FedCASIC 2019
April 16-17, 2019
Washington, DC

Federal Computer Assisted
Survey Information Collection Workshops
Program
Tuesday, April 16, 2019

Arrival and Registration: 8:00 am - 9:00 am
Central Reception Area

Plenary Session – Welcoming Remarks and Keynote Address: 9:00 am - 10:00 am
The Federal Data Strategy, the Evidence Act, and Statistical Activities
Rooms 1, 2, and 3
Nancy Potok, Chief Statistician of the United States, Office of Management and Budget

As we continue to modernize federal statistics, the data landscape continues to change rapidly. The Federal Data Strategy, the newly enacted Foundations of Evidence-based Policymaking Act, and other developments have created multiple opportunities for statistical agencies to add significantly to federal capacity to leverage data as a strategic asset.

Session 1A: 10:15 am - 11:30 am
Utilizing Paradata to Make Informed Decisions on Data Collection Procedures and Processes
Room 1
Chairs: Pam McGovern, National Agricultural Statistical Service
Chris Stringer, U.S. Census Bureau

The Role of Survey Paradata in a Federal Statistical Agency’s Commitment to Quality and Professional Standards of Practice
John Finamore, National Center for Science and Engineering Statistics
Jennifer Tancreto, U.S. Census Bureau
Rachel Horwitz, U.S. Census Bureau
Renee Reeves, U.S. Census Bureau

Federal statistical agencies, like the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation, have a principal function to serve as a clearinghouse for the collection, interpretation, analysis, and dissemination of objective data. The work and decisions at federal statistical agencies are guided by principles and practices that position the agency to provide relevant, accurate, timely, credible, and trustworthy statistical information that researchers, policymakers, and the public require (NASEM, 2017). One practice that is critical for the effective operation of a statistical agency is the commitment to quality and professional standards of practice. This presentation will provide an overview of the efforts that NCSES has taken to use paradata as part of this commitment.

Paradata are measurements related to the process of collecting survey data. These measurements can occur within any data collection mode (paper, telephone, face-to-face, or web) and with different levels of detail. Within a web survey, paradata can be collected on a variety of factors that describe a respondent’s interaction with the web instrument including completion time, breakoffs, changed answers, and navigation through the instrument. Within a telephone survey, paradata can be collected on specifics of the telephone data collection effort including the number of outbound call attempts. In this presentation, we present information on how NCSES has used web survey paradata for measurement error evaluation and respondent burden reduction. We also will present the results from our use of web and telephone survey paradata to inform data collection operations and adaptive design intervention decisions.
**Using Web Paradata to Understand and Improve the Data Collection Process**
Robyn Sirkis, National Agricultural Statistics Service
Pam McGovern, National Agricultural Statistics Service

Survey organizations are turning to paradata to help as a tool in understanding and improving the data collection process. Paradata, a term used to describe data about the data collection process itself, are collected at the respondent level. With an increasing focus on web instruments, the National Agricultural Statistics Service (NASS) began collecting and analyzing paradata for its web data collections. Paradata can give great insight into the response process in a self-administered instrument. In web surveys, paradata can be classified into several types: contact (date, time, and number of sessions), device-type (e.g., device type used, device orientation, and browser), and questionnaire navigation (e.g., change of answers, error triggers, breakoffs, and help access). This presentation will discuss results of a web paradata analysis conducted for the 2017 Census of Agriculture. In conjunction with the paradata, demographic data was also utilized to better understand the response process across respondents answering the Census via the web. The paradata analysis results are critical to understanding the quality of the web collection, informing improvements to the online instrument, and assessing the online data collection experience.

**Uses and Applications of Paradata in Web and Mixed-mode Surveys**
François Laflamme, Statistics Canada

The increasing use of Web and Mixed-mode surveys has provided a wide scope of new, ongoing and timely paradata. Over the past decade, many studies have demonstrated the usefulness of paradata in the data collection context while greatly improving the understanding of the complex and evolving data collection process. In particular, paradata is often used to identify and implement strategic opportunities for data collection improvements. At the same time, it is also recognized by many survey researchers that paradata can also be of great methodological use in many survey steps prior to, during, and after data collection. This presentation provides an overview of the use and applications of paradata to plan, assess, manage, monitor, and improve the survey data collection process for Web and mixed-mode surveys.

**Using Contact Data to Gain Operational Efficiencies**
Rachel Horwitz, U.S. Census Bureau
Renee Reeves, U.S. Census Bureau
Beth Newman, U.S. Census Bureau

Paradata are present in all data collection modes, though their uses may be different. In interviewer-administered surveys, paradata are often used to reduce nonresponse bias and to predict survey participation. Another common use is to gain efficiencies in data collection. Specifically, transaction data can be used to determine at what point contacts cease being productive. This has the potential to reduce the overall number of contacts and also give interviewers more time to focus on cases that are more likely to respond. This presentation provides two examples, one from a telephone survey and one from a face-to-face survey, where transaction data were used to demonstrate the potential to reduce the overall number of contacts without having a significant impact on response rates.
Did We Catch Your Eye? Using Eye Tracking to Inform Survey Design

Room 2
Chair: Jennifer Edgar, U.S. Bureau of Labor Statistics

Research in Practice: Using Eye-tracking to improve engagement, reach and customer satisfaction
Kath Straub, Usability.org
Sylvie Williams, Social Security Administration

Eye-tracking researchers uses information about where a person's eyes go (or fail to go) to draw inferences about readability or intuitiveness of a design or content. In this session, we will survey what eye-tracking captures, common (and uncommon) measures and metrics, and how eye-movement data is used to infer reader confidence, confusion and cognitive load. You will leave the talk with a more concrete and confident ability assess and use eye-tracking research to inform your web, content and survey design.

Optimal Placement of Instructions and Definitions in Survey Questionnaires
Heather Ridolfo, National Agricultural Statistics Service
David Biagas, U.S. Bureau of Labor Statistics

In establishment surveys, it is common to include detailed instructions and definitions with survey materials to ensure that all respondents are interpreting the survey constructs in the same way. Additionally, questions in establishment surveys are sometimes keywords reflecting the construct being measured followed by definitions (e.g., Intermediate Market (business or organization in the middle of the supply chain marketing locally – and/or regionally – branded products) (Haraldsen, 2013). However, it can be a challenge to get respondents to read the instructions/definitions. For example, Gernsbacher (1990) found that in the case of keywords followed by definitions, respondents tend to read the keywords and ignore the definitions that follow. Respondents are even less inclined to look for auxiliary information (e.g., instruction booklets, help screens) when responding, even if this information is necessary for answering the survey question (Conrad et al., 2006; Sloan and Ott, 2016).

The National Agricultural Statistics Service (NASS) often struggles with where to place instructions and definitions. When conducting cognitive interviews, respondents are usually asked if they read the instructions and definitions associated with a question. However, cognitive interviewing is not an ideal method for determining this because respondents are not very good at telling you what they have and have not read (Schall and Bergstrom 2013; Guan et al., 2006; Albert and Tedesco, 2010). In fact, it has been found that in cognitive interviews respondents will often answer yes to these types of probes, but upon further probing it becomes clear that they did not read the instructions/definitions; or if they did, they did not adhere to them.

A study was designed to determine the optimal placement of question instructions and definitions to increase the likelihood that they are attended to by respondents. Using Tobii eye tracking software, NASS examined four design features that are commonly used on various NASS survey questionnaires to determine which designs are most likely to be attended to by survey respondents. The four design features evaluated were: 1. Definitions preceding survey questions; 2. Definitions following survey questions; 3. Single banked include/exclude statements; and 4. Double banked include/exclude statements. In this presentation, we will present the results of this test.
Session 1B: 10:15 am - 11:30 am
Did We Catch Your Eye? Using Eye Tracking to Inform Survey Design
Room 2

Using Eye Tracking to Confirm Usability Issues with Forced-Choice Response Options and Branching Designs
Elizabeth Nichols, U.S. Census Bureau
Shelley Feuer, U.S. Census Bureau
Kathleen Kephart, U.S. Census Bureau

At the U.S. Census Bureau, eye tracking is one piece of data captured during usability testing of online household surveys. While eye tracking data are sometimes the primary metric when comparing A/B web page designs, the usefulness of the data is less clear when conducting typical usability testing of a survey. This presentation provides examples of when eye tracking data has provided quantitative evidence to complement observational data on respondent behavior during usability testing of online household surveys. In one example, we share how participants answered several forced-choice questions in a survey and then show how the eye tracking data matched our observations. In another example, we share the eye tracking data from two different skip sequence designs. While the observed behavior of the participants with the two designs was not that different, the eye tracking revealed possible subtle differences that need more investigation.

The Impact of Question Format on Reading and Response Behaviors
Jean Fox, U.S. Bureau of Labor Statistics

The format of survey questions can impact the response process as well as the responses themselves in a variety of ways. For example, previous research has found that respondents provide more affirmative responses to a forced-choice question (where respondents have to select “yes” or “no”) compared to a “mark all that apply” question. To provide additional insight in the impact of question format, this study used eye tracking to record participants’ eye movements as they answered survey questions. The study compared eye movements and response patterns with two sets of questions: (1) “yes/no” versus “mark all that apply” questions as well as (2) questions presented individually versus in a grid. This presentation will share the results of this study along with the potential implications for survey design.

Eye Tracking in Usability Testing
Silvia Salazar, National Cancer Institute
Jonathan Strohl, Fors Marsh Group

Eye-tracking technology is typically used during usability testing sessions at the National Cancer Institute (NCI) to analyze gaze patterns during interactions with digital products; however, the technology can also be used for observational purposes. In the spring of 2018, usability testing was conducted with healthcare professionals on three cancer risk assessment tools: breast cancer, colorectal cancer, and melanoma. Healthcare professionals were asked to input simulated patient information into the tools to calculate and assess cancer risk. During the testing sessions, we streamed the point of gaze of the participants to an observation room for stakeholders to view in real time. We will present this study’s usability testing findings and resulting recommendations and will discuss how eye tracking was used in this study for strictly observational purposes to increase stakeholder engagement.
Session 1C: 10:15 am - 11:30 am

Automatic Coding: Using Machine Learning to Structure Open-Ended Text Data
Room 3
Chair: Brandon Kopp, U.S. Bureau of Labor Statistics

Using Machine Learning to Autocode Products for Commodity Flow Survey
Mehdi Hashemipour, Bureau of Transportation Statistics*
Christian Moscardi, U.S. Census Bureau*

Since 1993, the Census Bureau and Bureau of Transportation Statistics (BTS) have partnered in conducting the Commodity Flow Survey (CFS), which provides important statistics about the movement of commodities in the United States. Every five years, tabulated estimates are generated for goods shipments, including the origin and destination, mode of transportation, tonnage and value of shipments, and average miles per shipment by commodity. These data are utilized by a wide number of users and stakeholders at all levels of government (federal, state, and local), as well as by transportation companies, consulting firms, and universities. In 2015, BTS and Census released the first public use microdata product consisting of approximately 4.5 million shipment records captured from the 2012 CFS.

To ease the respondent burden in the next CFS data collection and to improve current estimates, we used Machine Learning and Natural Language Processing (NLP) to analyze the commodity shipment text descriptions and automate the assignment of Standard Classification of Transported Goods (SCTG) codes to each shipment record. In this project, we used the bag-of-words model to vectorize documents and find the frequency occurrence of words. These vectorized words are used as features for training the final Machine Learning classifiers. According to the U.S. Census Bureau, using this model in the future CFS will approximately save respondents 50,000+ hours, and $2.1 million in respondent lookup costs.

Autocoding the Survey of Occupational Injuries and Illnesses – 5 years in
Alex Measure, U.S. Bureau of Labor Statistics

Five years ago BLS staff read and manually coded hundreds of thousands of written descriptions of work-related injuries and illnesses each year. Today, more than two thirds of these codes are now assigned by a deep neural network, which evaluations suggest, is substantially more accurate on average than trained human workers. We discuss how we addressed some of the many challenges inherent in this transition including how we built these new computer systems, how we evaluate their performance, how we decide when and how to use them, and how we monitor and maintain them to continually improve performance.

Automating Cause of Death Classification: Natural Language Processing in the Division of Vital Statistics
Patrick Drake, National Center for Health Statistics

Literal text on death certificates and hospital intake records can be a rich source of information for public health research and surveillance. Traditionally, processing literal text on vital statistics records is a manually intensive effort, but recent advances have been made to automate those processes. These natural language processing algorithms, once implemented, will improve the timeliness of data production, and reduce the resources needed to process mortality records. This talk will describe two recent efforts by the National Center for Health Statistics (NCHS) – the first project seeks to better quantify opioid involvement in infant deaths in the United States by identifying opioid drug use and exposure from the narrative description of the circumstances surrounding each death present on death certificates. The second project developed a system to automatically and accurately classify cause of death (based on ICD-10) from literal text for fetal death records.
Automatic Coding: Using Machine Learning to Structure Open-Ended Text Data
Room 3
Chair: Brandon Kopp, U.S. Bureau of Labor Statistics

Improving Estimates of Arrest Related Deaths with Machine Learning
Peter Baumgartner, RTI International
Jason Nance, RTI International
Duren Banks, RTI International
Erin Kennedy, RTI International

In 2015, the Bureau of Justice Statistics launched a pilot study evaluating a redesign of the Arrest Related Deaths (ARD) program, investigating a new approach that included identifying potential arrest related deaths from news articles. In this talk we’ll walk through how we used natural language processing, machine learning, and effective data processing to reduce incoming news article volume by 99% for manageable human review. This is a behind the scenes look at a project FiveThirtyEight declared one of the best data stories of 2016.
FedCASIC 2019 Workshops Program

Posters and Demonstrations: 12:30 pm - 1:00 pm
Central Reception Area

Use of Mobile Technology to Improve Quality and Efficiency in Field Surveys
Abie Reifer, Westat

Westat is beginning the fifth year of a program using mobile technology to increase efficiency and quality of field data collection. Our mFOS (Mobile Field Operations System) application provides convenient and easy-to-use functions for field interviewers to review current case information, quickly track work activities, and record contacts with participants. Additionally, mFOS features support a range of project needs by leveraging native mobile capabilities to improve data collection. Features have been added each year to meet evolving project needs as the technology is disseminated across the organization.

Applying Real World Eye-tracking to Paper-based Direct-to-consumer (DTC) Advertisements
Andrew Caporaso, Westat
Douglas Williams, Westat
Victoria Hoverman, Westat
Jennifer Crafts, Westat

Eye-tracking is frequently used as an empirical measure to assess attention to defined areas or features of visual fields. These range from survey forms to informational and marketing products. A common practice is to use fixed integrated eye-tracking device with a monitor to display the visual stimulus. This allows the researcher to control presentation of the stimulus, but diminishes real-world application for paper-based documents - especially those spanning multiple pages and requiring respondent interaction with the paper document. This presentation will report on the methodology and data quality for a pretest Westat conducted on behalf of the Food and Drug Administration (FDA). The objective of the pretest was to refine procedures for collecting eye-tracking measures of attention for comparison with responses to self-reported measures of information recall, recognition, and comprehension. Forty participants who had one of two medical conditions wore eye-tracking glasses while viewing a paper DTC advertisement related to their medical condition. Following the ad reading, participants answered a web questionnaire related to the advertisement and responded to debriefing probes about their experience with the eye-tracking equipment and survey. We will report on lessons learned with respect to achieving successful calibration and accurate eye tracking data. The findings will be informative for researchers interested in collecting eye-tracking data for real-world stimuli.

The Future of Accessing Census Bureau Data
Tyson Weister, U.S. Census Bureau
Zach Whitman, U.S. Census Bureau

Because the Census Bureau collects data for over 130 different programs disseminated through separate tools, it is easy for users to get lost when searching for the data they need. In response to this, the Bureau is developing a new platform to provide a streamlined experience to access all data and content. This vision in data dissemination is one where users will no longer need to know which tools to use - instead, they will simply start searching on census.gov to find statistics, maps, webpages, and other great Census Bureau products they need. In this session, we will demo the new interactive site on data.census.gov, designed as a preview for users to provide feedback. Attendees will explore the platform's latest tables, maps, and data visualizations in an easily digestible format, and learn how challenges in dissemination can help inform data collection and processing.
Posters and Demonstrations: 12:30 pm - 1:00 pm
Central Reception Area

Improvements to the Release of Economic Indicators
Willette Allen, U.S. Census Bureau
Mala Kline, U.S. Census Bureau
Adeline Tran, U.S. Census Bureau

The Census Bureau releases 13 principal economic indicators that are key to showing the health of America’s current economy and future economic performance. Business owners, researchers, investors, economists, and policymakers rely heavily upon these data to make decisions every day. Prior to the improvements, the indicators were not released until after the atomic clock showed the official release time and took up to thirty minutes for the data to be available. Meanwhile, the high value data was being released by other organizations before the Census Bureau could release it. The Census Bureau embarked on a project as an aspect of the Census Bureau's digital transformation to reduce the lag between the indicators' official release and when they are posted to the web by creating a custom application to meet consumers' demands. Since its deployment, we’ve had a 99.9% success rate with Economic Indicator data being available in as little as two milliseconds after the release allowing stakeholders to have more timely and equitable access. The new streamlined, automated method allows customers to access economic indicators on census.gov more expeditiously and efficiently by optimizing the process required to post economic indicator data to the Internet.

Mapping Connections: Visualizations of user behavior and question relationships in the American Housing Survey
Katie Gustafson, U.S. Census Bureau

The American Housing Survey (AHS) instrument software contains around 1400 questions, many of which are related to each other in ways that are not obvious. When changes are made to the instrument without awareness of these connections, errors in data collection or usability problems for interviewers may result. It is valuable to clarify these relationships in a way that can be useful for future waves of the survey. We used 2017 AHS paradata and instrument documentation to create data visualizations and charts identifying how interview questions and sections of the survey relate to each other, and pinpointed where in the instrument problems seemed to cluster. These visualizations were automated and applied to 2019 AHS instrument test data for analysis by the stakeholder branches during instrument testing and refinement.

Demonstration of Advanced Blaise Layout and Editing Features for Multimode Data Collection
Nikki Brown, Westat
Karen Robbins, Westat
Kathleen O’Reagan, Westat

Blaise 5 supports data collection across multiple modes from a single code base. This demonstration of a survey instrument in CAPI, CAWI and CADI modes will show how advanced layouts can be applied to the same Blaise instrument code operating on different devices (iOS, Android and Windows smartphones and tablets). Editing features in Blaise 5 enable users to develop customized editing layouts and move through the instrument for review and data editing.
Posters and Demonstrations: 12:30 pm - 1:00 pm
Central Reception Area

Using Google Analytics to Evaluate the Online Interaction between Survey Producer and Data User

Google Analytics (GA) is a tracking service that provides information about our users, how they get to the CE website, and their user behavior on the site. It reads the JavaScript that is written on each of our webpages which collects information about user's hardware, software, location, web navigation, and more, and then stores it on Google servers. BLS is able to query the database to view reports that use a 4% sample of our web traffic. BLS can also request unsampled reports from Google. BLS then uses the data generated by GA to make decisions on how to improve our website and web content. The Consumer Expenditure Surveys (CE) program uses GA to track web hits, pdf, Excel, and zip file downloads, and to track the most frequently viewed CE Beyond the Numbers (BTN) and Monthly Labor Review (MLR) articles published by the CE. This presentation will focus on the Google Analytics application that is used in the CE program to see what the CE top web hits are and web hits to specific CE web pages (Publication Tables, Public-Use Microdata, Gemini Redesign, and other web sites.) This will include a training session on how the GA application works and functionality of the CE GA quarterly report that tracks data users. This presentation will share information about how the CE Surveys Program uses Google Analytics to improve data user engagement, and will be of interest to other survey programs making data products available on their website.

Making Survey Data More Accessible Through Web Based Analytical Software
Ian S Thomas, RTI International

Nationally representative survey data, while remaining a gold standard and a rich source of information about the health and well-being of our citizens, is expensive and time-consuming to collect. Moreover, its full value can only be realized when made available to the largest audience - researchers, policymakers at local and national levels, journalists, and the general public. Accessibility issues surrounding survey data public use files (PUFs) and other forms of survey results, even when made available on government websites, can still create a serious barrier to knowledge and insight. In this talk we will discuss the obstacles to access and analysis of publicly-funded, population-based survey data and what can be done to make these data more accessible. We will show recent examples of our web-based survey data analysis tools and describe our approach to incorporating human-centered design and agile software development methodologies for developing these survey data products. Some of the most salient features include identifying target users, developing and refining the user interface for appeal to various audiences, conducting user experience testing, planning for accuracy and speed, and developing a proper feedback loop.

Integrating Administrative Records to Enhance Survey Responses
Juan Salazar, Databricks

Administrative Records (data already collected by the Government) come from a multitude of sources; other Gov't's, various Agencies, different databases...so each data set is going to come in a different format. Previously, it took an extensive effort to integrate this disparate data, in order to create Data Linkage between the records themselves, and individuals/entities that need to be matched up with each record. This is because a full data model needs to be designed up-front, so Gartner and the New York Times have reported that 60-80% of the time, effort, and funding will be spent during this data integration process - even before the data can be mined for results. The poster will offer an alternate method of integrating data, by ingesting the data as-is, and simply harmonizing only the data elements that are required for the specific survey - in an agile manner - therefore saving significant effort and time, so that analysis of the data can be done faster. Administrative Records can be protected by Title 13 and 26, plus other Government laws, so we will also show how the data is protected.
Administrative data have traditionally been collected and employed for uses such as administering and monitoring programs, providing for government transparency, tracking of service and benefit receipt, and meeting regulatory and reporting requirements. Recent technological advancements and a focus on less costly and time-consuming data collection methods have increased the public and private sectors’ use of administrative data for non-traditional uses, such as the evaluation or program impacts and predictive analytics. The capability of these advancements, efficiencies of administrative data, and accurate interpretation of analytic results are best realized if the data are properly collected, well documented, appropriately cleaned and integrated, and accessible to and correctly used by end users, such as program staff and external researchers. Drawing on data governance strategies and our experience collecting and processing administrative data for large-scale evaluations, we will present best practices for improving the quality of administrative data at five key stages of the data life cycle: (1) planning and preparation, (2) documentation and diagnostics, (3) cleaning and standardization, (4) integration, and (5) analysis and dissemination. These best practices can increase comfort with and trust of an entity’s administrative data and allow end users to experience the many benefits of having high-quality administrative data at their fingertips.

Collecting and monitoring data are key components of programs that provide social services to their communities. In addition to using data to conduct evaluations, assess performance, and track quality improvement, programs must often collect participant data as a condition of receiving federal grant funding. To address this growing need, Mathematica Policy Research developed RAPTER®, a scalable, cloud-based data system, to enable federal agencies and their grantees to consistently and effectively collect, track, and report participation data to inform program efficacy. RAPTER® is a secure web-based platform with the following customizable modular components: - Program enrollment - Random assignment - Participant and service tracking - Survey administration - Reporting This poster presentation will demonstrate RAPTER® instances that Mathematica is implementing on federal and foundation-sponsored evaluations using a range of capabilities and customizations. RAPTER® is NIST-compliant and FedRAMP-ready, and is compatible with Confirmit and other data collection systems. The evaluations that use RAPTER® include a randomized controlled trial of a program to improve employment outcomes for low-income individuals through targeted coaching and support. This presentation will demonstrate the flexibility and strength of Mathematica’s RATER® system to support consistent and effective data collection, analysis, and reporting across programs and over time.
A fundamental consideration of medical and public health research is ensuring that subjects understand the implications for their participation in the research and that they actively agree. §45 CFR 46 regulations require the inclusion of specific elements in the informed consent. Specifically, informed consent must begin with a concise description of the study purpose, voluntary nature of the study, why one would want to participate in the research, foreseeable risks and discomforts, benefits, alternative course of treatment (if applicable), duration, among other items. While informed consent is a standard and well-understood procedure, it presents some challenges in a study that uses a smartphone for data collection. That is, user experience is important, cognitive load can be extensive and often interrupted, and the size of a display screen may limit the amount of information shared. Moreover, new approaches to informed consent are needed with increasing use of multi-mode studies and the prevalence of smartphone usage. Therefore, ICF developed a standardized informed consent approach to ease the burden on research staff, to facilitate subject understanding, while meeting DHHS OHRP regulations. We present our approach to informed consent on smartphones, including seamless operation on differing devices and operating systems, the practical aspects of user experience, and the software development practices to ensure re-use and easy implementation between projects.
Session 2A: 1:00 pm - 2:15 pm
Top Three Management Challenges in Survey Technology and Programming Agencies and Organizations Face Today
Room 1
Chair: Karen Davis, RTI International

          Cheryl Lee, U.S. Energy Information Administration
          Tara Merry, Mathematica Policy Research
          Gina-Qian Cheung, Survey Research Center, Univ. of Michigan

Panelists will identify the top challenges facing their agencies or organizations today given the changing survey technology, data systems, and programming environments. Projects today often include innovative survey technologies, the use of specialized programming customizations, incorporated administrative and extant data sources, and the integration of different devices and technologies to support data collection. The panelists will discuss the ways that their organizations are dealing with the environmental changes that they have identified, and offer examples and lessons learned in addressing these challenges.

Session 2B: 1:00 pm - 2:15 pm
Considerations for Online Testing for Research at Federal Agencies
Room 2
Chair: Erica Yu, U.S. Bureau of Labor Statistics

          Paul Scanlon, National Center for Healthcare Statistics
          Jennifer Childs, U.S. Census Bureau
          Robert Sivinski, Office of Management and Budget

Online testing can offer 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. The four panelists in this session, who represent both the research and policy considerations of online testing at federal agencies, will answer your questions about how to conduct this research effectively and in compliance with federal oversight. The panel session will cover practical considerations for researchers both experienced and new on using these platforms at your own agency, including a discussion of when to use probability and non-probability panels and what regulations or restrictions may apply in different circumstances.
**FedCASIC 2019 Workshops Program**

**Session 2C: 1:00 pm - 2:15 pm**
**Administrative Records and Statistics**
Room 3

**Designing a Survey Tool to Determine When Administrative Data Can Substitute for Survey-Collect ed Data**
Rachel Carnahan, NORC, Univ. of Chicago
Andrea Mayfield, NORC, Univ. of Chicago
Felicia LeClere, NORC, Univ. of Chicago
Samantha Rosner, NORC, Univ. of Chicago

The Medicare Current Beneficiary Survey (MCBS) is a longitudinal panel survey of a nationally representative sample of the Medicare population. Data on facility-dwelling Medicare beneficiaries are collected in both Medicare certified facilities and other types of facilities. For certified facilities, the current data collection approach collects data that are also reported to the Centers for Medicare & Medicaid Services through the Long-Term Care Minimum Data Set (MDS) and the Certification and Survey Provider Enhanced Reports (CASPER). There is on average a two-month lag in posting MDS data, making MDS alone an unreliable source of administrative data for use during real-time data collection. The operational link between CASPER and MDS allows us to use CASPER as an indicator to identify whether MDS data will be available for a case once the MDS data are posted. This paper will provide an overview of how we will use CASPER as the source for a lookup tool, allowing us to identify in real-time whether a case will have administrative data available during post-processing. Once the lookup confirms the sampled person resides in a certified facility, we know that the case will have MDS data available. In these cases, 129 of 682 items can be skipped during data collection, resulting in a reduction in interview times of an average of 15 minutes. During post-processing, the administrative data will be incorporated with the survey data to substitute for the skipped items.

**Using Administrative Records to Produce Business Statistics: The nonemployer statistics by demographics series**
Adela Luque, U.S. Census Bureau
Renuka Bhaskar, U.S. Census Bureau
James Noon, U.S. Census Bureau
Kevin Rinz, U.S. Census Bureau
Victoria Udalova, U.S. Census Bureau

The quinquennial Survey of Business Owners (SBO) was the only comprehensive source of information on nonemployer and employer business demographics. To address increasing nonresponse rates and costs, and a rising demand for reliable, more frequent and timely data, the Census Bureau has consolidated three business surveys into a new survey, the Annual Business Survey. One of the consolidated surveys is the SBO. The nonemployer component of the SBO will now be accomplished through a new blended-data approach that leverages existing administrative (AR) and census records to assign demographic characteristics to the universe of nonemployer businesses, and produce an annual series that will become the only source of statistics of U.S. businesses with no paid employment by business owner demographics. This new series is starting to be known as the Nonemployer Statistics by Demographics series or NES-D. Demographics will include gender, race, Hispanic origin, veteran status, age and place of birth. Meeting the public's needs, NES-D will provide reliable estimates with no respondent burden on a more frequent and timely basis than the SBO. Using the 2015 vintage of nonemployer businesses and demographic information from the decennial census, the American Community Survey, the Census Numident and AR from the Department of Veteran Affairs, we discuss preliminary results on the initial stages of NES-D, the challenges encountered along the way, and next steps.
Session 2C: 1:00 pm - 2:15 pm

Administrative Records and Statistics
Room 3

Using Statistical Learning for Model Specification Selection
Steven Sawyer, U.S. Bureau of Labor Statistics
Alvin So, Formerly U.S. Bureau of Labor Statistics

Hedonic regression models are important tools used to capture technological change in price indexes. One challenge of developing hedonic models is variable selection. The choice of variables is important, because it determines the amount of price change the model will show. Traditionally, one way to select variables is to maximize the adjusted R² value of the model. However, the downside of this is that the model may perform poorly using data from a different sample. This poor performance may mean that the model is over-fitted. One way to overcome this problem is to use statistical learning. Statistical learning provides methods to evaluate a model by its ability to predict the dependent value of observations not used in the model, which is quantified by the mean squared error (MSE). Using the MSE to evaluate models allows the selection of a model that captures all of the relevant variables without being too specific to a particular dataset. This presentation will examine the use of statistical learning techniques to select a hedonic model for microprocessors, a product that sees rapid improvement that makes it difficult to quantify price change without a hedonic regression model. The statistical learning technique that will be focused on is repeated k-fold cross validation.

International Trade Data as a Source for Import and Export Price Indexes
Susan Fleck, U.S. Bureau of Labor Statistics

International trade transaction data is recorded by importers and exporters to conform to regulations that oversee customs and commerce, and individual records are required for each company for each product in every shipment by all modes of transportation. This rich administrative dataset is regularly used by the BLS International Price Program (IPP) as the sample frame for selection of representative products and companies that engage in trade. While price and quantity data are available, average prices by product area are subject to unit value bias, i.e., the change in price is influenced by changes in shipments, rather than changes in market conditions. IPP is researching the possibility of using the administrative trade data to create average prices for much more detailed break outs of company, product, and shipping routes, to determine whether the more detailed and greater number of prices track actual market conditions. The research has been carried out for two 'fairly homogenous' BEA end use indexes of dairy/eggs and vegetables. Statistical tests related to unit value bias provided valuable insights into price variability, across-month index value changes, and within-product price clustering and results are promising. The measures are based on a new methodology using the Tornqvist index formula at the lowest level of index calculation to account for substitution of goods and use of current period weights in order to be consistent with trade.
Session 3A: 2:25 pm - 3:25 pm
Software Advances
Room 1

Blaise 5 in Production
R. Suresh, RTI International
Emily Caron, RTI International
Lilia Filippenko, RTI International
Gil Rodriguez, RTI International
Mai Wickelgren, RTI International

Blaise 5 shows a lot of promise when it comes to the ultimate goal of survey software -- programming it once and deploying it on all modes and devices! After two years of intensive testing, RTI has rolled out Blaise 5 into production. This presentation will cover some projects where RTI has implemented Blaise 5 to show its varied features. These projects are implemented in a variety of environments, including an ACASI touch screen CAPI on Windows 10 laptops, a Web-based interview that requires upload of photos and possibly videos, and an iOS app with a lot of grid questions administered on iPads.

Rostering with Voxco
Matthew Bensen, RTI International
Nathan Sikes, RTI International
Jean Robinson, RTI International
David Schultz, RTI International

In developing Voxco survey instruments at RTI, programmers often employ a method of building the survey in Voxco Command Center and pushing the survey instrument to Voxco Online. This is essential when employing mixed-mode surveys involving CATI and web to keep the instruments consistent; additionally, organizations may choose to use the Command center as a repository for instruments. When pushing survey instruments with 'rostering' or 'looping' processes, the translation between Voxco Command Center and Voxco Online is not always straightforward nor does it consistently produce correct results in the translated instrument. This is not the only challenge with rostering in Voxco instruments even if the instrument is developed solely in Voxco Online. Voxco offers different methods to perform rostering and knowing which to use for the correct scenario is based on the programmer's best judgement and experience. Additionally, there are challenges to overcome when extracting data where instruments use rostering. This presentation will discuss: The pitfalls of the translation process between Voxco Command Center and Online with rostering and how to address these pitfalls; Different methods of rostering within Voxco Online and what works best for different instrument situations; The extraction of data from Voxco keeping the roster data in an understandable format and properly named variables so that data analysts can use the variables correctly. Examples will be provided.
Session 3A: 2:25 pm - 3:25 pm
Software Advances
Room 1

Web Screen Presentation Using Blaise 5
Michelle Amsbary, Westat
Rick Dulaney, Westat
Justin Kamens, Westat*
Joelle Michaels, U.S. Energy Information Administration

The Commercial Buildings Energy Consumption Survey (CBECS) periodically collects energy consumption and expenditure data from U.S. buildings. For several data collection cycles, dating back to 1999, CBECS has used Blaise for data collection. Beginning with the 2018 CBECS, the goal is to encourage respondents to complete their questionnaires on the web, which naturally suggests a move to Blaise 5. During the course of gathering requirements, the U.S. Energy Information Administration (EIA) and Westat identified several challenges around the conversion of a CAPI instrument to web. EIA and Westat initiated the development of best practices for web screen presentation on CBECS. We held seminars with methodologists, conducted a literature review, and surveyed publicly available common screen presentations, and from these activities we developed and refined a basic screen presentation template. We then identified representative CBECS questionnaire items for all item types and special situations. In addition to the core question types generally used throughout surveys - categorical, continuous, string, select all that apply, etc. - we also identified more complex types such as date pickers, grids, and lookups. We programmed these items in Blaise 5 for the web resulting in a web screen presentation that works well for the CBECS and generally for web data collection. In this presentation we will demonstrate key aspects of the CBECS web screen presentation.

Session 3B: 2:25 pm - 3:25 pm
Utilizing Online Research Panels for Federal Statistics
Room 2
Chair: Bob Torongo, Ipsos Public Affairs

Panelists: Larry Cohen, Strategic Business Insights
Nick Bertoni, American Trends Panel, Pew Research
Michael Lawrence, Ipsos Public Affairs
David Grant, American Life Panel, RAND Corp.

Discussant: Robert Sivinski, Office of Management and Budget

Survey researchers often struggle with the cost of collecting relevant representative survey data and the difficulty of reaching the population of interest. Emerging approaches, such as online statistically recruited survey research panels, pose a more cost-effective data collection method than ‘traditional’ survey methods. A high caliber probability recruited panel can can provide valuable insight into population characteristics. We will examine case studies, such as regulatory reporting, evaluating ROI or other high profile policy issues, where an online panel approach was utilized to satisfy data collection needs. We will discuss the planning, conceptualization, formation of research questions, design and implementation of online panels for statistical research. Our panel will help statistical research professionals learn strategies to improve their methods, approach, and implementation for delivering relevant statistics to their stakeholders and clients. Our methods can be applied across a wide range of industries and policy areas and will benefit both new and seasoned practitioners alike.
Session 3C: 2:25 pm - 3:25 pm
Administrative Records Policy Considerations
Room 3

Considerations Associated with Building, Marketing, Maintaining and Managing a Big Data Repository Project
Helen P. Ray, RTI International
Adam Weeks, RTI International
William Savage, RTI International

Data repositories offer advantages to researchers by providing a valuable resource in which they can answer new questions by using previously collected data. The considerations associated with building a big data repository present complexities and challenges from a project management perspective. Anticipating challenges can ease the stress and time it takes to make important decisions. During this session the participants will traverse through ten years of experience building and managing data repositories from a project management perspective. Using two Federally funded project examples, we will examine the most important considerations required for success. The exploration of these projects will include data issues, access constraints, metadata, outreach efforts and project management. An understanding of the audience for the repository and the intended goals of the resource are key points that will reverberate throughout the presentation.

The Future Is Now: How surveys can harness social media to address 21st-century challenges
Amelia Burke-Garcia, Westat
Brad Edwards, Westat
Ting Yan, Westat

Conducting survey research is increasingly difficult. Researchers face an ever more challenging combination of technological and social change and increasing costs -- making it more difficult to recruit sampled respondents initially and to retain them for panel studies. The extremely rapid evolution of communication and information technologies and the synergies between them are changing human social interaction and may ultimately render the current survey paradigm (researcher-centered ‘structured conversation’ with a willing randomly-selected respondent) outmoded. We envision a new survey world - that uses multiple data sources, multiple modes, and multiple frames (Lyberg & Stukel). It leverages social media to make survey research more people-centered, efficient, and relevant. While early research is promising, a comprehensive review of the ways social media can support, even transform, surveys is lacking. As the whole world is now getting hard to survey, we leverage Tourangeau’s framework of defining difficult-to-survey populations to explore social media’s potential for changing survey research across the survey life cycle stages: sampling, identifying, reaching, persuading, and interviewing. Real life case studies will be discussed including how social media may: (1) be a source for research panels; (2) help pinpoint data to replace or supplement self-report; (3) augment or replace survey data; and (4) reach and persuade survey respondents to participate.
Session 3C: 2:25 pm - 3:25 pm
Administrative Records Policy Considerations
Room 3

Practical Evaluation of Proposals for Integration of Multiple Data Sources
John L. Eltinge, U.S. Census Bureau

In recent years, governmental, academic and private-sector organizations have expressed interest in the development and use of statistical information products and services that are based on the integration of multiple data sources. These often include both sample surveys and non-survey sources (sometimes described as “non-designed,” “organic” or “big” data), e.g., administrative or commercial records, images and social media traces. Practical production-level work in this area is generally acknowledged to require objective evaluation criteria that can be communicated to, and used by, a wide range of stakeholders. This paper outlines some criteria for these purposes. Following a brief review of applicable literature, seven areas receive principal emphasis:

1. Clear statements of inferential goals and related quality requirements aligned with well-defined stakeholder needs.
2. Extensions of traditional “total survey error” models to evaluate the predominant components of variability that are important for evaluation of the quantitative dimensions of data quality.
3. Criteria for assessment of non-quantitative dimensions of data quality like relevance and timeliness.
4. Criteria for management of notable dimensions of risk arising from, e.g., prospective loss or degradation of one or more data sources; problems with design, implementation and maintenance of production systems; and disclosure risks.
5. Realistic assessment of the dominant fixed and variable components of cost of each of the steps in the design, implementation, maintenance and improvement of the production process.
6. Demonstrated capabilities of the organizations and personnel assigned to each of the primary tasks required for production of the anticipated statistical information on a high-quality, sustainable and cost-effective basis.
7. Current limitations on the information available for each of (1)-(6), and practical options for improvement of that information in high-priority areas.

Two running examples illustrate some of the main ideas developed here. The paper closes with a proposed checklist that includes specific sets of questions aligned with each of areas (1)-(7).

Session 4A: 3:35 pm - 4:50 pm
Monitoring Data Collection
Room 1

Response Rate Alternatives in Data Collection Monitoring
Kevin Tolliver, U.S. Census Bureau

With response rates declining in demographic surveys (Brick & Williams, 2013) and response rates alone not providing enough evidence low nonresponse bias (Groves & Heeringa 2006; Groves & Peytcheva 2008), there is a desire to observe real-time data collection monitoring metrics that are to response rates. The R-indicator (Schouten et al., 2009; Schouten et al., 2011) and coefficient of variation are two such measures that can be more indicative of nonresponse bias issues. We have developed a tool that uses Demographic Surveys Sample Database information and U.S. Planning Database, which includes 5-year American Community Survey block-group level estimates, as auxiliary data used to assess representativeness of data throughout data collection. This presentation illustrates some different uses on real data and discusses how this can be used for various demographic surveys.
For the 2017-18 data collection cycle, the National Teacher and Principal Survey (NTPS) explored the use of unconditional incentives in an effort to increase teacher response rates and overall sample balance. The teacher incentives experiment was designed to test the effectiveness of incentives sent to teachers and/or school-level coordinators. The experiment was conducted in two phases, based on whether or not a school returned the Teacher Listing Form (TLF) early. NTPS samples more than one teacher in a school, and the coordinator is a school staff member who helps get the questionnaires to teachers. The teacher sampling process for NTPS is a two-stage design and is conducted on a flow basis. First, a school coordinator from a sampled school must return the TLF, which includes the names and subject matters for every teacher within the school. From the TLF, teachers are then sampled weekly for the teacher questionnaire. In an effort to equally disperse the teachers across all treatment groups, logistic modeling was used to create response propensity models to predict the likelihood that a school will return the TLF, which allows for teachers to then be sampled from that school. Results from the prior cycle of NTPS were used to inform the models. The models will utilize information known prior to data collection to predict which schools are likely to return the TLF. Based on the results of these models, schools were assigned to eight incentive treatment groups.

Callback appointments are a common occurrence in telephone surveys when the sampled respondent is unavailable or requests another time to participate. In the general population, text message reminders have shown promise in helping people meet scheduled medical appointments (Talonen, Aistrich, & Borodulin, 2014; Gurol-Urganci et al., 2013). Applied to survey research, appointment reminders may increase the likelihood of keeping the original appointment, and benefit response rates. To test this hypothesis, scheduled callback records were randomly assigned to one of three treatment groups: (1) callback appointments were scheduled according to the set time, (2) callback appointment scheduled for a set time and a text message reminder sent in advance of the appointment, (3) same treatment as group 2 and, if an interviewer was unavailable, a text message sent that a new appointment was necessary. We examine the effectiveness of the text message intervention using two metrics: a) meeting scheduled callback times and b) survey response rates. We hypothesize that communicating with the respondent at the agreed time (even in the absence of completing the survey) and offering the opportunity to schedule a new time maintains engagement and benefits response rates. Our analysis adds to the body of research on meeting callback appointments and its effect on survey response rates.
Session 4A: 3:35 pm - 4:50 pm
 Monitoring Data Collection
 Room 1

Balancing Sample Size and Cost: An innovative solution in the absence of sampling mechanisms for IVR surveys
Kristen Flaherty, ICF*
Kelli Keith, ICF*
Adam Lee, ICF

Federal survey clients always want the largest sample size possible for a given cost. However, with large sample sizes comes the growing cost of processing every survey response, especially with technological sampling limits. The federal client sought to collect a defined number of completed interviews from respondents, while still offering the survey to every person calling into or receiving calls from four different customer service call centers. ICF designed and hosted the IVR survey, but these call centers were solely responsible for forwarding participants to the survey, and they lacked the technological capabilities to randomly flag calls for an interview. With a much higher than anticipated number of completed surveys, we developed an approach that allows all participants to be offered a survey, but randomly routes a percentage of participants to a shorter version of the survey. This provides the client with more customer satisfaction data than if we had to limit the number of participants that could take part in the interview, and lowers the cost overall. This paper presents the implementation of an innovative approach to controlling the selection of calls for an interactive voice response (IVR) customer satisfaction survey in the absence of a sampling mechanism and the tradeoffs of alternate approaches.

Session 4B: 3:35 pm - 4:50 pm
A Focus on Data Quality
Room 2
Chair: Catherine Haggerty, NORC, Univ. of Chicago

A Delicate Transaction: The Data Quality Program for the 2016 and 2019 Survey of Consumer Finances (SCF)
Kate Bachtell, NORC, Univ. of Chicago
Catherine Haggerty, NORC, Univ. of Chicago
Shannon Nelson, NORC, Univ. of Chicago
Richard Windle, Board of Governors of the Federal Reserve System
Joanne Hsu, Board of Governors of the Federal Reserve System

Through unwavering commitment to improvement and decades of refinement, the quality program for the Survey of Consumer Finances (SCF) is regarded as one of the most rigorous in social science research. The program for this triennial survey has been developed to overcome the challenges associated with collecting complex financial data from households that cannot be gleaned from administrative data alone. It is a delicate transaction. In this paper, we first describe the core features of the SCF data quality plan: validation, monitoring, evaluation, feedback, and continuous learning. We then share select findings from the 2016 SCF. During the 2016 SCF, monitoring of missing and/or invalid values (“don’t know,” “refused,” etc.) among dollar variables was expanded. We use multivariate regression modeling to identify the strongest predictor of “missingness” among dollar variables and discuss the results. Finally, we share plans for improving the data quality program for the 2019 SCF. Our findings and discussion may be of interest to survey practitioners, federal statistical agencies, and users of financial data.
Session 4B: 3:35 pm - 4:50 pm
A Focus on Data Quality
Room 2

Respondent Use of Records and Impact on Data Quality
Shannon Nelson, NORC, Univ. of Chicago
Catherine Haggerty, NORC, Univ. of Chicago
Kate Bachtell, NORC, Univ. of Chicago
Kevin Moore, Board of Governors of the Federal Reserve System
Jesse Bricker, Board of Governors of the Federal Reserve System

The Survey of Consumer Finances (SCF) encourages respondents to reference financial records throughout the interview so as to ensure that the most accurate and precise data are reported. Respondents are asked to use a variety of records, such as tax records, those they receive about their bank accounts, retirement accounts, and brokerage accounts and about accounts in other financial institutions. Using data from past rounds of the SCF, we explore the characteristics of respondents who use vs. do not use financial records to complete the interview and examine the effects of record use on data quality. We address the following research questions: Are respondents with more complex finances more likely to use records? Are early responders, who cooperate with requests to complete the survey more readily than late responders, more likely to use records? Do the numbers reported by those who use records appear more precise, that is, are they less likely to be rounded? Are there fewer “don’t know” responses when records are used? Are respondents who are reported by interviewers to be less interested in, or more suspicious of, the study more or less likely to use records? These findings will be of interest to those who collect financial data and may inform strategies for interacting with respondents and their use of records.

Designing a Web Survey for the 2019 SCF
Lisa Lee, NORC, Univ. of Chicago
Shannon Nelson, NORC, Univ. of Chicago
Catherine Haggerty, NORC, Univ. of Chicago

The Survey of Consumer Finances (SCF) is the premier source of information on the financial circumstances of American households. Conducted every three years, the SCF has traditionally been fielded as an interviewer-administered CAPI survey. For 2019, the Federal Reserve Board (FRB) has added the SCF Web Test, a feasibility test of collecting a subset of the SCF data via web to be conducted with an oversample of 2500 households in the AP sample. The SCF collects personal financial data that is both complex and sensitive. Since the FRB relies heavily on SCF data for trend analyses, the use of the web mode must be evaluated carefully due to the need for comparability across modes.

Items from several sections of the SCF were selected to be included in the web survey, with the remaining sections to be administered in a follow-up CAPI interview. In designing the web instrument, an important consideration was mirroring the rigorous approach to collecting dollar amounts that is implemented in the CAPI survey. The web survey underwent cognitive and usability testing. Then in a field pretest, respondents first completed the web survey independently and then participated in a CAPI interview to complete the remaining sections of the SCF. In this presentation we will describe the process of designing and testing the web instrument. We will also discuss lessons learned, preliminary findings from the fielding of the 2019 SCF Web Test, and initial thoughts on the feasibility of including a web component in future rounds of the SCF.
Session 4B: 3:35 pm - 4:50 pm
A Focus on Data Quality
Room 2

Data Quality through the Lens of Field Staff: Identifying Data Problems and Improving Data Quality in the Survey of Consumer Finances
Micah Sjoblom, NORC, Univ of Chicago
Catherine Haggerty, NORC, Univ of Chicago
Kevin Moore, Board of Governors of the Federal Reserve System

The Survey of Consumer Finances (SCF) delves into every financial detail of a household’s finances, therefore, maximizing data quality is a constant challenge. It is important to deliver feedback to interviewers about the quality of their work, especially to interviewers new to the SCF. As is the SCF tradition NORC reviews all call record entries, comments interviewers record during the interview, and debriefing notes interviewers are required to record after the interview. We also measure the number of missing values, the number of responses given as a range as opposed to a precise numeric value, and we look for interviews potentially conducted with the wrong respondent. In addition, economists at the Board of Governors of the Federal Reserve closely examine all of the substantive data and provide feedback after each weekly data delivery. Data quality issues or data errors are systematically reported in a timely fashion. Supervisors develop remedial action which includes one-on-one, small group, or self-directed electronic training. This research will examine two aspects of the data quality/ improvement process: Interviewer assessment of the improvement process and whether or not the interviewers improve their performance over time.

We will describe our methodology for upcoming qualitative and quantitative examination of the data to evaluate the outcome of our data quality process and data improvement over time.

Session 4C: 3:35 pm - 4:50 pm
Management Challenges related to Staffing, Recruiting and Retention focusing on Data Science and Artificial Intelligence
Room 3

Management Challenges Related to Staffing, Recruiting and Retention Focusing on Data Science and Artificial Intelligence
Chair: Jane Shepherd, Westat

Panelists: Kristen Monaco, U.S. Bureau of Labor Statistics
Anup Mathur, U.S. Census Bureau
Nora Paxton, Mathematica Policy Research
Karen Davis, RTI International
Gonzalo Rivero, Westat

This panel will discuss the challenges of recruiting, retention, and staff development for those with highly sought after skills in areas such as data science and artificial intelligence. Panelists will explore strategic approaches, best practices, and examples of ways organizations have addressed staff development, training, engagement and retention for individuals with these in-demand skills and capabilities.
Wednesday, April 17, 2019

Arrival and Registration: 8:00 am - 9:00 am
Central Reception Area

Session 5A: 9:00 am - 10:15 am
Improving Field Operations
Room 1

On Using Cognitive Computing and Machine Learning Tools to Improve Call Center Quality Control
Lew Berman, ICF
John Boyle, ICF
Don Allen, ICF
Josh Duell, ICF
Matt Jans, ICF
Ronaldo Iachan, ICF

Total survey error explicates sampling and non-sampling error as two primary components. In telephone surveys, non-sampling error has several sources including measurement error (ME). ME occurs when the method of collecting data is affected by the call center agent, the respondent, or the questionnaire, and the interactions between them. ME may in part be reduced thru quality assurance (QA) and quality control (QC) procedures. We define QC to be those procedures used after field work has begun to improve data quality. This includes expert staff manually reviewing agent conducted telephone interviews for effective speech rate, properly reading a question, and accurately recording a response. General industry practice is to review 5-10% of all calls. However, these practices are labor intensive, subjective, and the evidence for this range of review is anecdotal. Thus, this paper will focus on automating call center QC. Cognitive computing (CC) and Machine Learning (ML) generally attempt to emulate human problem solving using clever algorithms, logical reasoning, signal processing, or statistical methods. We are experimenting with the Microsoft CC and ML tools to automate call center QC. Under certain conditions, these tools may lead to increasing the volume of call review, objective measures of agent performance, and lower costs. We describe our experimental approach, the tools used, and results of our initial experiments.
Session 5A: 9:00 am - 10:15 am
Improving Field Operations
Room 1

Re-Engineered Field Operations Powered by a Unified Platform
TJ Oleksiak, Pegasystems
Sherif Elshayeb, Pegasystems
Zack Rudman, Pegasystems
Maks Takele, Pegasystems
Michael Hemmer, Pegasystems
Raghu Govindaraj, Pegasystems
Thomas Kaufmann, Pegasystems

The 2020 Census Operational Plan includes a goal of re-engineering field operations to efficiently and effectively manage the 2020 Census fieldwork, and as a result, reduce the staffing, infrastructure, and brick and mortar footprint required for the 2020 Census. The Pega Platform is enabling Census to realize this goal through the delivery of automation. Pegasystems has been working with Census to provide fieldworkers with the capability to work completely remotely and perform all administrative and data collection tasks directly from a handheld device. Supervisors will also be able to work remotely and communicate with their staff via these devices. Together, the Field Operational Control System (FOCS) and Enumeration (ENUM) mobile application deliver automation that enables significant changes to how cases are assigned, field enumeration is performed and the supervision of field staff. This talk introduces these innovations and their impacts for 2020 and beyond.

As demonstrated in the 2018 End to End Test, enumerator productivity increased with the use of these technologies and automated processes. As stated at a 2020 Census Program Management Review (PMR) by Al Fontenot, Associate Director of Decennial Census Programs, “We’ve now successfully tested these technological advances that will allow for a more efficient use of our workforce for the 2020 Census. In the 2018 Census Test, we were able to complete 1.56 cases per hour worked, compared to 1.05 in the 2010 Census.”

Text messages to field interviewers
Amanda Nagle, U.S. Census Bureau

In the Survey of Income and Program Participation (SIPP) 2019 data collection, text messages will be sent to interviewers to test if text message communications can nudge interviewers to be compliant with data collection instructions. The text messages will facilitate communications between headquarters and the interviewers and could boost compliance while reducing effort expended by field supervisors. We will test two message styles and 3 message frequencies. We will be able to understand the type and intensity of message that is effective for different types of respondents. This is an innovative approach to communicating with interviewers and we are not aware of any other federal agencies or surveys sending text messages to their interviewers. This presentation will discuss why SIPP and its interviewers are a good candidate for text communications, the content of the text messages, the mechanisms used to send the messages, the current state of the project, and its limitations.
Security in Corporate Networks
Eugene M Farrelly, RTI International

Security is a top priority in the design of any survey system and ultimately those systems are maintained by users in a corporate environment. How do you extend security at the organization level? Bring-your-own-device (BYOD) programs provide undeniable flexibility to employees and increased productivity. How do you balance those benefits against the sobering reality of a technological climate of increased risks of data breach? We'll describe how we are taking steps to further secure our corporate networks and modify BYOD programs to leverage new technologies that don't sacrifice flexibility and still ensure organizational level security.

Challenges and Opportunities in Using Data for Public Good while Protecting Privacy and Data Confidentiality: A National Statistician's quality review
Gentiana D. Roarson, Office for National Statistics*
Iain Dove, Office for National Statistics*

New legislation enables government statisticians in the UK to better share and use data across organizational boundaries, at a time of dramatic increase in the volume of data available from a wide range of sources. This has brought about major changes to the way organizations process personal data, encouraging greater transparency and accountability. On the other hand, against a backdrop of increasingly sophisticated attacks being developed by intruders and mounting public concern over data breaches, the data revolution presents significant new challenges to protecting privacy and confidentiality. It is vital that the statistical community understands and addresses these evolving challenges to provide a solid foundation for innovation to take place and draws upon the full range of expertise in this fast-developing field. These developments present an unprecedented opportunity for the Government Statistical Service in the UK to innovate with data, while safeguarding privacy and fostering public trust. The Office for National Statistics joined forces with leading experts in privacy and data confidentiality to explore the latest advances in these methods and to set out the state-of-play in using data for public good while ensuring data privacy. This paper outlines the challenges faced by statisticians as well as recommendations for next steps and further research needed to take full advantage of the new opportunities created by technological developments.
Session 5B: 9:00 am - 10:15 am
Data confidentiality
Room 2

Proposed Model for Tailoring Confidentiality Information
Heather Ridolfo, National Agricultural Statistics Service
Rebecca L. Morrison, National Center for Science and Engineering Statistics
Stephanie Willson, National Center for Health Statistics
Cleo Redline, National Center for Education Statistics
Casey Eggleston, U.S. Census Bureau
Jennifer Hunter Childs, U.S. Census Bureau
Jacob Bournazian, U.S. Energy Information Administration*

Data protection statements for surveys are commonly drafted based on concepts and assumptions relating to trust and privacy. Agencies with statistical collections strive to assure respondents that their data will be protected from unauthorized access and re-identification in order to satisfy privacy concerns. A common issue in drafting confidentiality statements is how much information to provide to respondents to satisfy their privacy concerns. A group of U.S. federal agencies developed a model for assessing survey respondent reactions to different confidentiality pledges and analyzed the amount of information to provide to satisfy respondents’ concerns about the confidentiality of the information they provide. Themes in the research data revealed that providing simple and understandable confidentiality language assures some respondents, while others need additional information. On the other hand, providing too much information in the beginning of an interview sometimes raises additional privacy concerns rather than assuage those concerns. Following the pattern of reactions to the proposed data protection statements and the types of concerns that respondents expressed, the theoretical model provides a framework for tailoring confidentiality information to individual respondent concerns. The model provides a useful approach for developing data protection statements that are appropriately tailored to the subject matter and data collection process. This presentation is based on the paper Proposed Model for Tailoring Confidentiality Information.
Session 5C: 9:00 am - 10:15 am
Machine Learning Applications for Surveys and Beyond
Room 3

Deep Learning for Development of Household List Frames in Low and Middle Income Countries
Robert Chew, RTI International
Kasey Jones, RTI International
Jennifer Unangst, RTI International
James Cajka, RTI International
Justine Allpress, RTI International
Safaa Amer, RTI International
Karol Krotki, RTI International

While nationally representative surveys are a rich source of data for indicators of sustainable development, it is costly and time consuming to build the household sampling frames that support these studies. This is particularly problematic in low and middle income countries (LMICs), where demand on surveys to respond to donors requirements are high and household census information is often severely out-of-date. In this case study, we explore the use of deep learning object detection models to identify and enumerate individual buildings directly from satellite imagery in the Kaduna state of Nigeria. In the test set, our model attained a mean average precision (mAP) of 0.48 for detecting structures, with relatively higher values in areas with lower building density (mAP = 0.65). Furthermore, when model predictions were compared against recent household listings from the 2017 Alive and Thrive Nigeria survey, the predictions showed high correlation with household coverage (Pearson = 0.70; Spearman = 0.81). With the need to produce comparable, scalable indicators of sustainable development in LMICs, these methods could accelerate production by helping to develop timely household frames.

Using Computer Vision to Process Vehicle Dashboard Displays in Transportation Safety
Kristin Jiating Chen, Westat
Alexander Cates, Westat
Rick Huey, Westat
Marcelo Simas, Westat
James Jenness, Westat
Gonzalo Rivero, Westat

This presentation will focus on the methodological aspects of a pilot project aimed at extracting data from video recordings of vehicle dashboard displays in order to understand driver behavior in the context of driver-assist technologies. More specifically, we will show how we used computer vision techniques in order to identify the presence of icons from video recordings of the central dashboard display of instrument vehicles. This information allowed the research team to know when driver-assist technologies were used by the driver and then correlate these data with other parameters such as vehicle speed, time of day, location, road functional class, etc. To accomplish this, we developed a machine learning pipeline using R and Python which utilized OpenCV (an open-sourced computer vision engine) to perform frame extraction from videos, image pre-processing (i.e., perspective transformation and denoising), and feature matching between a reference icon set and the extracted frames. Results of the application of this pipeline were monitored using an interactive tool that allows the user to visualize the results and modify the parameters of the feature matching system. Results on a small test set are promising although additional research is needed to reduce the number of false positives.
Session 5C: 9:00 am - 10:15 am
Machine Learning Applications for Surveys and Beyond
Room 3

**Application of Machine Learning and Remote Sensing in Analyzing U.S Port Performance**
Mehdi Hashemipour, U.S. Department of Transportation
Julie Parker, U.S. Department of Transportation

With approximately 360 commercial U.S ports, documenting freight movement and port performance can be costly and challenging. The survey-based programs conducted by the transportation agencies to collect information such as the number of containers loading daily, amount of good stowed, number of trucks in a queue, type of cranes, yard equipment, etc. in each commercial port. By relying only on survey crew collected data, it will be difficult to monitor the freight movement and collect the port performance in a time series manner and the gathered data will be inconsistent. Remote sensing imagery combined with Machine Learning algorithms could support the reduction of these inconsistencies. Machine Learning offers the potential for effective and efficient classification of spatial imagery. The Machine learning algorithms can handle data of high dimensionality and to map classes with complex characteristics. In our project, we utilize different Machine Learning and Deep Learning (including a variety of algorithms such as ANN, SVM, SOM, and DT) to classify the ground objects existing in a case study port. Then, we apply the trained classifier as a baseline model to satellite imageries of all U.S ports to collect the desired data. The combination of using a time series satellite imagery and Machine Learning in a system help to detect spatial patterns changing over time and find the significant hot spots, anomalies, and outliers in any selected U.S port.

Session 6A: 10:25 am - 11:40 am
Incentives and Respondent Engagement
Room 1

**Implementing Non-Monetary Incentives in the National Sample Survey of Registered Nurses**
Renee Stepler, U.S. Census Bureau
Daniel Eklund, U.S. Census Bureau
Margaret Behlen, U.S. Census Bureau
Elizabeth Sinclair, U.S. Census Bureau

Using non-monetary incentives can present an attractive alternative to cash incentives by lowering the cost of the incentives, avoiding security risks associated with handling monetary incentives, and providing something engaging for the target sample. During 2018, the U.S. Census Bureau conducted the National Sample Survey of Registered Nurses (NSSRN) on behalf of the Health Resources and Administration’s (HRSA) National Center for Health Workforce Analysis (NCHWA) using non-monetary incentives to test their efficacy in increasing response rates, reducing follow-up costs, and decreasing non-response bias. Pens and lanyards were used as incentives since nurses indicated during cognitive interviewing that these items would be useful for everyday work activities. However, non-monetary incentives present their own set of challenges ranging from acquisition of the incentives to ensuring packaging is appropriate for delivering the incentives unharmed. This presentation provides a brief summary of the impacts of the non-monetary incentives, details processes taken to effectively implement the incentives for the 2018 survey, and presents lessons-learned for future federal surveys.
PayPal? An Incentive to Check Out?

Jeff Franklin, RTI International
Alyson Miller, RTI International
James Pruitt, RTI International
Patrick Tucker, RTI International
Jennifer Wine, RTI International

The National Center of Education Statistics National Postsecondary Student Aid Study (NPSAS) is a large-scale, nationally representative study of students enrolled in postsecondary education. Like many other studies, NPSAS utilizes incentives as part of the strategy to motivate survey participation, improve response rates, and minimize nonresponse bias. In 2015, to match the rise in mobile device usage and the digitization of money, the NPSAS team sought a means to offer a more convenient and immediate incentive option to sample members as a thanks for completing the NPSAS survey. PayPal, with 254 million registered users and an annual payment volume of over $450 billion in 2017, was selected due to its popularity and name recognition in the U.S. Since integrating PayPal's application program interface (API) into our survey engine and offering PayPal incentive payments in 2015, we have continued to fine-tune our processes through four additional rounds of data collection on longitudinal studies related to NPSAS. In this presentation, we will share our experience gained in establishing PayPal as an incentive option. We will describe the various options for providing PayPal incentives, the user’s experience of receiving an incentive via PayPal, and procedures we’ve developed for tracking and reconciling payments. In addition, we will share lessons learned and the unanticipated challenges we faced along the way.

National Health and Nutrition Examination Survey (NHANES): Using text reminders to encourage fasting for medical exams

Sheryl A.B. Wood, Westat
Tatiana Nwankwo, National Center for Health Statistics

Texting is very common and an increasing mode of communication. Sending text reminders to study respondents is relatively easy and inexpensive. In May of 2011, the National Health and Nutrition Examination Survey (NHANES) set up text reminders to encourage participants with morning exams to arrive having fasted. Fasting is required to administer the Oral Glucose Tolerance Test and to accurately test cholesterol. This presentation summarizes the survey’s experience with these text prompts over the past 6 years. NHANES is an address based sample with in-person screening for eligibility. Participation includes an in-home interview plus a 4-hour exam at a Mobile Exam Center (MEC). Text is a preferred way to remind people about an appointment, and is used increasingly for this purpose by health care and personal care providers. In May of 2011, fasting text reminders were added to the NHANES protocol. While we cannot say that the texting is the direct cause for higher fasting rates in the group that receives texts, as those giving permission to text may be a more compliant group, we have seen the rates stay steady and it has not had a detrimental effect. Text reminders continue to be an inexpensive way to remind participants of appointments and other study participation requirements. Data for different age, gender and race/ethnic groups and by year are also available and will be included in the presentation if there are interesting findings.
Text Interactions on an RDD Study Using an SMS-Enabled Outbound Number
Thomas Brassell, ICF
Josh Duell, ICF
Randal ZuWallack, ICF
Matt Jans, ICF

The proliferation of mobile devices has been a central focus of survey research over the past decade. Random digit dial (RDD) telephone studies increasingly implement dual-frame sample designs, web surveys are mobile optimized, and some researchers are exploring the utility of administering entire surveys via short message service (SMS; text messaging). Our study explores the value of using an SMS-enabled outbound number in a nationwide telephone health survey to engage sampled respondents that text the outbound number after receiving a telephone call. We collected approximately 900 of these text replies during our first wave of data collection. Our research finds that the messages contain over 15 times more negative comments than positive ones. However, some responses (e.g., "I'll call as soon as I can." and "Can you call back later?") suggest that these replies could be used to schedule follow-up calls, disposition cases, or potentially conduct an interview with the sampled phone number. For subsequent waves, we engaged respondents with non-terminal responses in an effort to solicit a survey complete. We will present a breakdown of the content of these messages, as well as our analysis of the impact to study dispositions and response rates through respondent engagement (e.g., scheduling callbacks, dispositioning refusals).

Applying UI/UX Best Practices for Web Surveys
Rebecca Watkins, RTI International

Web-based surveys are becoming an increasingly common method of capturing data. Survey professionals are already aware of the various advantages and disadvantages of web-based surveys, but all can agree that survey methodologies should not be abandoned on the web. However, achieving recommended design guidelines for surveys online can be a struggle. RTI International recently went through the process of creating and documenting a Cascading Style Sheet (CSS) template that can be implemented across surveys to quickly and easily enforce certain important visual design best practices, such as shaded grids, white answer spaces against a colored background, spacing non-substantive response options apart from substantive options, and others. Further, the template ensured that these evidence-based design principles were in place at all screen sizes. Finally, a consistent brand and color scheme contributes to user confidence in the authenticity of the survey.
An Integrated Approach of Qualitative and Quantitative Methods in Usability Testing of a Survey Data Dissemination Tool

Alda Rivas, U.S. Census Bureau
Erica Olmsted-Hawala, U.S. Census Bureau
Shelley Feuer, U.S. Census Bureau
Lin Wang, U.S. Census Bureau

The U.S. Census Bureau is developing a public-facing data dissemination tool. The tool is designed to improve the public's access to demographic and economic data collected through official surveys. It is thus crucial to ensure good usability of the tool so that a public user can acquire the information of interest with ease. A series of usability testing sessions have been conducted to assess users' effectiveness (accuracy), efficiency, and satisfaction while using the tool. We used an integrated approach of both qualitative and quantitative methods to evaluate usability with different user groups (e.g., novice vs. expert) and devices (e.g., laptop vs. smartphone). We then presented the results and recommendations to stakeholders. In this presentation, we will share our methodologies and findings with the audience using real testing data.

Not a Copy and Paste: Makings of non-English data collection instruments

Brianda Perez, U.S. Census Bureau
Lily Kapaku, U.S. Census Bureau

The 2020 Census will be the first census to provide an option to respond online, not only in English, but in 12 non-English languages. With this innovation comes complexities, as providing non-English translations within data collection instruments is not as simple as just displaying translations on screen: English content must be designed with translation in mind, so dynamic content in English, such as a switch between verb tenses, does not become impossible in non-English languages.; Formatting used for emphasis in English, such as bold, all caps, italics and other styles, must be chosen carefully to apply (or apply with modifications) across non-English languages.; Elements within the design may need to be adjusted to account for text expansion, including headers and buttons.; Right-to-left languages need development modifications and adjustments for images such as directional arrows. This presentation will focus on these necessary customizations, adjustments to non-English user interfaces, and tips on developing non-English translations for data collection instruments.
Session 6B: 10:25 am - 11:40 am

Questionnaire Design

Room 2

Collecting and Using Web Paradata to Explore Measurement and Nonresponse Among Nonprofit Organizations

Matt Jans, ICF
Adam Lee, ICF
Summer Brenwald, ICF
Jennifer Greer, ICF
Arlen Rosenthal, ICF
Sherri Mamon ICF
Rebecca Morrison, NSF
Ronda Britt, NSF
Jock Black, NSF
John Jankowski, NSF

Web survey paradata are perhaps some of the most complex paradata systems in practice. This presentation will show a sophisticated paradata system designed for the Nonprofit Research Activities (NPRA) Survey. This survey of US-based nonprofit research organizations asked questions about the amount and type of research that they perform or fund. One specific feature of the data collection system is that it performed multimode sample management and data collection functions, and was designed to be accessed by both project staff and respondents. The paradata collected included user agent strings (e.g., information about the device type, operating system, and web browser used by someone logging in), events (i.e., interim dispositions and actions that reflect an organization's status at any point in time during the fielding period), and user actions (e.g., clicks on links and buttons while completing the questionnaire). This presentation describes the overall functions and features of the system, but focuses on analyzing paradata to answer the following research questions: 1) Does the amount of time spent answering questions vary across types of nonprofits, 2) Do certain question types take longer to answer than others, and 3) Do break-offs occur at particular questions or points in the questionnaire. The results are framed in the context of Total Survey Error and OMB burden guidelines.
FedCASIC 2019 Workshops Program

Session 6C: 10:25 am - 11:40 am
More Machine Learning Applications for Surveys and Beyond
Room 3

Machine Learning for Medical Coding in Health Care Surveys
Emily Hadley, RTI International
Rob Chew, RTI International
Jason Nance, RTI International
Peter Baumgartner, RTI International
Rita Thissen, RTI International
David Plotner, RTI International
Christine Carr, RTI International
Aerian Tatum, HealthCare Resolution Services

Manually coding free-form text responses in surveys can be a time-intensive process subject to human error and high labor costs. For health care surveys, the process of medical coding is particularly complex due to the need for specialized domain knowledge, the varying quality of clinical notations, and the large number of classification codes (nearly 70,000 diagnoses are present in the ICD-10-CM classification system alone). Given the challenges posed to medical coders and the constraints placed on statistical agencies to develop high-quality estimates within budget, machine learning techniques offer potential gains in both efficiency and quality. In this talk, we describe a machine learning approach for assigning medical codes to clinical verbatim text found in medical records for patient visits from the 2016 and 2017 National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medical Care Survey-Emergency Department (NHAMCS-ED). We discuss the process of creating machine learning models, evaluating the performance of a benchmark model, and potential use cases and future work. From the results of this approach, we would not recommend replacing human coders; however, our results suggest the trained machine learning model could be used in complement with manual coding through use of a code recommendation system or automated quality assurance checks for medical coders.

Coding And Tagging of Qualitative Interview Transcripts Using a Natural Language Processing (NLP)-Based Machine Learning Algorithm
Josh Levy, RTI International
Kevin Wilson, RTI International*
Justin Landwehr, RTI International
Julianne Payne, RTI International

Human coding and tagging of recorded focus group and interview transcripts is an expensive and time-consuming activity that is nevertheless necessary for diverse areas of survey research. In response to the growing demand for human coding along with shrinking research budgets, we have developed an NLP machine-based method for initial coding of qualitative interview transcripts: TaCT (transcript and coded tagging), with the aim of reducing human-coding time and yet still maintain a high standard of coding quality. The TaCT pipeline has unique features that make it especially useful when multiple coders are working with the same transcripts. For example, TaCT can detect and merge codes assigned by different coders to the same text passage. This greatly increases both precision and recall statistics, leading to more robust code assignments. After text cleaning, the support vector machine (SVM) is trained, and coding applied to test sets. For initial development, we split training and test sets into multiple runs using between 1% and 90% of the available text for training purposes. In this talk we will discuss the results, and the impact of TaCT on future transcript coding work.
More Machine Learning Applications for Surveys and Beyond

A Coding Application

Gina Cheung, Survey Research Center, Univ. of Michigan
Cheng Zhou, Survey Research Center, Univ. of Michigan*

Survey Research Operation (SRO) in University of Michigan developed a coding application back in 2002. The tool has outdated with modern technology and also it was designed to tightly couple with one data collection software. In the presentation, we will discuss and demo a new coding application we have developed, which incorporates the new development of software, database, and machine learning technology with the aim to achieve efficiency, consistency, enhanced reporting functions, low maintenance and versatility. The new application is suitable to do any kind of coding activities and it is independent from any data collection software.

Predicting Missing Levels Using Supervised Machine Learning


The National Compensation Survey (NCS) is an establishment survey program that provides comprehensive measures of occupational wages, employment cost trends, and benefit incidence and detailed plan provisions. For each occupation sampled from the non-Federal sector, the NCS assigns a level that is equivalent to the General Schedule (GS) grade level used in the Federal sector. This is accomplished by using the four-factor system provided by the Office of Personnel Management (OPM). For each sampled occupation, points are assigned to each factor, and the aggregated points are placed in one of the fifteen point ranges corresponding to a specific GS grade level. Consequently, an occupation cannot be leveled if any of the four factor points is missing, and the GS grade level suffers from item nonresponse. However, unlike the other survey items in the NCS, there currently is no imputation procedure in place to fill in the missing level information. This study first examines the pervasiveness of partially coded factor points resulting in missing GS grade levels, then it utilizes supervised machine learning to fill in those missing levels by exploiting the partially available factor points. Lastly, this study compares the results from machine learning approach against those from a more traditional approach.
Surveys are universally experiencing falling response rates and rising data collection costs. While many people choose not to respond to surveys, many people do. Finding out the reasons why they participate will inform ways to motivate more people to respond. Sparked by the ICF market research conducted for the 2010 Census, we investigate attitudinal motivators and barriers to government survey participation, particularly health surveys. The respondents for this research were recruited from a national online panel. The questions focused on interest in health, interest and trust in surveys (general and government-sponsored), community involvement, trust in government generally, and demographics. Using the results of the survey, cluster respondents into four segments: a) Compliers, who are most likely to respond to surveys on any topic; b) Fence Sitters, who are willing to take surveys, but generally not as enthusiastic as the compliers; c) Reluctant Participants, who tend to keep to themselves and have low trust in government; and d) Distrusters, who have low trust in government and are concerned about data security.

While our research is limited to surveying survey participants about survey participation, we find distinct differences in likelihood to respond to a government survey on a topic they considered to be important. Ninety-five percent of compliers reported they probably or definitely would, while only 59% of Reluctant Participants reported they probably or definitely would.

The focus of our presentation is on responses to questions about trust in government and attitudes about government surveys as well as the characteristics and stated behavior of the respondent segments who score low on these metrics.
Session 7A: 12:45 pm - 2:15 pm
Audience Segmentation for Adaptive Design: An Approach to Improving Nonresponse in Surveys
Room 1

Topic Salience and Propensity to Respond to Surveys: Findings from a National Mobile Panel
John Boyle, ICF
Lew Berman, ICF
Jamie Dayton, ICF
Ronaldo Iachan, ICF
Matt Jans, ICF
Robynne Locke, ICF
Randy ZuWallack, ICF

In response to declining response rates, it is important to reexamine why people participate in surveys. Leverage-salience Theory posits that people will respond to surveys as a function of the importance of a survey design (e.g., topic, incentive, duration, mode), and respondent awareness of that feature. To explore topic salience within a broader framework of survey participation, we conducted a study among a national sample of nearly 2,000 adults from a mobile panel. After answering questions about general propensity to respond to surveys, respondents were asked how important it would be to participate in government surveys on seven randomized topics, including: education, aging, health, voting, and issues facing the nation. They were subsequently asked how likely they would be to participate in a government survey on: (1) a topic they endorsed as being important; and (2) a topic endorsed as not important. The likelihood of responding to a survey is directly related to the importance of the topic to the respondent. The importance of the seven topics varies across the population. Hence, survey topic should be described in terms with the broadest interest for the highest response rate. At the same time, salience of specific topics vary by demographics and lifestyle, so researchers should consider tailored protocols to compensate for population segments with lower interest in the topic. Since topic importance can be partially predicted by demographics and lifestyle, sample information and paradata may be useful in a tailored design. Additional information collected as part of screening, could substantially improve subgroup targeting within a responsive design.
Audience Segmentation for Adaptive Design: An Approach to Improving Nonresponse in Surveys
Room 1

Willingness to Participate in Physical Measures and Specimen Collection in Health Surveys
John Boyle, ICF
Lew Berman, ICF
Jamie Dayton, ICF
Ronaldo Iachan, ICF*
Matt Jans, ICF
Randy ZuWallack, ICF

There is increasing interest in adding objective physical measures and specimen collection to health surveys to augment or replace some of the self-reported health measures. Unfortunately, we have little information on the willingness of the public to participate in these extensions of traditional health surveys; which they are willing to do; and who is willing to participate in them. Our adaptive approach investigates which subgroups or segments may consider different forms of collection of physical measures within a framework of health survey participation. The research used data from a national mobile panel survey conducted for a Census balanced sample of nearly 2,000 adults who responded through the panel’s smart-phone app.

After asking respondents about their willingness to participate in surveys with different topics, modes, sponsorship, and incentives, respondents were asked about their willingness to participate in health studies that included physical measures and specimen collection. These included measurement of height and weight, blood pressure, waist circumference, and finger stick blood draw by a health representative. They were also asked about their willingness to measure their blood pressure themselves using a machine in their home and the willingness to install an app on their phone for a health study.

We will explore key predictors of willingness to participate in these measures, including demographics, attitudes, lifestyles, and health status. The analyses using classification and regression trees allow for interactions between these predictors and a prioritization of the main predictors, suggesting solutions for the next generation of health surveys.

Designing, implementing, and analyzing Leverage Saliency Theory experiments
Matt Jans, ICF
Julia Sumner, ICF
Rachael Forando, ICF
John Boyle, ICF
James Dayton, ICF
Lew Berman, ICF

Randomized experiments are a disciplinary best practice for understanding the effects of survey design features on participation, but they are often difficult to conduct. This presentation describes, in detail, the conceptualization and implementation of experimentally-designed survey scenarios to explore how interview duration and incentive amount affect participation decisions. Respondents in MFour’s Surveys-on-the-Go® panel were asked about their likelihood to respond to an interesting and uninteresting survey with interview duration (1 hr v. 2 hrs) and incentive level ($25 v. $50) randomized factorially. We discuss how the experiment was designed and fielded, including screen shots of the various question versions, and how we analyzed the data to understand the effects of duration and incentive amount while controlling for other respondent characteristics using Stata’s margins and margins plots commands with weighted survey data. We also share initial results. The presentation will include an experimental design and analysis workflow checklist for such experiments so that the audience leaves with a concrete tool for implementing and analyzing experiments on survey features in CASIC environments.
Session 7B: 12:45 pm - 2:15 pm
Quantitative and Qualitative Pre-Testing
Room 2

Using Online Testing and Wearable Devices to Pretest Diary and Stylized Sleep Measures
Brandon Kopp, U.S. Bureau of Labor Statistics

Data on sleep duration is of great interest to researchers, government, and health organizations, as sleep can impact important health and social outcomes. One common data collection method is time diaries such as the American Time Use Survey (ATUS). In the ATUS, interviewers use a set of scripted, open-ended questions to walk respondents chronologically through their activities during the prior 24-hour day, including sleep. In contrast, other surveys use stylized questions that ask people about the “average, normal, or typical” amount of sleep they get in a given timeframe. Although in theory diary and stylized questions should measure the same construct, a sleep gap has been observed where diary measures tend to exceed stylized measures of sleep. We present multi-method research investigating possible reasons for the sleep gap. We conducted online research where participants (N=1200) completed a time-diary and answered stylized questions about their sleep. We found that how participants defined sleep and the order of the diary and stylized sleep questions affected reports of sleep duration. The findings also suggested that some features of diary and stylized sleep measures may lead to different types of measurement error. Although this study provided rich insights into the sleep gap, we were unable to determine whether one measure was more accurate than the other. To get at accuracy, we conducted a validation study using a wearable device that tracked participants’ sleep duration over a one-week period. Participants (N = 35) completed two interviews about one week apart. At visit one, participants answered basic demographic questions and then were given instructions to wear the device at all times for one week. At visit two, participants completed a time-diary and answered stylized questions about their sleep. We then compared participants’ self-reported sleep duration from the time diary and stylized measures to the device-recorded sleep data. During debriefing, we asked participants about their response process across the diary and stylized measures, and probed about any discrepancies from the device-recorded sleep data. We discuss participants’ perceptions of the accuracy of the device-recorded sleep data, and how in some instances the data served as a useful memory cue for participants in recalling their sleep and wake times more precisely. We summarize the findings, implications, and logistics involved in conducting sleep validation research using wearable devices, and how these methodologies enabled us to gain a deeper understanding of measurement error in questionnaire pretesting.

The Role of Micro-surveys in Supporting Agile Software Development
Margo Kabel, Veterans Health Administration
Danielle Kato, Veterans Health Administration

Veterans’ Health Administration (VHA) is developing an integrated tool to facilitate management of medication orders/prescriptions in both inpatient and outpatient settings. VHA Pharmacy Benefits Management has been charged with designing a graphical use interface to replace the current text based system. The development team was interested in getting feedback from staff pharmacists in medical facilities on the design of this application. Typical methods used to obtain user feedback early in the development process include focus groups and interviews. Both of these methods (which require between 30 minutes to 1.5 hours) were unacceptable to Pharmacy Supervisors because they felt it the time required would adversely impact patient care. In response, we developed ‘Micro-Surveys’. A micro-survey (or data call) is an online questionnaire that focuses on a single topic and can be completed in less than 10 minutes. This paper will discuss the use of micro-surveys to provide just in time user feedback in Agile software development for health care.
Session 7B: 12:45 pm - 2:15 pm
Quantitative and Qualitative Pre-Testing
Room 2

Using Instrument Paradata to Assess Question Wording Revisions
Jasmine Luck, U.S. Census Bureau
Jonathan Katz, U.S. Census Bureau
Mary Davis, U.S. Census Bureau
Matthew Virgile, U.S. Census Bureau

Instrument paradata can help inform whether survey questions are problematic in web and CATI/CAPI instruments. After problematic questions are identified, they can then be pretested to help inform changes to wording or formatting that could represent improvements. If the recommended change is implemented, replicating the paradata research used to detect the original problem would demonstrate whether there was an improvement. Staff from the U.S. Census Bureau have conducted questionnaire pretesting on the National Teacher and Principal Survey (NTPS), for three survey cycles. In this presentation, we will be focusing primarily on the web instrument. First, we will present a history of the revisions to web survey questions that have been problematic. We will discuss the findings from multiple phases of testing used to inform the question wording revisions. Second, we will evaluate whether the question wording revisions for the 2017-2018 cycle were successful by looking at select metrics of the 2015-2016 web paradata and comparing that to the 2017-2018 web paradata. Finally, we will discuss future research goals in using web paradata to examine the performance of question wording revisions.

Marrying Operational Data And Customer Feedback Data To Save Taxpayer Dollars
Anahita Reilly, U.S. General Services Administration

GSA is the "B2B" agency of the federal government, delivering value and savings in real estate, acquisition, technology, and other mission-support services across government. Through this session, we will share examples of how we are using an enterprise-wide Voice of the Customer program to align feedback from both agency customers and industry partners with operational data, informing how we design our products and services. We will also touch on the governance structure set in place to enable self-service feedback collection processes while adhering to a consistent enterprise-wide brand that minimizes survey burden to the customer.
Session 7C: 12:45 pm - 2:15 pm
Advances in the Use of Location Data
Room 3
Chair: Beth Nichols, U.S. Census Bureau

Using Computer Vision to Detect Housing Units from Satellite Imagery
Stephanie Eckman, RTI International
Qiang Qiu, Duke University

We draw upon recent advances in computer vision algorithms and in the availability and resolution of geospatial big data to improve the way data are collected for social science research. High quality surveys require a list of dwellings from which to select a sample. Many studies undertake considerable investments of time and money to create a list of households, and this costly work is repeated by many studies each year in the US, leading to inefficiencies and redundancies across projects. Research has shown that the most common methods of housing unit listing underrepresent rural households. The unique characteristics of the rural population, combined with the continued use of enumeration methods that miss some households, can bias estimates. Our approach uses computer vision techniques to detect dwellings in satellite images. Satellite images have become increasingly available at high resolutions and reduced cost in the last few years. During the same time, advances in computer vision have improved the ability of computers to find and identifying objects. We trained an algorithm to detect housing units in Wake County, NC. The approach achieves 83% accuracy and does not require large amounts of training data. We will continue to refine the algorithm over the coming months and expand it to other areas.

Development and Maintenance of Location-Based Data to Support Statistical Programs
Michael Ratcliffe, U.S. Census Bureau

All activity measured by the Census Bureau occurs within or across space. A census or survey response is associated with a specific location, whether a residence, a place of work, an establishment, a business, etc. In the case of functional relationships, response information may be associated with multiple locations that are linked by a specific activity; for example, a journey to work, a commodity flow, a migration flow. As such, an accurate framework of geospatial and location-based data is critical to the Census Bureau’s mission. The fundamental geospatial attribute is the latitude and longitude (x,y) coordinate. This allows for assignment of any respondent or respondent attribute to a unique location on the surface of the earth. This also allows for determination of the relationship between the respondent and attributes and all geographic entities used to tabulate and disseminate data. From the coordinate, all other geospatial relationships are possible. In this presentation, I will discuss ongoing efforts, including research and innovation activities, within the Census Bureau to collect, develop, and manage location-based data as part of the geospatial frame supporting statistical data collection, tabulation, and dissemination. This includes acquisition of location-based data from tribal, state, and local government partners as well as coordinate locations based on structure outlines or interpolated based on sequencing along a network.
FedCASIC 2019 Workshops Program

Session 7C: 12:45 pm - 2:15 pm
Advances in the Use of Location Data
Room 3

Predicting Segments and Household Characteristic Using Satellite and Street View Images
Raphael Nishimura, Survey Research Center, Univ. of Michigan

Many face-to-face surveys need to screen households to identify eligible respondents for their studies, which tends to be very time- and cost-intensive. We can improve the efficiency of such efforts by using auxiliary information about the sampled units to predict their eligibility and use these predictions to assist the sample design, with stratification and differential selection probabilities, for instance. For this purpose, public-available satellite and street view images can be used as auxiliary information to feed machine-learning algorithms. In this presentation we will show initial results of such attempts. We initially trained a deep learning algorithm to predict household eligibility based on Google Street View images from geo-coded sampled households, which we know their eligibility status from previous surveys. Then, we evaluate the results of the algorithm using a validation dataset of households with their corresponding Google Street View images. This same approach is also applied to predict sampled segments using satellite images.

An Exploration of User Map Preferences and Behaviors
Jessica Holzberg, U.S. Census Bureau
Elizabeth Nichols, U.S. Census Bureau
Lin Wang, U.S. Census Bureau

Every ten years, the Census Bureau sends enumerators to housing units that have not responded to the census for nonresponse followup. During this in-person followup phase, some of these housing units are found to be vacant. If the Census Bureau could obtain information about the occupancy status of a housing unit earlier, personal visits to vacant housing units could be avoided, thus reducing fieldwork costs. To accomplish this, one idea is to ask people who have already responded to the census to identify neighboring homes that are vacant. The potential use of this strategy raises several questions, including whether census respondents can find themselves and their neighbors on a map or address list, accurately identify units around them that are vacant, and report that information willingly. In this presentation, we discuss findings from a qualitative exploratory study to gather more information about these issues. Study participants were shown different types of housing unit displays (e.g., satellite maps and road maps) and were asked about their preferences. We also asked questions to assess how knowledgeable participants are about their neighboring housing units and how willing they are to report on their behalf, as well as how participants use online maps in their daily living. We will conclude with recommendations for researchers interested in collecting data from users using maps.
Session 7C: 12:45 pm - 2:15 pm
Advances in the Use of Location Data
Room 3

Best Practices in Consent to Capture Geolocation Data in Self-Administered Web Surveys
Scott D. Crawford, SoundRocket

Capturing geolocation data in web-based surveys is no longer the 'new thing' - survey researchers have been doing it now for several years. Navigating the ethical and consent related issues of doing so, however, is not yet thoroughly discussed and established. We will share our research, and summarize others research, to begin to build a set of best practices around implementing geolocation capture in self-administered web surveys. In this presentation we will remain focused on web-based survey data collection - but will emphasize efforts where mobile devices are used to access the web surveys (as those are most commonly where geolocation data is relevant). However, we will not discuss issues relevant to mobile apps. Where data exists, we will discuss how these practices contribute to data quality.

Session 8A: 2:25 pm - 3:25 pm
Emerging Survey Applications for Mobile
Room 1

A Smartphone App to Record Food Purchases and Acquisitions
Marcelo Simas, Westat
Ting Yan, Westat
Janice Machado, Westat

The proportion of American adults owning a smartphone has almost doubled since 2011; now about two thirds of American adults own a smartphone of some kind and for many of them their smartphone becomes an important tool in their lives (Pew Research Center, 2015). Given this trend, researchers are also increasingly using smartphones for various research purposes such as ecological momentary assessment, transportation and time-use diary studies, health monitoring, and passive mobile data collection (e.g., Link, Lai, and Bristol, 2014; Sonck and Fernee, 2013; Revilla, Ochoa, and Loewe, 2016). This paper describes an innovative use of smartphones to collect food purchases and acquisitions. Food purchases and acquisitions used to be collected through paper diaries. We developed a smartphone app that allows respondents to enter all foods and drinks they have obtained on a smartphone. The app takes advantage of smartphone functionalities to reduce reporting burden. For instance, the app tracks respondents' geo locations passively and uses these locations to cue food acquisition. In addition, we linked up with various databases to reduce the burden of reporting name and address. Furthermore, the app allows respondents to scan barcodes, to take pictures of food items, and to use auto-complete to get food item descriptions. We will also present the results of internal usability testing conducted in which the burden of using the app was compared to that of paper questionnaires.
Pre/Post Hurricane Surveys: Measuring hurricane preparations and post storm impacts using geofencing methods
James Dayton, ICF
Thomas Brassell, ICF

Surveying recent hurricane victims is a difficulty prospect, let alone survey potential victims. This paper reports on a novel technology that can track the location of consenting non-probability panel participants before, during and after the storm event and trigger surveys prior to and following the event. Combining an anticipated storm path with panel member residence and geo-location information can provide a useful sampling frame for potential and actual victims. Prior to Hurricane Florence, ICF fielded a pre-storm survey with nearly 2,200 respondents along the expected storm path just prior to the hurricane’s scheduled landfall. The pre-storm questionnaire asked about storm preparation, evacuation plans, previous hurricane experiences, and current health and psychological stress. The post-storm survey included the same measures, as well as questions about the impact of Hurricanes Florence and Michael, completing post-storm questionnaires with over 60% of pre-storm respondents and 750 new respondents in the region impacted by Florence and Michael. These results were analyzed to provide data to disaster planning and response organizations on the pre and post storm preparations, the extent and nature of damage, estimated cost of damage (covered and not covered by insurance), planned and actual evacuation information as well as health and psychological effects. These findings have major implications for the methodology of disaster-related surveys.

Path 2 Post-Disaster Behavior: An evaluation of the utility of geolocation tracking and geofencing to assess activity post-disaster
Thomas Brassell, ICF
James Dayton, ICF

The behavior of survivors immediately following a disaster is not well understood. While post-disaster surveys typically look to address the degree of impact and needs of the surviving population, at minimum these surveys are often distributed a few weeks following the event. Although this information is valuable, so is understanding the immediate activities of the population impacted within the first few days following the disaster when aid may be scarce and needs high. Using the geolocation of mobile panel members located in areas impacted by hurricanes Florence and Michael, we identified members via geolocation tracking who visited geofenced home improvement stores post-disaster. We surveyed 250 respondents regarding purchases related to home repairs needed due to the hurricanes, specifically collecting information regarding type of purchase and amount spent, as well as the presence of insurance coverage for the repairs and the potential financial impact. Geofencing has typically been applied to marketing research activities, with organizations collecting information on the purchasing habits of their own or their competitor’s consumers. However, its utility could be expanded to geofence disaster support centers, community centers, local shelters, and medical service providers. Implications of the results of this study and the future application of this technology in understanding the needs of survivors of disasters will be discussed.
Session 8B: 2:25 pm - 3:25 pm
Towards Web Surveys
Room 2

Ongoing Experimentation in Transition from a National Paper and Pencil Survey to a Multi-Mode Web and Paper and Pencil Survey
Sarah Grady, Department of Education

Between 2007 and 2012, the National Household Education Surveys, sponsored by the National Center for Education Statistics, transitioned from a random-digit dial (RDD) phone survey to a paper-and-pencil (PAPI) survey using an Address-Based Sample (ABS) frame. The NHES is a complex two-stage survey. To address issues of rising cost and declining response rates, as well as reduce the complexity of administration, an experiment in 2016 tested the feasibility of adding the web mode to the NHES design. Based on the results of this work, NHES will be administered starting in January 2019 with most sampled addresses receiving a request to complete the survey by web before being offered paper. Several experiments are embedded in the NHES:2019 design to test the effectiveness of multi-mode design options and increase response rates. This presentation will provide an overview of the experiments planned for NHES:2019 including the reasons for the experiments, the design, and some very preliminary results if available. Experiments include tests of advanced mailings, use of a contingent $10 or $20 incentive to encourage web response, measures of response gain from using a pressure-sealed mailer with web credentials in lieu of a postcard, and targeting households modeled as being least likely to respond by web with PAPI only. This presentation is intended to spur discussion about methodological challenges in multi-mode household survey administration and potential solutions.

Examining the Effectiveness of Push-To-Web Mixed Mode Approaches on Response Rates
William G. Lehrman, Centers for Medicare & Medicaid Services*
Layla Parast, RAND Corp.
Megan Mathews, RAND Corp.
Marc Elliott, RAND Corp.
Anagha Tolpadi, RAND Corp.
Kirsten Becker, RAND Corp.*
Elizabeth Flow-Delwiche, Centers for Medicare & Medicaid Services
Debra Stark, Centers for Medicare & Medicaid Services

There is growing interest in using surveys to measure patient experience in all healthcare settings, and to employ new technologies to economically reach patients. Recent work involving survey administration among emergency department (ED) patients using single-mode approaches (e.g. mail-only, web-only) has demonstrated very low response rates. High response rates are desirable to accurately represent the experience of care of the entire ED patient population. Using an experimental design, we tested the viability of a web-based survey within the context of mixed mode survey protocols, both with and without survey invitations, with particular interest in response rates and respondent characteristics. Study Design: Patients who had visited the ED and were discharged home (i.e., discharged to the community [DTC]) were sampled and randomized to one of nine survey administration arms. The reference arm was standard mixed mode (mailed survey with telephone follow-up). All other arms involved a web survey with one or more of the following methods of invitation: an emailed link to the web survey, a texted link to the web survey and/or a mailed paper invitation with information on how to access the web survey. If the survey had not been completed within a specified timeframe, the web survey was followed by (1) a mail survey, (2) a telephone survey, or (3) both.
Session 8B: 2:25 pm - 3:25 pm  
Towards Web Surveys  
Room 2

AURA System – NASS’s Survey Solution  
Ramonia Davis, National Agricultural Statistical Service

The National Agricultural Statistics Service (NASS) is responsible for conducting hundreds of surveys every year covering virtually every aspect of U.S. agriculture. Production and supplies of food and fiber, prices paid and received by farmers, farm labor and wages, farm finances, chemical use, and changes in the demographics of U.S. producers are only a few examples. NASS is committed to providing timely, accurate, and useful statistics in service to U.S. agriculture. To accomplish this mission, NASS develops paper and web surveys to collect the data from producers across the continental US and Puerto Rico that is necessary to estimate production and identify production trends.

The realization that NASS must utilize current technology to fulfill its mission has contracted for the development of new software, Aura Survey Designer that replaces the legacy software that was developed in 2002, known as the Question Repository System (QRS). The new Survey Designer is a Windows client application providing a user friendly, respondent web application for our clients. It also provides an enumerator web application and an upcoming enumerator mobile application. The new software allows the generation of multiple versions of data collection instruments for paper and web modes of data collection.

Session 8C: 2:25 pm - 4:15 pm  
Workshop on Rapidly Prototyping a Machine Learning Pipeline  
Rooms 9-10

Christian Moscardi, U.S. Census Bureau

Have you ever wondered how machine learning could improve your team's workflow? Have you thought about "creative reuse" of data to take advantage of its inherent structure and patterns? In this workshop, we'll use unstructured text data (free-text product descriptions) to develop a model that can accurately predict product codes. We'll also include some core machine learning techniques, concepts, and metrics... and we'll show how easy it is to develop a model in Python. We hope that after this workshop, you'll be eager to learn more and maybe even try out some machine learning on your team's data.

Please note that if you bring a laptop with Anaconda/Python installed, we'll have a Jupyter Notebook published online, and you can follow along in-person.
Session 9A: 3:15 pm - 4:15 pm
Workshop on Basic Usability Testing
Room 3

Jean E. Fox, U.S. Bureau of Labor Statistics
Brooke Dine, Department of Health and Human Services

In this workshop, we will cover the basics of usability testing. We will explain what usability testing is (and isn’t), and how to plan, conduct, and analyze a basic usability test. We will look at how to identify tasks, recruit participants, run test sessions, and evaluate the results of a simple usability test. We will also recommend resources for those interested in learning more on the topic.

Session 9B: 3:15 pm - 4:15 pm
Workshop on Eye Tracking
Room 7-8

Erica Olmsted-Hawala, U.S. Census Bureau
Elizabeth Nichols, U.S. Census Bureau
Lin Wang, U.S. Census Bureau
Temkia Holland, U.S. Census Bureau

Usability studies of web surveys investigate user performance on the survey in terms of accuracy, efficiency, and satisfaction. Eye tracking is one tool researchers use to assess the usability of the survey. It can assist the researcher as s/he attempts to understand how the survey works for respondents. It can also provide quantitative data for empirical testing of design differences. With eye-tracking, one may be able to determine how the user engaged with the survey including whether the user looked at the question stem, instructions, response options as well as navigation features. This hands-on demonstration will allow conference attendees to experience different eye-tracking stations including variations of eye tracking on laptops, and different methods to eye track on mobile devices: using an eye-tracking stand where the mobile device has to remain fixed in one spot and using eye-tracking glasses where the respondent wears the glasses and the mobile device can be held in a more natural position. Challenges of each method will be highlighted during the demo along with instruction and feedback on how to gather eye-tracking data including calibration, importance of a high sampling rate, and analysis of the data. Finally, participants can watch a 10-minute slideshow that covers the topics of the basics of eye-tracking methodology including the human visual system, information processing, principles of eye movement recording, and data analysis and interpretation.