approaches to improve digital strategies over time
- Successes and challenges of each

A Penny for Your Thoughts: Research opportunities presented by Amazon Mechanical Turk
Jesse Chandler, *Mathematica Policy Research*

Online labor markets like Amazon Mechanical Turk (MTurk) make it possible for anyone to recruit large and diverse groups of online study participants rapidly and at low cost, with the tradeoff that study design and sampling are almost entirely "do-it-yourself" In recent years, MTurk in particular has become incredibly popular among social scientists. This talk will provide an overview of what is currently known about the platform, the range of applications it has been successfully used for and the challenges of using it to conduct research. In particular, I will present evidence that MTurk workers are diverse, but by no means representative of any particular population and that the composition of samples can vary depending on the sampling strategy. The flexibility afforded by MTurk's "do-it-yourself" approach to

Session 6A: 12:45 pm – 1:45 pm
Using Online Samples
Room: Auditorium 1 & 2

survey fielding has increased the range of tasks that researchers can ask respondents to complete. At the same time, it has made it easy to overlook potential threats to study validity.

When to use Non-Probability – An evaluation of the use of a non-probability mobile panel in a post-disaster area in comparison to a probability sample
Thomas Brassell, *ICF International*

Non-probability panels offer an alternative survey methodology for post-disaster areas considering the logistical challenges and costs associated with constructing representative probability frames that can account for population displacement and infrastructure damage (Kessler et al., 2008). Specifically, non-probability studies offer solutions to the practical challenges of surveying in post-disaster areas, provided internet access is attainable (Hugelius et al., 2017). Studies examining post-disaster media access have found smartphones to be the primary source connectivity to the internet (Kaigo, 2012), and the optimal communication path for post-disaster government aid (Federal Emergency Management Agency, 2013). Considering this past research, mobile panels, may offer the best method for obtaining responses post-disaster given the frequency of use of smartphone devises by the impacted population.
Our study explores the use of a non-probability mobile panel, in comparison to a traditional random digit dial (RDD) study, as a measure of population displacement, and attitudes and health outcomes post-disaster in Harris County, TX. We will conduct a survey using Mfour's Mobile Research Panel, comparing responses from the mobile panel to those obtained through the RDD telephone survey. In addition, we will obtain demographic information on the mobile panel pre- and post-Hurricane Harvey to assess population displacement within the panel. We will compare these data to those obtained through the RDD telephone survey (pre- and post-Hurricane Harvey), and most recent census data to assess the mobile panel's ability to gauge population displacement.
The results of this study provide vital support for the use of non-probability mobile panels in post-disaster areas as a less costly and quick means of assessing the population's mental and physical health, as well as the damage sustained to their area. In addition, it could provide information on the degree of population displacement experienced in an area, which could be vital for estimation purposes.

Session 6B: 12:45 pm – 1:45 pm
Roundtable: Call Center Technologies
Room: Auditorium 3
Moderator: Michael Gerling, *National Agriculture Statistics Service*
Panelists: Kurt Johnson, *ICF International*
James Dayton, *ICF International*
Amy Bailey, *National Agriculture Statistics Service*

This audience driven session discusses call center infrastructure improvements, software & hardware improvements, VOIP vs. analog, automated dialing. Both small and large call centers will benefit from discussing their challenges and successes that make a successful call center.

Session 6C: 12:45 pm – 1:45 pm
Software

Room: Conference Rooms 3 & 4
Chair: Dave Sheppard, *U.S. Census Bureau*

Lessons Learned - When mobile data collection apps cannot store data on the device

Matthew Boyce, *RTI International*

Mobile devices can be used by both field staff and respondents to gather data, participate in surveys, and provide relevant feedback or intervention. Some apps work continuously online, or via a web portal, which can be problematic if connectivity cannot be maintained. Others, such as our PHIT app platform, work offline, storing data locally in an encrypted database and uploading later whenever a data connection becomes available. Sometimes the security risks of storing data locally, even when encrypted, can outweigh the benefits. In these cases, an application may need to operate in an online-only mode, storing nothing locally. We recently came across a situation that required us to extend our PHIT platform for real-time data storage to a secure central database. In this presentation, we will discuss some of the challenges faced and lessons learned during this transition, as well as future concepts to support both local and remote secure data.

Dusting off the Frame: An open-source computing approach to high quality frame cleaning

Christopher Griggs, *RTI International*
Rich Zemonek, *RTI International*

The success of large census surveys can greatly depend on the accuracy of the sample frame and the sources from which the frame data is populated. In this presentation, we detail the use of Microsoft's newest open source platform, .NET Core, to create a cross-platform, database agnostic frame cleaning application for use on a 22,000-member roster of law enforcement agencies. Using both user-entered and machine imported data, the application layers these submissions and routes them to data managers for vetting and selection. This web application architecture uses a code-first approach, creating database objects rapidly after completion of design documentation. Coupled with the application itself this lets developers to deploy into any platform (AWS, Azure, on-site) in minutes. See how Amazon Web Services (AWS) played a critical part in the development cycle, and the role that great user design and workflow are playing to increase data quality of this large frame.

Jenkins

Federico Vazquez, *U.S. Census Bureau*

Jenkins is on the bleeding edge of technology today. It is also one of the most compelling technologies of the last decade in terms of its approach to software development and operation practices. The Jenkins Continuous Integration solution has become a standby in organizations of all sizes that want to increase productivity and streamline software development in the era of Agile. It has extensive community support has the extended the core functionality of Jenkins by developing thousands of useful plugins. An ecosystem of more than 1,100 plug-ins has emerged, enabling customers to add all sorts of functionality and integrate Jenkins with everything from Active Directory to GitHub to Tomcat. Jenkins is becoming a

Session 6C: 12:45 pm – 1:45 pm
Software
Room: Conference Rooms 3 & 4

must tool for DevOps. It allows companies to build very sophisticated build pipelines very quickly, thus greatly reducing the risk within the software development lifecycle. Tons of companies have already been using Jenkins to implement continuous integration pipeline. In this new paradigm, product teams push their work to customers as quickly as possible so that they can collect feedback and improve upon the previous iteration of their products. Concepts such as minimum viable product (MVP), release candidate, velocity, etc. are all derived from these new approaches. In contrast, product teams using older paradigms like waterfall development might not hear back from customers for months and, quite often, not until the product is commercialized. Continuous Integration Continuous integration (CI) is a software development practice in which developers regularly merge their code changes into a central repository, after which automated builds and tests are run. Continuous integration most often refers to the build or integration stage of the software release process and entails both an automation component (e.g., a CI or build service) and a cultural component (e.g., learning to integrate frequently). The key goals of continuous integration are to find and address bugs quicker, improve software quality, and reduce the time it takes to validate and release new software updates. The basic challenges of continuous integration include maintaining a single source code repository, automating builds (and building fast), and automating testing. Additional challenges include testing on a clone of the production environment, providing visibility of the process to the team, and allowing developers to obtain the latest version easily. Continuous Delivery and Deployment Continuous delivery (CD) is a software development practice where code changes are automatically built, tested, and prepared for production release. It expands upon continuous integration by deploying all code changes to a testing environment, a production environment, or both after the build stage has been completed. When continuous delivery is properly implemented, developers always have a deployment-ready build artifact that has passed through a standardized test process. With continuous deployment, revisions are deployed to a production environment automatically without explicit approval from a developer, making the entire software release process automated. This, in turn, allows for the product to be in front of its customers early on, and for feedback to start coming back to the development teams. Why Use Jenkins? Jenkins is a very popular product among developers who want to automate their CI/CD pipelines. It accomplishes all of the phases of CI and CD. It integrates very well across languages, platforms, and operating systems. It's open-source software.

Session 7A: 2:00 pm – 3:00 pm
Mode Effects
Room: Auditorium 1 & 2
Chair: Adam Safir, *Bureau of Labor Statistics*

New Developments in ARSweb Collection Efforts for Unclassified Establishments in FY 2018
John J. Kane, *Bureau of Labor Statistics*
Mark Sauer, *Bureau of Labor Statistics*

The Annual Refiling Survey (ARS) is used in conjunction with the Unemployment Insurance tax-reporting system in each State so that changes in the industrial and geographical compositions of our economy are

Session 7A: 2:00 pm – 3:00 pm
Mode Effects
Room: Auditorium 1 & 2

captured in a timely manner and reflected in BLS statistical programs. The purpose of the ARS is to verify and update information for eligible businesses - mailing addresses, physical locations including counties, and their main business activities. There are nearly 10.1 million establishments on file with the States. Roughly one-third of private-sector establishments are reviewed and updated annually. BLS conducted the entire FY 2017 ARS using only one-page web letters and emails to respondents. The ARS includes trying to collect information about unclassified establishments for which there is no information available about their current industrial activities. Recognizing that online collection has made it easier to pursue data from unclassified establishments, BLS has piloted new procedures for the FY 2018 ARS to target unclassified establishments. More contacts are attempted than in past years. BLS has plans to incorporate this new process into full production and to examine more ways to assign industry (NAICS) codes to unclassified units. These efforts are consistent with BLS continually pursuing options to reduce employer burden and costs and to take advantage of technological innovations. The primary focus of this presentation will be to describe the new procedures related to unclassified establishments as well as to examine how the impact of the move to online collection exclusively is leading to exploring other options for improving data quality.

Should Specific Households be Targeted for FedEx Reminders? Evidence from a national survey
Mahi Megra, *American Institutes for Research*
Becca Medway, *American Institutes for Research*

The 2016 National Household Education Survey was a two-stage self-administered survey. While most mailings were sent using the United States Postal Service (USPS), the second screener-stage reminder was mailed using FedEx. Preliminary analysis showed that the FedEx mailing led to the highest increase in the screener response rate out of the three screener reminders that were sent. Additionally, bivariate analysis revealed that the respondents to the FedEx reminder were more likely to come from underrepresented groups (such as non-White, less educated, low-income, and rural households), potentially decreasing nonresponse bias. However, although these results suggest that the FedEx mailing is a worthwhile nonresponse follow-up option, using this service can be very expensive. Instead of sending all nonresponding households the second reminder via FedEx, subgroups of households could be targeted with FedEx mailings as part of a tailored design (with the others receiving a cheaper USPS reminder mailing). The current study will expand the preliminary analyses by developing a model to determine which household characteristics available on the frame are the most predictive of a particularly high likelihood of responding to the FedEx mailing. The results will then be used to identify those households that should continue to get a FedEx mailing in future administrations and those that could receive a less costly mailing. Identification of these households will take into account (1) whether the case improves the representativeness of the respondent pool (reducing bias) and (2) the relative cost of FedEx versus USPS mailing for that case. An analysis will be included that assesses the effect of this targeted approach on the expected response rate and cost per complete for future administrations. The results of this analysis will be of interest to researchers considering the cost-benefit tradeoffs of more expensive mailing strategies – particularly those conducting surveys that include FedEx mailings.

Session 7A: 2:00 pm – 3:00 pm
Mode Effects
Room: Auditorium 1 & 2

A Mixed Mode and Incentive Experiment using Administrative Data in the Making Ends Meet
Scott L. Fulford, *Consumer Financial Protection Bureau*
Brian Bucks, *Consumer Financial Protection Bureau*
Mick P. Couper, *Survey Research Center at the University of Michigan*
This paper discusses the impact on survey cost, response rates, and possible nonresponse bias of two experiments: (1) using concurrent or sequential online and then mail survey modes; and (2) different incentive amounts given to initial survey nonrespondents. The experiments were conducted in a pilot version (n=2000) of the Consumer Financial Protection Bureau's Making Ends Meet (MEM) survey. The MEM survey examines US consumers' use of financial products, especially when their finances are tight. The sampling frame is an administrative data set consisting of a 1 in 48 sample of the universe of credit records which includes rich detail on consumers' credit characteristics. The unusually rich information for the sampling frame provides a strong basis for understanding correlates of survey response. In the first experiment, half of the sample was randomly allocated to receive only a letter inviting them to take the survey online or on a mobile phone (online-first). The other half received a letter with the same online instructions as well as a paper version and postage-paid return envelope (mail and online). Both groups received a pre-paid five dollar incentive with the initial survey invitation letter and a reminder postcard in week two. Everyone who had not yet responded from both groups received another letter with a full paper survey and an additional incentive in week five. The second experiment varied the amount of this second incentive: we allocated half of the sample to receive five dollars and half to receive ten dollars. At week five, the response rate for the online-first group was much lower, 3.9 percent compared with 11 percent for the mail-and-online group. The observable demographic and credit characteristics of respondents generally did not differ statistically between the two groups. The exception is age, as respondents in the online-first group were younger. After the week five mailing, the response rates for the online-first group quickly increased, although their ultimate response rate, 18.6 percent, was nearly two percentage points lower than the final response rate for the mail-and-online group. The online-first treatment did yield a much higher fraction of online responses, 30 percent compared with 17 percent for the mail-and-online group. We find no other statistical differences in observable demographic and credit characteristics between the groups for the final sample. Given the lower response rates, we calculate that mail processing costs would have to be nearly eight dollars per returned survey to justify the sequential web then mail approach based on cost. In our incentive experiment, we find that increasing the incentive to $10 in week five had little, and possibly harmful, effect on response rates. In the online-first group, the response rate was 4.2 percentage points lower for those who received ten dollars compared with those who got five dollars. In contrast, in the mail-and-online group, those who received ten dollars had a slightly higher response rate. The differences in response rates by incentive amount, however, were statistically insignificant in both cases.

Session 7B: 2:00 pm – 3:00 pm
Findings from Research with Special Populations

Room: Auditorium 3
Chair: Jean E. Fox, *Bureau of Labor Statistics*

Measuring Experience of Care for People with Disabilities

Coretta Mallery, *American Institutes for Research*
Kerry M. Lida, *Centers for Medicare and Medicaid Services*
Melanie Brown, *Centers for Medicare and Medicaid Services*
Christopher Pugliese, *American Institutes for Research*
Elizabeth Frentzel, *American Institutes for Research*
Susan Raetzman, *IBM*

This research examined the most inclusive methodology for conducting survey research with individuals with disabilities. The research team designed and tested the Consumer Assessments of Health Providers and SystemsÂ® (CAHPS) Home and Community-Based Services (HCBS) or HCBS CAHPS survey to be a cross-disability instrument that assesses services for individuals who are frail elderly, individuals with a physical disability, individuals with an intellectual or developmental disability, individuals with a brain injury, and individuals with serious mental illness. The HCBS CAHPS survey fills a critical need in long-term services and supports (LTSS) quality assurance because it allows for comparisons about the quality of services and supports across Medicaid HCBS programs or between managed care organizations or other subgroups. The 19 NQF-endorsed measures derived from the survey address critical HCBS domains including person-centeredness, community integration, staff communication, unmet needs, and personal safety. The field test of the instrument included 31 HCBS programs in ten states during 2014-2015 (n= 3,226). The survey asks Medicaid beneficiaries about their experiences with paid staff that support and/or provide care to them. The field test included two randomized experiments to test for potential mode and response option effects in individuals who have physical or mental disabilities. This presentation will include results from these experiments and the implications for administering surveys to individuals with disabilities including the potential of including proxy responses. This research will inform future research efforts to accurately conduct survey research with people with physical and mental challenges.

Collecting Data in Colonias Settlements: A Pilot Study in Hidalgo County, Texas

Catherine C. Haggerty, *NORC at the University of Chicago*
William Donner, University of Texas Rio Grande Valley
Katie Archambeau, *NORC at the University of Chicago*
Ned English, *NORC at the University of Chicago*
Colm O'Muircheartaigh, *NORC at the University of Chicago*

In late 2016 NORC and the University of Texas Rio Grande Valley partnered to conduct a pilot test with residents of Colonias in Hidalgo County, Texas. A Colonia is a makeshift border settlement often established in flood plain areas, without adequate government services and oversight, leaving residents particularly vulnerable to natural disasters and health risks. The study was intended to help Rio Grande Valley residents better prepare for and respond to the flood hazards they routinely experience in the region. In a two-stage sample selection process, we utilized purposive sampling to select Colonias and Address Based Sampling (ABS) to select housing units for interviewing. The sample included 12 Colonias

Session 7B: 2:00 pm – 3:00 pm
Findings from Research with Special Populations
Room: Auditorium 3

and we interviewed 96 households. Our presentation will include a description of our methodology, a profile of the Colonias residents, experiences with extreme weather events, their preparedness for hurricanes and floods, and the agreement and disagreement between the data residents reported and that provided in government records.

Turnover Intentions: Wishful thinking or realistic preview
Judah Frank, *Office of Personnel Management*
The relationship between turnover intention and turnover behavior among Federal employees was analyzed in this study. The Federal Employee Viewpoint Survey (FEVS) and the Enterprise Human Resources Integration (EHRI) were used as the sources of data. Results indicated that turnover intention had emergent properties, such that individual level analyses revealed an insignificant relationship between turnover intention and behavior, while group level analyses revealed a significant relationship. Additionally, individual level discrepancies between turnover intention and turnover behavior were shown in their respective relationships with other third variables, such as engagement, as measured in the FEVS through the Employment Engagement Index, and job satisfaction, as measured in the FEVS through the Global Satisfaction Index.

Session 7C: 2:00 pm – 3:00 pm
Mobile Technologies
Room: Conference Rooms 3 & 4
Chair: Melissa Therrien, *U.S. Census Bureau*

Do Fences Really Make Good Neighbors? A side-by-side comparison of RDD and Geofencing
James Dayton, *ICF International*
Matt Jans, *ICF International*

Traditional behavioral health survey research relies on respondent recall of past behaviors and experiences and their ability to predict future actions. Recall can be imperfect, especially over time. Intentions are also relatively poor predictors of actual behavior. Additionally, screening to identify respondents prone to engagement in a particular activity can be costly. Can we improve the accuracy of both types of health measures by asking respondents about their experiences in situ (e.g., at the doctor's office)? Can we know when respondents might be likely to engage in a particular behavior and ask them to report in that moment? How would the resulting data and statistical estimates compare to traditional survey methods? ICF tested this innovative survey approach using an app-based smartphone panel to provide panel members with a survey opportunity when entering a target location. Response rates, survey completion rates, key health estimates, and data quality indicators were compared among three data collection approaches: in-store, geo-fenced responses; web-based responses from an ABS and mobile panel frame; and the traditional CATI-based, risk-factor survey. In addition, we assessed respondent willingness to perform tasks in addition to answering survey questions, such as taking a

Session 7C: 2:00 pm – 3:00 pm
Mobile Technologies
Room: Conference Rooms 3 & 4

picture of alcohol product in-store signage. Findings may lead to a rethinking of methodological approaches to sampling and questionnaire design for public health surveys.

Survey Data Collection and Management Using Participant's Own Smartphones
Paresh Patel, *SurveySignal, LLC*

As social science researchers, program analysts, program managers, and market researchers, we need to know the why, when, and how of human experience. For nearly a century, researchers have used surveys to capture data. Today's advancement in technology gives us a powerful, new way to connect with survey participants/citizens/customers/employees wherever they are, during the natural course of their day. Ecological Momentary Assessment (EMA) method captures situational variation as it happens in real time. Experience sampling has become an increasingly popular method of survey data collection (structured and unstructured) in social and personality psychology, social science research, market research, and beyond. Ecological Momentary Intervention (EMI) method provides interventions to people during their everyday lives (i.e. in real time) and in natural settings (i.e. real world). With the ubiquity of smartphone ownership and recent technical advances, conducting research studies on participant's own devices have become increasingly easy to do. Signaling participants, collecting "in the moment" responses from them and delivering interventions can be done without installing software on their smartphones. In this presentation, we present an example of one reliable, user-friendly, highly customizable, and cost-effective solution that has been used for hundreds of EMA/EMI studies around the world. That example, SurveySignal, is a web-based application that generates Short Message Service (SMS), Multimedia Messaging Service (MMS) or Email messages as signals and reminders, according to fixed, random or mixed schedules, with interactive SMS Data Collection or links to mobile surveys designed with any widely used online survey systems including Qualtrics and SurveyMonkey. We describe the method and customizable parameters, and then present evaluation results from multiple studies conducted using SurveySignal. An experimental manipulation of the reminder signal in one study showed how reminder SMS messages led to a 10% increase in response rates.

Integrating Smartphone Apps into Small Regional Household Travel Surveys
Jane Shepherd, *Westat*
Marcelo G. Simas, *Westat*
Laura Wilson, *Westat*

The growing percentage of persons who own a smartphone in the United States combined with recent technological advances in terms of processing capacity and battery management have made it possible for small regional household travel surveys (HTS) to use them as an alternative to traditional GPS data loggers. This paper presents a brief review of the evolution of smartphone technology application in HTS and introduces a low-cost method for integrating smartphone apps into existing travel reporting modes in HTS, which is suitable for use in small regional HTSs. This method is demonstrated by presenting preliminary results from a recently concluded HTS in Billings, MT where all age-eligible household persons

Session 7C: 2:00 pm – 3:00 pm
Mobile Technologies
Room: Conference Rooms 3 & 4

were invited to download and use smartphone apps to capture and report their travel details. As a result, GPS, accelerometer and multi-day travel data were collected from 626 persons belonging to 513 households out of the 1,066 complete households in the final delivery sample. Summary statistics of the collected response and travel data are presented and discussed and the paper concludes with recommendations for future research. Keywords: Household Travel Surveys, Smartphone Apps, GPS

Session 8A: 3:15 pm – 4:15 pm
Networking Session: Survey Design and Privacy, Confidentiality, & Disclosure
Room: Auditorium 1 & 2

This interactive session provides an opportunity to network with innovators and colleagues to develop awareness, engage in professional collaboration, and share knowledge. This is a great way to meet others with similar interests and develop potential partnerships.
Topics: Survey Design
Privacy, Confidentiality, & Disclosure

Session 8B: 3:15 pm – 4:15 pm
Roundtable: What Makes a Good Interviewer? Metrics and Methods for Ensuring Data Quality
Room: Auditorium 3
Moderator: Matt Jans, *ICF International*
Panelists: Dave Roe, *Abt Associates*
Mary Penn, *ICF International*
Tammy Cook, *Westat*
Kelly Lynn, *RTI International*
Samantha Daugherty, *SSRS*
David Morgan, *U.S. Census Bureau*
Patty Maher, *Survey Research Center at the University of Michigan*

To ensure data quality, most survey organizations use multiple metrics and methods to a) hire the best interviewer workforce possible, and b) ensure that their interviewers perform to established quality and performance standards. Many factors affect the standards chosen and their implementation, including survey mode, sample design, and data collection resources available. For example, an organization might seek different interviewer qualities for face-to-face interviews than phone interviews. When monitoring the interviewers they've hired, organizations must create assessment protocols that reward productive behavior and correct unproductive behavior. Strong interviewers are often in short supply, creating a challenge for survey managers. If they set standards too high, few interviewers will meet them, creating a limited interviewer pool. If they create lower or more flexible standards to fill the interviewer pool, they risk data quality. There is no perfect balance. This facilitated roundtable discussion, which includes

Session 8B: 3:15 pm – 4:15 pm
Roundtable: What Makes a Good Interviewer? Metrics and Methods for Ensuring Data Quality

Room: Auditorium 3

leading survey companies, statistical agencies, and academic survey organizations, will address these challenges by reflecting on questions such as the following:

1) What metrics or standards do you use when hiring phone interviewers? Field interviewers? Please discuss both minimum criteria and higher, preferred standards (i.e., what does a "minimally-qualified" interviewer look like v. an "exceptionally-qualified one").
2) What metrics or standards do you use to monitor interviewers during production? Again, please discuss both minimum criteria and higher, preferred standards. a. What metrics do you use to monitor sample yield or unit nonresponse issues (e.g., refusal rates or cooperation rates)? b. What about metrics for measurement issues (e.g., reading pace, fluency)?
3) How do you incentivize and re-enforce positive behavior by interviewers and project management staff on these metrics? What management processes or mentoring/training techniques do you use?
4) For both phone and field, how have you developed your preferred set of metrics? Which metrics are based on scientific research v. arbitrary (organizational or client-based) standards v. "tried and true company standards"
5) How do you use CASIC technologies to monitor and intervene with interviewers (e.g., audio recordings, survey manager dashboards/reports, instant messaging)? What pros and cons have you experienced using any of these?

Session 8C: 3:15 pm – 4:15 pm
Networking Session: Big Data, Paradata, and Software & Development

Room: Conference Rooms 3 & 4

This interactive session provides an opportunity to network with innovators and colleagues to develop awareness, engage in professional collaboration, and share knowledge. This is a great way to meet others with similar interests and develop potential partnerships.

Topics: Big Data
Paradata
Technology, Software, & Development