



# **Federal Committee on Statistical Methodology 2018 Research and Policy Conference**

## **Final Program and Abstracts**

**Walter E. Washington Convention Center  
March 7 – 9, 2018**

### **Major Sponsors**

Bureau of Economic Analysis  
Bureau of Justice Statistics  
Bureau of Transportation Statistics  
Department of Veterans Affairs  
Economic Research Service  
Energy Information Administration  
National Agricultural Statistics Service  
National Center for Education Statistics  
National Center for Science and Engineering Statistics  
National Center for Veteran Analysis & Statistics  
Office of Research Evaluation and Statistics, Social Security Administration  
Statistics of Income Division, Internal Revenue Service  
U.S. Census Bureau

### **Also sponsored by**

Mathematica Policy Research  
NORC at the University of Chicago  
RTI International  
Westat

### **Hosted by:**

Council of Professional Associations on Federal Statistics

## **FCSM 2018 RESEARCH AND POLICY CONFERENCE PROGRAM COMMITTEE**

Andrew Zukerberg, Co-Chair, National Center  
for Education Statistics

David Kashihara, Co-Chair, Agency for  
Healthcare Research and Quality

Denise Abreu, National Agricultural Statistics  
Service

Jennifer Beck, National Center for Science and  
Engineering Statistics

Edgardo Cureg, Energy Information  
Administration

Irena Dushi, Social Security Administration

Maria Garcia, U.S. Census Bureau

Jessica Graber, U.S. Census Bureau

Robin Kaplan, Bureau of Labor Statistics

Grace Kena, Department of Justice

Dave Kinyon, Energy Information  
Administration

Lynn Langton, Department of Justice

Yan Liu, Statistics of Income Division, Internal  
Revenue Service

Janine McFadden, Department of  
Transportation

Paul Scanlon, National Center for Health  
Statistics

Erik Scherpf, Economic Research Service

George Sheldon, Department of Veterans  
Affairs

Darius Singpurwalla, National Center for Science  
and Engineering Statistics

Makram Talih, National Center for Health  
Statistics

Scott Wentland, Bureau of Economic Analysis

Andrew White, National Center for Education  
Statistics

Alice Yu, Bureau of Labor Statistics

## **CONFERENCE COORDINATORS**

Krista Cannon, U.S. Census Bureau

Mae Pattison, Council of Professional Associations on Federal Statistics

LoWanda Rivers, U.S. Census Bureau

Elaine Russell, U.S. Census Bureau

## FEDERAL COMMITTEE ON STATITICAL METHODOLOGY

John Eltinge (Chair), U.S. Census Bureau

Nancy Bates, U.S. Census Bureau

Jonaki Bose, Substance Abuse and Mental Health Services Administration

Patrick Cantwell, U.S. Census Bureau

Chris Chapman, National Center for Education Statistics

Jennifer Edgar, Bureau of Labor Statistics

Dennis Fixler, Bureau of Economic Analysis

Barry Graubard, National Institutes of Health

Michael Hawes (CDAC Chair), Department of Education

Jennifer Madans, National Center for Health Statistics

Rochelle (Shelly) Martinez, Office of Management and Budget

Wendy Martinez, Bureau of Labor Statistics

Jaki McCarthy, National Agricultural Statistics Service

Pamela McGovern (Webmaster), National Agriculture Statistics Service

Grace Medley, Substance Abuse and Mental Health Services Administration

Peter Miller, U.S. Census Bureau

Renee Miller, Energy Information Administration

Jeri Mulrow, Bureau of Justice Statistics

Jennifer Parker, National Center for Health Statistics

Polly Phipps, Bureau of Labor Statistics

Mark Prell, Economic Research Service

Joseph Schafer, U.S. Census Bureau

Rolf Schmitt, Bureau of Transportation Statistics

Marilyn Seastrom, Department of Education

Joy Sharp, Bureau of Transportation Statistics

G. David Williamson, Centers for Disease Control and Prevention

Gordon Willis, National Institutes of Health

Linda Young, National Agriculture Statistics Service

## COUNCIL OF PROFESSIONAL ASSOCIATIONS ON FEDERAL STATISTICS (COPAFS)

John Thompson, Executive Director

Corinna Turbes, Associate Director of Strategic  
Communications and Policy Analysis

Mae Pattison, Executive Assistant



## **2018 FCSM Research and Policy Conference**

The 2018 Federal Committee on Statistical Methodology (FCSM) Research and Policy Conference is one of the activities of the FCSM. The FCSM is a federal statistical interagency committee organized under the auspices of the Office of Management and Budget's Statistical and Science Policy Office and is dedicated to improving the quality of federal statistics. The committee's major goals are to:

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

The 2018 FCSM Research and Policy Conference has two themes. The first is to provide a forum for experts from around the world to discuss and exchange current research and methodological topics relevant to federal government statistical programs. The second is to learn about and contribute to the modernization initiatives that the Office of the Chief Statistician of the United States is sponsoring. The conference will offer papers on a wide range of topics that address these themes, including: survey design and methodology; new ways to leverage administrative records; combining data from multiple sources; emerging issues in confidentiality, informed consent, and disclosure avoidance; improvements to economic statistics; artificial intelligence and machine learning; Bayesian and non-probability statistical methods; advances in imputation, estimation and record linkage; and initiatives to modernize the Federal Statistical System.

Sessions feature participation from government, private sector, and academic researchers from three countries. All paper sessions include an open discussion; some sessions include a formal discussion. Papers will be made available on the FCSM Web site (<http://fcsm.sites.usa.gov/>) and presentations will be on the site of the Council of Professional Associations on Federal Statistics (<http://copafs@copafs.org>) following the conference.

There will be two keynote speakers:

- Dr. Sharon Lohr, Professor Emerita, Arizona State University
- Dr. Nancy Potok, Chief Statistician of the United States

Federal Committee on Statistical Methodology (FCSM) Research and Policy Conference  
Washington, DC – March 7 - 9, 2018

**Wednesday, March 7**

7:30 a.m.

**Registration**  
(Concourse)

8:30 - 9:30 a.m.

**Welcoming Remarks and  
PLENARY SESSION I**  
(Rooms 146ABC)

9:30 - 10:00 a.m.

**Break**  
(Concourse)

10:00 - 11:45 a.m.

**CONCURRENT SESSION**

- A-1:** Room 146A
- A-2:** Room 146B
- A-3:** Room 146C
- A-4:** Room 145AB
- A-5:** Room 147AB

11:45 a.m. - 1:15 p.m.

**Lunch on Your Own**

1:15 – 3:00 p.m.

**CONCURRENT SESSION**

- B-1:** Room 146A
- B-2:** Room 146B
- B-3:** Room 146C
- B-4:** Room 145AB
- B-5:** Room 147AB

3:00 – 3:15 p.m.

**Break**  
(Concourse)

3:15 – 5:00 p.m.

**CONCURRENT SESSION**

- C-1:** Room 146A
- C-2:** Room 146B
- C-3:** Room 146C
- C-4:** Room 145AB
- C-5:** Room 147AB

- *Medical Assistance available in Room 149A.*
- *Admin Office in Room 148.*

**Thursday, March 8**

7:30 a.m.

**Registration**  
(Concourse)

8:30 - 10:15 a.m.

**CONCURRENT SESSION**

- D-1:** Room 146A
- D-2:** Room 146B
- D-3:** Room 146C
- D-4:** Room 145AB
- D-5:** Room 147AB

10:15 - 10:30 a.m.

**Break**  
(Concourse)

10:30 a.m. – 12:15 p.m.

**CONCURRENT SESSION**

- E-1:** Room 146A
- E-2:** Room 146B
- E-3:** Room 146C
- E-4:** Room 145AB
- E-5:** Room 147AB

12:15 p.m. - 1:45 p.m.

**Lunch on Your Own**

1:45 – 3:30 p.m.

**CONCURRENT SESSION**

- F-1:** Room 146A
- F-2:** Room 146B
- F-3:** Room 146C
- F-4:** Room 145AB
- F-5:** Room 147AB

3:30 – 4:00 p.m.

**Break – Light Refreshments**  
(Concourse)

4:00 - 5:30 p.m.

**PLENARY SESSION II**  
(Rooms 146ABC)

**Friday, March 9**

7:30 a.m.

**Registration**  
(Concourse)

8:30 - 10:15 a.m.

**CONCURRENT SESSION**

- G-1:** Room 146A
- G-2:** Room 146B
- G-3:** Room 146C
- G-4:** Room 145AB
- G-5:** Room 147AB

10:15 - 10:30 a.m.

**Break**  
(Concourse)

10:30 a.m. – 12:15 p.m.

**CONCURRENT SESSION**

- H-1:** Room 146A
- H-2:** Room 146B
- H-3:** Room 146C
- H-4:** Room 145AB
- H-5:** Room 147AB

12:15 p.m. - 1:45 p.m.

**Lunch on Your Own**

1:45 – 3:30 p.m.

**CONCURRENT SESSION**

- I-1:** Room 146A
- I-2:** Room 146B
- I-3:** Room 146C
- I-4:** Room 145AB
- I-5:** Room 147AB

3:30 – 3:45 p.m.

**Break**  
(Concourse)

3:45 – 5:15 p.m.

**CONCURRENT SESSION**

- J-1:** Room 146A
- J-2:** Room 146B
- J-3:** Room 146C
- J-4:** Room 145AB
- J-5:** Room 147AB

This conference will have two Plenary sessions.  
The first on Wednesday morning and the second on Thursday afternoon.

7:30 a.m. – 5:00 p.m.

**Registration**  
(Concourse)

8:30 - 8:40 a.m.

**Welcoming Remarks**

**Location:** Rooms 146A and 146B and 146C

8:40 – 9:30 a.m.

**PLENARY SESSION I**

**Location:** Rooms 146A and 146B and 146C

**Keynote Speech:**  
**Combining Data and Crime**

**Keynote Speaker:**  
**Sharon Lohr** (Professor Emerita at Arizona State University)

---

9:30 - 10:00 a.m.

**Break**  
(Concourse)

---

10:00 - 11:45 a.m.

**Concurrent Session A-1**  
**Advances in Web Scraping for Federal Statistics**

**Location:** Rooms 146A

**Organizer:** Linda J. Young (National Agricultural Statistics Service)

**Session Chair:** Paul Scanlon (National Center for Health Statistics)

**Web-scraping & Similar Tools to Capture Information for BLS Uses**

David Friedman (Bureau of Labor Statistics),  
William Thompson (Bureau of Labor Statistics)

**SABLE: Tools for Web Crawling, Web Scraping, and Text Classification**

Brian Dumbacher (U.S. Census Bureau), Lisa Kaili Diamond (U.S. Census Bureau)

**Evaluating the Use of Web-Scraped List Frames to Assess Undercoverage in Surveys**

Linda J. Young (National Agricultural Statistics Service)

**Seizing Opportunities and Preserving Principles**

Bob Sivinski (OMB)

**Museum Frame Development – A Universe is Comprised of Many Worlds**

Lisa M. Frehill (Institute of Museum and Library Services), Jason Enos (Institute of Museum and Library Services), Matthew Birnbaum (Institute of Museum and Library Services)

10:00 - 11:45 a.m.

**Concurrent Session A-2**  
**Does Survey Response Change Respondent Behavior?**

**Location:** Rooms 146B

**Organizer and Chair:** Alice Yu (Bureau of Labor Statistics)

**Speeders and Sloths: The Effects of Respondent Timing on Web Survey Data Quality**

Michael Planty (RTI International), Lynn Langton (Bureau of Justice Statistics), Marcus Berzofsky (RTI International), Christopher Krebs (RTI International), Christine Lindquist (RTI International)

**Response Burden and Bias in the First National Food Acquisition and Purchase Survey (FoodAPS-1): An Empirical Analysis Based on Respondent Feedback Survey**

Xingyou Zhang (Economic Research Service), John Kirlin (Economic Research Service), Mark Denbaly (Economic Research Service), Elizabeth Larimore (Economic Research Service)

**Changes in Interview Length over a Data Collection Period: Interviewer Learning or Shifting Respondent Characteristics?**

James Dahlhamer (National Center for Health Statistics), Aaron Maitland (National Center for Health Statistics), Stephanie Coffey (U.S. Census Bureau)

**Smart Math Saves Time and Improves  
Communication**

Laura Nielsen, Ph.D. (Environmental Protection Agency), Susan Day (Eastern Research Group)

**10:00 - 11:45 a.m.**

**Concurrent Session A-3**

**New Findings on Wealth of U.S.**

**Households**

**Location:** Rooms 146C

**Organizer:** Jonathan Eggleston (U.S. Census Bureau)

**Session Chair:** Irena Dushi (Social Security Administration)

**Economic Experiences of Disadvantaged Families  
in the Great Recession and its Aftermath**

Lisa Dettling (Federal Reserve Board), Joanne Hsu (Federal Reserve Board)

**A Better Asset and Debt Data Infrastructure for an  
Ethnically Plural America**

William Darity Jr (Duke University), Darrick Hamilton (The New School)

**Retirement Adequacy and Wealth Distribution  
Among Early Savers**

Alice Henriques (Federal Reserve Board), Lindsay Jacobs (Federal Reserve Board), Elizabeth Llanes (Federal Reserve Board), Kevin Moore (Federal Reserve Board), Jeffrey Thompson (Federal Reserve Board)

**Evaluating Wealth Data in the Redesigned 2014  
Survey of Income and Program Participation**

Jonathan S. Eggleston (U.S. Census Bureau), Michael Gideon (Amazon)

**Discussant:** Howard Iams (Social Security Administration)

**10:00 - 11:45 a.m.**

**Concurrent Session A-4**

**Non-Probability Sampling and Estimation**

**Location:** Rooms 145AB

**Organizer:** Jill A. Dever (RTI International)

**Session Chair:** David Kashihara (Agency for Healthcare Research and Quality)

**Redirected Inbound Call Sampling (RICS) – An  
Example of Fit for Purpose Non-Probability  
Sample Design**

Burton Levine (RTI International), Karol Krotki (RTI International)

**Benchmark Assessment of Respondent Driven  
Sampling Data for Foreign-Born Korean Americans**

Sunghye Lee (University of Michigan), Sooin Lee (Mathematica Policy Research), Michael Elliott (University of Michigan), Z. Tuba Suzer Gurtekin (University of Michigan)

**Comparing Alternative for Estimation from  
Nonprobability Samples**

Richard Valliant (Universities of Michigan & Maryland)

**Combining Probability and Nonprobability  
Samples to form Efficient Hybrid Estimates**

Jill A. Dever (RTI International)

**Discussant:** Courtney Kennedy (Pew Research Center)

10:00 - 11:45 a.m.

### Concurrent Session A-5

#### Transparent Quality Reporting in the Integration of Multiple Data Sources

Location: Rooms 147AB

Organizer: John Eltinge (U.S. Census Bureau)

Session Chair: Mary Bohman (Economic Research Service)

##### Overview of Transparent Quality Reporting

John L. Eltinge (U.S. Census Bureau)

##### Quality of Input Data

Mark Prell (Economic Research Service)

##### Quality of Data Processing

Joseph Schafer (U.S. Census Bureau)

##### Quality of Output Statistical Products and Services

Jennifer Parker (National Center for Health Statistics)

---

11:45 a.m. - 1:15 p.m.

Lunch on Your Own

---

1:15 - 3:00 p.m.

### Concurrent Session B-1

#### Advances in Web Probing

Location: Rooms 146A

Organizer and Session Chair: Paul Scanlon (National Center for Health Statistics)

##### An Automated Refusal Conversion Strategy for Web Surveys

Taylor Lewis (U.S. Office of Personnel Management), Mark Gorsak (U.S. Office of Personnel Management), Naomi Yount (Westat)

##### Using Close-Ended Web Probes to Inform the Redesign of the National Health

Paul Scanlon (National Center for Health Statistics)

##### The Roles of Typicality, Specificity, and Set Size in Providing Examples for Survey Questions: Evidence from an Online Study Using Self-Administered Comprehension Probes

Erica Yu (Bureau of Labor Statistics), Amy Swallow (Bureau of Labor Statistics)

##### Benefits and Drawbacks of Using Crowdsourcing Techniques for Cognitive Testing

Sarah Cook (RTI International), Rachel E. Morgan (Bureau of Justice Statistics), Christopher Krebs (RTI International), Lynn Langton (Bureau of Justice Statistics)

##### Impact of Dependent Interviewing on Consistency of Answers in the American Housing Survey

Katie Gustafson (U.S. Census Bureau), Evan Brassell (U.S. Census Bureau)

1:15 - 3:00 p.m.

### Concurrent Session B-2

#### Topics in Editing and Imputation: Automated Systems, Machine Learning, Hot Deck, Response Propensities, and Edit Reduction

Location: Rooms 146B

Organizer: Maria Garcia (U.S. Census Bureau)

Session Chair: Darcy Steeg Morris (U.S. Census Bureau)

##### Innovative Approaches to Generating National and Regional Estimates of Crime and Arrest with Incomplete Data

KiDeuk Kim (Urban Institute), Ashlin Oglesby-Neal (Urban Institute), Dean Obermark (Urban Institute)

##### Evaluating Hot Deck with Propensity Matching for the Advance Monthly Retail Trade Survey (MARTS)

Katherine Jenny Thompson (U.S. Census Bureau), Laura Bechtel (U.S. Census Bureau), Nicole Czaplicki (U.S. Census Bureau)

##### Assessing the Automated Imputation of Missing and Erroneous Survey Data: A Simulation-Based Approach

Larkin Terrie (Bureau of Economic Analysis)



**Edit Reduction Research in the U.S. Census Bureau's Economic Directorate**

Lisa Kaili Diamond (U.S. Census Bureau), Brian Dumbacher (U.S. Census Bureau)

**1:15 - 3:00 p.m.**

**Concurrent Session B-3**

**Measuring and Imputing Incomes**

**Location:** Rooms 146C

**Organizer:** Mark Klee (U.S. Census Bureau)

**Session Chair:** Janine McFadden (Bureau of Transportation Statistics)

**Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions**

Jeff Larrimore (Federal Reserve Board), Jacob Mortenson (Joint Committee on Taxation), David Splinter (Joint Committee on Taxation)

**Calculating a Supplemental Poverty Measure in the Survey of Income and Program Participation**

Ashley Edwards (U.S. Census Bureau), Liana Fox (U.S. Census Bureau), Lewis Warren (U.S. Census Bureau)

**Improving Earnings Imputations in the Survey of Income and Program Participation**

Gary Benedetto (U.S. Census Bureau), Rebecca Chenevert (U.S. Census Bureau), Jonathan Eggleston (U.S. Census Bureau), Mark Klee (U.S. Census Bureau), Joanna Motro (U.S. Census Bureau), Robert Munk (U.S. Census Bureau)

**Are Bracket Responses Accurate? Using IRS W-2 Records to Design Earnings Edits for the Redesigned CPS ASEC**

C. Adam Bee (U.S. Census Bureau)

**Discussant:** John Czajka (Mathematica Policy Research)

**1:15 - 3:00 p.m.**

**Concurrent Session B-4**

**Innovations in Sample Design**

**Location:** Rooms 145AB

**Organizer:** David Kashihara (Agency for Healthcare Research and Quality)

**Session Chair:** Matthew Thompson (U.S. Census Bureau)

**Sampling Design and Variance Estimation for the National Survey of WIC Participants, Wave 3**

Stanislav Kolenikov (Abt Associates), David Judkins (Abt Associates)

**Implications of Repeated Sampling in a Crime Survey**

Mike Brick (Westat), Grace Kena (Bureau of Justice Statistics), Pam Broene (Westat)

**Assessing How Well Students in Economics Experiments Model Decision-Making for Non-Student Populations**

Stephanie Rosch (Economic Research Service), Jacob Fooks (Economic Research Service), Daniel Hellerstein (Economic Research Service), Lori Lynch (Economic Research Service), Kent Messer (University of Delaware), Sharon Raszap Skorbiansky (Economic Research Service), Collin Weigel (Johns Hopkins University), Eliana Zeballos (Economic Research Service)

**Using Social Media for a Probability Sample: Is It Possible?**

Marcus Berzofsky (RTI International), Tasseli McKay (RTI International), Patrick Hsieh (RTI International), Rob Chew (RTI International), Katie Grimes (RTI International), Amanda Smith (RTI International), Natasha Latzman (RTI International), Marni Kan (RTI International), Mindy Stahl (RTI International)

**Comparing Random and Nonrandom Samples Using Model-Implied Randomization**

Vladislav Beresovsky (National Center for Health Statistics)

1:15 - 3:00 p.m.

**Concurrent Session B-5**

**Successful Approaches to Developing Data Science Capabilities in Organizations**

**Location:** Rooms 147AB

**Organizer and Chair:** Brian Moyer (Bureau of Economic Analysis)

**Moderator:**

Sally Thompson (Deputy Director, Bureau of Economic Analysis)

**Panelists:**

**Practical Implementations of Data Science in the Public Sector**

Jeff Chen (Chief Innovation Officer, Bureau of Economic Analysis)

**Preparing the Federal Statistical Workforce for Business Transformation**

Ron Prevost (Senior Statistician for Data Integration, U.S. Census Bureau)

**Tools & Technology for Embracing Data Science**

Robin Thottungal (Chief Data Scientist, US Environmental Protection Agency)

**Economics and Data Science at Hiring Lab**

Martha Gimbel (Research Director, Indeed.com Hiring Lab)

---

3:00 - 3:15 p.m.

**Break**

(Concourse)

---

3:15 - 5:00 p.m.

**Concurrent Session C-1**

**Contact Strategies**

**Location:** Rooms 146A

**Organizer and Chair:** Jennifer Beck (National Science Foundation)

**Effect of Contact Strategy on Response Mode Selection**

Yuliya Romanyuk (Statistics Canada), Cilanne Boulet (Statistics Canada)

**Topic Interest in Voting Surveys: Experimental Evidence**

Colin MacFarlane (Fors Marsh Group), Carl Turner (Fors Marsh Group), Jonathan Mendelson (Fors Marsh Group), Krysha Gregorowicz (Fors Marsh Group)

**Adaptive Designs - Tailoring of Contact Materials to Increase Survey Participation**

Rebecca J. Powell (RTI International), Antje Kirchner (RTI International), Emilia Peytcheva (RTI International)

**If We Boost It, Will They Come?: Evaluating Efforts to Promote Web Completion in a Multi-Mode National Establishment Survey**

Lauren Harris-Kojetin (National Center for Health Statistics), Christine Caffrey (National Center for Health Statistics), Melissa Hobbs (RTI International), Angela Greene (RTI International), Manisha Sengupta (National Center for Health Statistics)

**Reducing Collection Effort While Maintaining Data Quality in Business Surveys**

Jessica Andrews (Statistics Canada), Pierre Daoust (Statistics Canada), Matei Mireuta (Statistics Canada)

3:15 - 5:00 p.m.

**Concurrent Session C-2**

**Applications of Artificial Intelligence and Text Analysis in Federal Statistics**

**Location:** Rooms 146B

**Organizer and Chair:** Paul Scanlon (National Center for Health Statistics)

**A Mixed-Methods Impact Analysis of Financial Regulations on the Banking Industry**

Ricky Rambharat (Office of the Comptroller of the Currency), Douglas Robertson (Office of the Comptroller of the Currency), Roger Tufts (Office of the Comptroller of the Currency), Desiree Schaan (Office of the Comptroller of the Currency), Eugene Floyd (Office of the Comptroller of the Currency), Bianca Werner (Office of the Comptroller of the Currency), Richard Moylan (Office of the Comptroller of the Currency)

**Coding Verbatim Responses Using an Auto-coding Program Based on a Two-step Matching Process: National Hospital Ambulatory Medical Care Survey Emergency Department Data, 2015**

Akintunde Akinseye (National Center for Health Statistics), Brian W. Ward (National Center for Health Statistics)

**Comparison of Machine Learning Algorithms to Build a Predictive Model for Classification of Survey Write-in Responses**

Andrea Roberson (U.S. Census Bureau), Justin Nguyen (U.S. Census Bureau)

**Using Machine Learning Techniques to Interpret Open-ended Responses in Web Surveys**

Laura Wronski (SurveyMonkey)

3:15 - 5:00 p.m.

**Concurrent Session C-3**

**Supplemental Poverty Measures**

**Location:** Rooms 146C

**Organizer:** Erik Scherpf (Economic Research Service)

**Session Chair:** Trudi Renwick (U.S. Census Bureau)

**Moving to the Median: The Case for Changing the Range of Expenditures Underlying SPM Thresholds**

Liana Fox (U.S. Census Bureau), Thesia I. Garner (Bureau of Labor Statistics)

**Controlling for Prices before Estimating SPM Thresholds and the Impact on SPM Poverty Statistics**

Thesia I. Garner (Bureau of Labor Statistics), Juan Munoz (Bureau of Labor Statistics)

**Enhancing the Supplemental Poverty Measure to Estimate the Impact of Health Insurance Benefits on Poverty**

Dahlia K. Remler (Baruch College, CUNY), Sanders Korenman (Baruch College, CUNY), Rosemary Hyson (Baruch College, CUNY)

**Supplemental Poverty Measure: A Comparison of Geographic Adjustments with Regional Price Parities vs. Median Rents from the American Community Survey: An Update**

Trudi Renwick (U.S. Census Bureau), Bettina Aten (Bureau of Economic Analysis), Eric Figueroa (Bureau of Economic Analysis)

**Discussant:** Trudi Renwick (U.S. Census Bureau)

3:15 - 5:00 p.m.

**Concurrent Session C-4**

**Small Area Estimation: Applications and Practical Demonstrations**

**Location:** Rooms 145AB

**Organizer:** Andreea L. Erciulescu (National Institute of Statistical Sciences and National Agricultural Statistics Service)

**Chair:** Valbona Bejleri (National Agricultural Statistics Service)

**Designing the 2017 Programme for the International Assessment of Adult Competencies to Produce Direct and Indirect Estimates**

Leyla Mohadjer (Westat), Tom Krenzke (Westat), Jianzhu Li (Westat), Wendy Van de Kerckhove (Westat), Lin Li (Westat)

**Benchmarking Options for Model-Based County-Level Estimation of Agricultural Cash Rental Rates**

Michael Bellow (National Agricultural Statistics Service), Andreea L. Erciulescu (National Institute of Statistical Sciences and National Agricultural Statistics Service), Nathan Cruze (National Agricultural Statistics Service)

**Small Area Estimation in the Annual Survey of Public Employment and Payroll (U.S. Census Bureau)**

Bac Tran (U.S. Census Bureau)

**County-Level Estimates of Mortality and Natality Indicators from the National Vital Statistics System**

Lauren Rossen (National Center for Health Statistics), Diba Khan (National Center for Health Statistics)

**Bayesian State-level Estimates of Diabetes Prevalence in the United States, 2006-2015**

Diba Khan (National Center for Health Statistics), Rong Wei (National Center for Health Statistics), Yulei He (National Center for Health Statistics), Hee-Choo Shin (National Center for Health Statistics), Donald J Malec (National Center for Health Statistics)

3:15 - 5:00 p.m.

**Concurrent Session C-5**

**New Methods to Evaluate Response Error**

**Location:** Rooms 147AB

**Organizer and Chair:** Erik Scherpf (Economic Research Service)

**Measuring Systematic Income Misreporting Heterogeneity Using a Novel Dataset**

Christian Imboden (University of Oregon), John Voorheis (U.S. Census Bureau), Caroline Weber (University of Oregon)

**Respondents' Reporting Accuracy of Social Security Benefits and its Implications in the Health and Retirement Study**

Irena Dushi (Social Security Administration), Howard Iams (Social Security Administration)

**Using Administrative Records to Evaluate Absolute and Relative Reporting Accuracy in Surveys**

Joanne Pascale (U.S. Census Bureau), Kathleen Call (State Health Access Data Assistance Center), Angela Fertig (University of Minnesota), Don Oellerich (Health and Human Services)

**Understanding Differences in the Disability Prevalence Across Federal Surveys: Why the 2014 Survey of Income and Program Participation Stands Out**

Heide Jackson (U.S. Census Bureau), Danielle Taylor (U.S. Census Bureau)

**Discussant:** Bruce Meyer

7:30 a.m. – 5:00 p.m.

**Registration**  
(Concourse)

8:30 – 10:15 a.m.

**Concurrent Session D-1**

**Hierarchical Bayes Small Area Estimation  
for Domains Defined by Demography,  
Geography or Industry**

**Location:** Rooms 146A

**Organizer:** Andreea L. Erciulescu (National Institute of Statistical Sciences and National Agricultural Statistics Service)

**Session Chair:** Don Malec (National Center for Health Statistics)

**Small Domain Estimation Using Probability and  
Non-Probability Survey Data**

Adrijo Chakraborty (NORC at the University of Chicago), Nada Ganesh (NORC at the University of Chicago)

**On Increasing the Number of County-Level Crop  
Estimates**

Andreea L. Erciulescu (National Institute of Statistical Sciences and USDA National Agricultural Statistics Service), Nathan Cruze (USDA National Agricultural Statistics Service), Habtamu Benecha (USDA National Agricultural Statistics Service), Valbona Bejleri (USDA National Agricultural Statistics Service), Balgobin Nandram (Worcester Polytechnic Institute)

**Small Area Estimation for Measures Related to  
Tobacco Use and Policies Using the Tobacco Use  
Supplement to the Current Population Survey**

Benmei Liu (National Cancer Institute), Isaac Dompok (Census Bureau)

**Small Area Co-Modeling of Point Estimates and  
Their Variances for Domains in the Current  
Employment Statistics Survey**

Julie Gershunskaya (Bureau of Labor Statistics), Terrance Savitsky (Bureau of Labor Statistics)

**Discussant:** Yulei He (National Center for Health Statistics)

8:30 – 10:15 a.m.

**Concurrent Session D-2**

**Probing Nonresponse Bias in Surveys -  
Measures and Consequences**

**Location:** Rooms 146B

**Organizer and Chair:** Peter Miller (U.S. Census Bureau)

**Lessons from Nonresponse Bias Studies Involving  
Federal Surveys**

Elise Christopher (National Center for Education Statistics), Morgan Earp (Bureau of Labor Statistics), Tala Fakhouri (National Center for Health Statistics), Steven Frenk (National Center for Health Statistics), Kathryn Piscopo (Substance Abuse and Mental Health Services Administration), Peter Miller (U.S. Census Bureau)

**Nonresponse Bias Indicators and Adjustments  
Employed in Federal Surveys**

Joseph Schafer (U.S. Census Bureau), Stephanie Coffey (U.S. Census Bureau), James Dahlhamer (National Center for Health Statistics), Jeffery Gonzalez (Bureau of Labor Statistics and Peter Miller (U.S. Census Bureau)

**Alternative Indicators of the Risk for Nonresponse  
Bias**

Raphael Nishimura (Abt Associates), James Wagner (University of Michigan), Michael Elliott (University of Michigan)

**Discussant:** Roger Tourangeau (Westat)

8:30 – 10:15 a.m.

**Concurrent Session D-3**

**Novel Approaches to Using Federal and Local Data**

**Location:** Rooms 146C

**Organizer:** Jessica Graber (U.S. Census Bureau)

**Session Chair:** Amber Noel (FEMA)

**Examining the Utility of Educational Administrative Records for Research and Improving Survey Operations: A Pilot Project at the U.S. Census Bureau**

Nikolas Pharris-Ciurej (U.S. Census Bureau), Quentin Brummet (U.S. Census Bureau), Thurston Domina (University of North Carolina, Chapel Hill), Andrew Penner (University of California, Irvine), Emily Penner (University of California, Irvine), and Sonya Porter (U.S. Census Bureau)

**Leveraging Access to and Use of Department of Defense (DoD) Data: A Case Study of Unraveling Military Attrition Through New Approaches to DoD Data Integration**

Josh Goldstein, David Higdon, Sallie Keller, Bianica Pires, and Stephanie Shipp (Social and Decision Analytics Laboratory, Biocomplexity Institute of Virginia)

**Measuring Innovation in New Ways using Non-Traditional Data Sources**

Stephanie Shipp, Sallie Keller, Gizem Korkmaz, and Bianica Pires (Social and Decision Analytics Laboratory, Biocomplexity Institute of Virginia), and Carol Robbins (National Center for Science and Engineering Statistics, National Science Foundation)

**Interdisciplinary Insights for Investigating the Intersection of Race/Color and Social Outcomes among Diverse Hispanics Communities: Implications for Statistical Measurements and Analysis**

Howard Hogan PhD (U.S. Census Bureau), Nancy López PhD (University of New Mexico), Ruth Enid Zambrana PhD (University of Maryland, College Park)

**Discussant:** Amber Noel (FEMA)

8:30 – 10:15 a.m.

**Concurrent Session D-4**

**Current Research for Improving Race and Ethnicity: Federal Data Maintenance; Collection; and Presentation**

**Location:** Rooms 145AB

**Organizer and Session Chair:** John Thompson (Council of Professional Associations on Federal Statistics)

**Research Examinations within the Federal Statistical System for Improving Federal Race/Ethnicity Data**

Carolyn Hronis (Energy Information Administration)

**Update on Race and Ethnicity Question Proposals for the 2020 Census**

Nicholas Jones (U.S. Census Bureau)

**Potential Improvements for 2020 Census Race and Ethnicity Conversion and Bridging**

Benjamin Bolender (U.S. Census Bureau)

**Discussant:** Lauren Musu-Gillette (National Center for Education Statistics)

8:30 – 10:15 a.m.

**Concurrent Session D-5**

**Advancing Disclosure Limitation Methods in Federal Data Releases**

**Location:** Rooms 147AB

**Organizer and Chair:** Rolando Rodriguez (U.S. Census Bureau)

**Towards Developing Synthetic Datasets for the Economic Census**

Katherine J. Thompson (U.S. Census Bureau), Hang Kim (University of Cincinnati)

**Formal Privacy and Synthetic Data for the American Community Survey**

Michael H. Freiman (U.S. Census Bureau), Rolando Rodriguez (U.S. Census Bureau), Jerome P. Reiter (Duke University, U.S. Census Bureau), Amy Lauger (U.S. Census Bureau)

**An Integrated Approach to Providing Access to Confidential Data**

Jerry Reiter (Duke University and U.S. Census Bureau)

**Challenges and Experiences Adapting Differentially Private Mechanisms to the 2020 Census**

Simson L. Garfinkel (U.S. Census Bureau)

**Discussant:** Lisa Singh (Georgetown University)

---

**10:15 – 10:30 a.m.**

**Break**

(Concourse)

---

**10:30 a.m. – 12:15 p.m.**

**Concurrent Session E-1**

**Cognitive Lab, Usability, and Survey**

**Development Research**

**Location:** Rooms 146A

**Organizer and Chair:** Robin Kaplan (Bureau of Labor Statistics)

**Evidence-based Standards and Guidelines for Mobile Survey Instrument Design**

Lin Wang (U.S. Census Bureau), Christopher Antoun (U.S. Census Bureau), Russell Sanders (U.S. Census Bureau), Elizabeth Nichols (U.S. Census Bureau), Erica Olmsted Hawala (U.S. Census Bureau), Brian Falcone (U.S. Census Bureau), Ivonne Figueroa (U.S. Census Bureau), Jonathan Katz (U.S. Census Bureau)

**Usability Testing Methodology for the 2017 Economic Census Web Instrument**

Rebecca Keegan (U.S. Census Bureau), Temika Holland (U.S. Census Bureau), Krysten Mesner (U.S. Census Bureau), Aryn Hernandez (U.S. Census Bureau)

**Defining Bullying: A Split-Ballot Experiment Across Three Federal Agencies**

Melissa Cidade (Avar Consulting), Rachel Hansen (National Center for Education Statistics, Department of Education)

**Health Insurance in the American Community Survey: Multiple Types of Coverage and Respondent Write-ins**

Edward R. Berchick (U.S. Census Bureau), Monica S. Wiedemann (U.S. Census Bureau)

**Improving the Anchoring Vignette Methodology with Visual Vignettes**

Mengyao Hu (University of Michigan), Sunghee Lee (University of Michigan)

**10:30 a.m. – 12:15 p.m.**

**Concurrent Session E-2**

**Combating Nonresponse an Update from the Field**

**Location:** Rooms 146B

**Organizer:** Linda J. Young (National Agricultural Statistics Service)

**Session Chair:** Jennifer Beck (National Science Foundation)

**Comparing Response Rates Across Surveys**

John Dixon (Bureau of Labor Statistics), Benjamin Cover (Bureau of Labor Statistics), Kirk Hagemeyer (Bureau of Labor Statistics), Nicholas Johnson (Bureau of Labor Statistics), Andrew Kato (Bureau of Labor Statistics), Randall Powers (Bureau of Labor Statistics), Demetrio Scopelliti (Bureau of Labor Statistics), Jason Tehonica (Bureau of Labor Statistics)

**Moving Towards a User-Friendly Expenditure Diary Survey**

Safia Abdirizak (Bureau of Labor Statistics), Brett McBride (Bureau of Labor Statistics)

**Stemming the Rising Tide of Nonresponse**

Barbara Rater (National Agricultural Statistics Service), Linda J Young (National Agricultural Statistics Service)

**Cleaning Out the Gutter: Identifying and Eliminating Deadwood from a Sampling Frame Using Trees**

Andrew J. Dau (National Agricultural Statistics Service), Gavin R. Corral (National Agricultural Statistics Service), Jodie M. Sprague (National Agricultural Statistics Service), Linda J. Young (National Agricultural Statistics Service)

**Discussant:** Rosario Palmieri (OMB)

**10:30 a.m. – 12:15 p.m.**

**Concurrent Session E-3**

**Issues of Informed Consent for Using Administrative Records**

**Location:** Rooms 146C

**Organizer:** Denise Abreu (National Agricultural Statistics Service)

**Session Chair:** Darcy Miller (National Agricultural Statistics Service)

**Pursuing Consent for Record Linkage in an Establishment Survey: Results from a National Survey**

Lauren Harris-Kojetin (National Center for Health Statistics), Manisha Sengupta (National Center for Health Statistics)

**Assessing Consent Bias in Linkage Studies**

Lisa B. Mirel (National Center for Health Statistics), Cordell Golden (National Center for Health Statistics), Cindy Zhang (National Center for Health Statistics)

**Challenges to Informed Consent from Administrative Data Linkage and Secondary Usage**

Carl Ramirez (U.S. Government Accountability Office)

**10:30 a.m. – 12:15 p.m.**

**Concurrent Session E-4**

**Estimation Challenges in Complex Surveys**

**Location:** Rooms 145AB

**Organizer:** Makram Talih (National Center for Health Statistics)

**Session Chair:** Vladislav Beresovsky (National Center for Health Statistics)

**Incorporating the Finite Population Correction into the Variance Estimation of a National Business Survey**

Sadeq R. Chowdhury (Agency for Healthcare Research and Quality), David Kashihara (Agency for Healthcare Research and Quality), Matthew Thompson (U.S. Census Bureau)

**Finding an Estimator that Minimizes Revisions in a Monthly Indicator Survey**

Nicole Czaplicki (U.S. Census Bureau), Yarissa Gonzalez (U.S. Census Bureau), Laura Bechtel (U.S. Census Bureau)

**Effect of Nearest Neighbor Imputation on Variances Calculated by Fay's Balanced Repeated Replication**

Bradley Rhein (Bureau of Labor Statistics), Leland Righter (Bureau of Labor Statistics), Chester Ponikowski (Bureau of Labor Statistics)

**Investigation of the NCHS Data Presentation Standards for Proportions: A Simulation Study**

Frances McCarty, PhD (National Center for Health Statistics), Jennifer Parker, PhD (National Center for Health Statistics)

**An Alternative Way of Estimating a Proportional-Odds Model with Complex Survey Data**

Phillip S. Kott (RTI International), Peter Frechtel (RTI International)



10:30 a.m. – 12:15 p.m.

**Concurrent Session E-5**

**Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps**

**Location:** Rooms 147AB

**Organizer:** Brian Harris-Kojetin (National Academies of Sciences, Engineering, and Medicine)

**Session Chair:** Andrew Zukerberg (National Center for Education Statistics)

**Statistical Methods for Combining Multiple Data Sources**

Trivellore Raghunathan (University of Michigan)

**Quality Frameworks for Statistics Using Multiple Data Sources**

Robert Groves (Georgetown University)

**Combining Data Sources While Protecting Privacy**

Brian Harris-Kojetin (National Academies of Sciences, Engineering, and Medicine)

**Discussants:** Peter Miller (U.S. Census Bureau)  
Shelly Wilkie Martinez (Office of Management and Budget) (invited)

---

12:15 – 1:45 p.m.

**Lunch on Your Own**

---

1:45 – 3:30 p.m.

**Concurrent Session F-1**

**Question Evaluation and Cognitive Interviewing**

**Location:** Rooms 146A

**Organizer:** Kristen Miller (National Center for Health Statistics)

**Chair:** Jennifer Edgar (Bureau of Labor Statistics)

**Cognitive Interviewing Methodology in 2018: Current Trends and Recent Challenges**

Kristen Miller (National Center for Health Statistics)

**Results of a Cognitive Interview Evaluation of the Revised Race Question, with Special Emphasis on the Newly Proposed Middle Eastern/North African Response Option**

Stephanie Willson (National Center for Health Statistics), Sheba Dunston (National Center for Health Statistics), Merarys Rios (U.S. Census Bureau)

**Training Cognitive Testing Interviewers in Different Settings and Languages**

Alisú Schoua-Glusberg (Research Support Services and IMPAQ International)

**Survey Translation and the Place of Expert Review in the Question Evaluation Tool Kit: Development of Best Practices**

Patricia Goerman (U.S. Census Bureau), Mikelyn Meyers (U.S. Census Bureau), Yazmin García Trejo (U.S. Census Bureau)

**Recently Resettled Refugees and their Experience with the Annual Survey of Refugees – Findings from Cognitive and In-Depth Interviews**

David Dutwin (SSRS), Hamutal Bernstein (Urban Institute), Susan Sherr (SSRS), Arina Goyle (SSRS), Robert Santos (Urban Institute), Nicole Deterding (Business Strategy Consultants), Erica Meade (U.S. Department of Health and Human Services)

1:45 – 3:30 p.m.

**Concurrent Session F-2**

**Conducting Randomized Experiments in Establishment Surveys**

**Location:** Rooms 146B

**Organizers:** Diane K. Willimack (U.S. Census Bureau), Jaki S. McCarthy (USDA / National Agricultural Statistics Service)

**Session Chair:** David Kinyon (Energy Information Administration)

**Conducting Experiments in Establishment Surveys: Obstacles and Opportunities**

Jaki S. McCarthy (USDA National Agricultural Statistics Service), Diane K. Willimack (U.S. Census Bureau)

**Obstacles in Planning Establishment Survey Experiments - Census of Agriculture Content Test and Agricultural Resource Management Survey**  
Kathy Ott (USDA National Agricultural Statistics Service), Tyler Wilson (USDA National Agricultural Statistics Service)

**Using Email to Solicit Response in an Establishment Survey**  
Joshua Langeland (Bureau of Labor Statistics)

**Experimenting with Alternative Question Designs for "Other, Specify" Product Information in Establishment Surveys**  
Diane K. Willimack (U.S. Census Bureau), Kevin Linares (formerly U.S. Census Bureau), Brian Kriz (JPSM/UMd alumnus)

**Discussant:** Jennifer Beck (National Science Foundation)

**1:45 – 3:30 p.m.**

### **Concurrent Session F-3**

#### **Extending Research through Access and Application of Federal Data**

**Location:** Rooms 146C

**Organizer and Chair:** Lynn Langton (Bureau of Justice Statistics)

**Researching the Psychology Workforce using Federal Statistics**  
Luona Lin (American Psychological Association), Karen Stamm (American Psychological Association), and Peggy Christidis (American Psychological Association)

**Intersectionality and Quantitative Methods for Better Serving Vulnerable Communities by Understanding the Simultaneity of Race-Gender-Class Inequalities: College Graduation and Income for College Graduates in New Mexico**  
Nancy López PhD (University of New Mexico), Michael O'Donnell JD, MS (University of New Mexico), Carmela Roybal MA (University of New Mexico), Lucas Pedraza MA (University of New Mexico), Jeffrey Mitchell PhD (Bureau of Business and Economic Research, BBER, University of New Mexico)

**Data Linking and Large Scale Assessments: Linking PIAAC and BLS Data for Labor Policy**  
Katie Seely-Gant (Energetics Technology Center), Lisa M. Frehill (Energetics Technology Center)

**Examining the Principles of Open Government**  
Jennifer Goode, Ph.D. (U.S. Census Bureau), Marie-Claude Jiguet-Akhtar, Ph.D. (Howard University)

**1:45 – 3:30 p.m.**

### **Concurrent Session F-4**

#### **Improvements in Data Collection Methodology**

**Location:** Rooms 145AB

**Organizer and Chair:** Andrew White (National Center for Education Statistics)

**Cost Effective Mail Survey Design**  
Doug Williams (Westat), Sherm Edwards (Westat), Pamela Giambo (Westat), Grace Kena (Bureau of Justice Statistics)

**Integration of Multiple Data Sources to Inform a Responsive Design**  
Peter Siegel (RTI International), Jennifer Wine (RTI International)

**Optimal Sample Size Allocation to Mixed Modes: A Case Study Using the Residential Energy Consumption Survey**  
Peter Frechtel (RTI International), Phillip S. Kott (RTI International)

**Design Issues for a Longitudinal Employer Health Insurance Survey to Facilitate Analysis of Policy Changes**  
Steven Machlin (Agency for Healthcare Research and Quality), David Kashihara (Agency for Healthcare Research and Quality)

**1:45 – 3:30 p.m.**

**Concurrent Session F-5**

**Next Steps in Advancing the  
Recommendations of the Commission on  
Evidence-Based Policy Making**

**Location:** Rooms 147AB

**Organizer and Session Chair:** John Thompson (Council  
of Professional Associations on Federal Statistics)

**Panel Discussion**

**Nick Hart**, Bipartisan Policy Center

**Bruce Meyer**, the McCormick Foundation Professor  
at the University of Chicago Harris School of Public  
Policy, and Former Commissioner

**Shelly Martinez**, Office of Management and  
Budget

**John Righter**, Deputy Staff Director, Office of  
Senator Patty Murray

**Katy Rother**, Counsel, Office of Representative  
Mark Meadows

---

**3:30 – 4:00 p.m.**

**Break**

(Concourse)

---

**4:00 – 5:30 p.m.**

**PLENARY SESSION II**

**Location:** Rooms 146A and 146B and 146C

**Keynote Panel:  
Initiatives to Modernize the Federal  
Statistical System**

**Keynote Panel Moderator:**  
**Nancy Potok** (Chief Statistician of the  
United States)

**Additional Panelists Include:**

**Mary Bohman** (Economic Research Service)

**Jeri Mulrow** (Bureau of Justice Statistics)

**Sally Thompson** (Bureau of Economic Analysis)

**William Wiatrowski** (Bureau of Labor  
Statistics)

7:30 a.m. – 5:00 p.m.

**Registration**  
(Concourse)

8:30 – 10:15 a.m.

**Concurrent Session G-1**  
**Challenges and Solutions to Collecting**  
**Information on Sexual Violence**

**Location:** Rooms 146A

**Organizer and Chair:** Shannan Catalano (Bureau of Justice Statistics)

**Improving the Measurement of Sexual**  
**Victimization Among Children through a Redesign**  
**of the National Survey of Children’s Exposure to**  
**Violence**

J. Michael Brick (Westat), Jennifer Bronson (Bureau of Justice Statistics), David Finkelhor (University of New Hampshire), Darby Steiger (Westat), Brecht Donoghue (Office of Juvenile Justice & Delinquency Prevention)

**Key Issues When Collecting and Publishing**  
**General Population Estimates of Rape and Sexual**  
**Assault**

David Cantor (Westat), Darby Steiger (Westat), John Hartge (Westat), Shannan Catalano (Bureau of Justice Statistics)

**Collecting Data on Rape and Sexual Assault in an**  
**Institutional Setting**

Jessica Stroop (Bureau of Justice Statistics)

**Discussant:** Howard Snyder (Subject Matter Expert)

8:30 – 10:15 a.m.

**Concurrent Session G-2**  
**Innovations in Measuring and Reducing**  
**Respondent Burden**

**Location:** Rooms 146B

**Organizer and Chair:** Heather Ridolfo (National Agricultural Statistics Service)

**Assessing Respondents’ Perceptions of Burden in**  
**the American Community Survey**

Jessica Holzberg (U.S. Census Bureau), Jonathan Katz (U.S. Census Bureau), Gerson Morales (U.S. Census Bureau), Mary Davis (U.S. Census Bureau)

**Estimating Reporting Burden for Statistical**  
**Surveys**

Anna Hamelin (Energy Information Administration), Brian Hewitt (Energy Information Administration)

**Exploring Sampling Techniques to Reduce**  
**Respondent Burden**

Yijun Wei (National Agricultural Statistics Service), Valbona Bejleri (National Agricultural Statistics Service)

**Response Likelihood to an Establishment Survey**  
**with a Simple Questionnaire Following an**  
**Establishment Survey with a Complex**  
**Questionnaire**

Joseph Rodhouse (National Institute of Statistical Sciences), Tyler Wilson (National Agricultural Statistics Service), Heather Ridolfo (National Agricultural Statistics Service)

**Discussant:** Michael Planty (RTI International)

8:30 – 10:15 a.m.

### Concurrent Session G-3

#### Bayesian Methods for Surveys

**Location:** Rooms 146C

**Organizer:** Nathan B. Cruze (USDA National Agricultural Statistics Service)

**Session Chair:** Kimberly Henry (Internal Revenue Service)

##### Bayesian Estimation Under Informative Sampling with Unattenuated Dependence

Matthew R Williams (Substance Abuse and Mental Health Services Administration), Terrance D. Savitsky (U.S. Bureau of Labor Statistics)

##### Scalable Bayes Clustering for Outlier Detection Under Informative Sampling

Terrance Savitsky (Bureau of Labor Statistics)

##### Some Exercises in Covariate Selection for a Bayesian Crop Yield Forecasting Model

Habtamu Benecha (USDA National Agricultural Statistics Service), Nathan B. Cruze (USDA National Agricultural Statistics Service)

##### Are We Under-Estimating Food Insecurity?

Christian Gregory (Economic Research Service, USDA)

**Discussant:** Nathan B. Cruze (USDA National Agricultural Statistics Service)

8:30 – 10:15 a.m.

### Concurrent Session G-4

#### Combining Different Levels of Data

##### Sources

**Location:** Rooms 145AB

**Organizer:** Denise Abreu (National Agricultural Statistics Service)

**Session Chair:** Glendon Haynes (Energy Information Administration)

#### Potential Uses of Individual-level Administrative Records Data in the 2012 Survey of Business Owners

Adela Luque (U.S. Census Bureau, presenter), Renuka Bhaskar (U.S. Census Bureau), Sharon Ennis (U.S. Census Bureau), James Noon (U.S. Census Bureau)

#### Linking USGS Water Use Data to Detailed Industries for Environmental Input-Output Modeling of the U.S. Food System

Sarah Rehkamp (USDA Economic Research Service), Patrick Canning (USDA Economic Research Service), Catherine Birney (University of Texas-Austin)

#### Incorporating OSHA Administrative Records in the Survey of Occupational Injuries and Illnesses

Brooks Pierce (U.S. Bureau of Labor Statistics)

#### Data Linkage with an Establishment Survey

Jennifer Sayers (National Center for Health Statistics), Scott Campbell (NORC), Clinton Thompson (National Center for Health Statistics), Geoff Jackson (National Center for Health Statistics)

#### Evaluation of Health Care Event Reporting in a National Household Survey Using Medical Provider Data

Jerrod Anderson (Agency for Healthcare Research and Quality), Emily Mitchell (Agency for Healthcare Research and Quality), Adam Biener (Agency for Healthcare Research and Quality)

8:30 – 10:15 a.m.

### **Concurrent Session G-5**

#### **Disclosure Limitation and Privacy Protection**

**Location:** Rooms 147AB

**Organizer:** David Kinyon (Energy Information Administration)

**Session Chair:** Ray Popovic (U.S. Census Bureau)

##### **Marrying Demand for Statistical Information with Disclosure Control: The Canadian Experience in Developing an Automated Dissemination Tool in an Open-Data World**

Zixin Nie (Statistics Canada), Claude Girard (Statistics Canada)

##### **Disclosure Control and Random Tabular Adjustment**

Mark Stinner (Statistics Canada)

##### **A Note on Multiplicative Noise Perturbation for Privacy Protection**

Xiaoyu Zhai (Department of Statistics, George Washington University), Tapan Nayak (Department of Statistics, George Washington University)

---

10:15 – 10:30 a.m.

#### **Break**

(Concourse)

---

10:30 a.m. – 12:15 p.m.

### **Concurrent Session H-1**

#### **New Uses of Combining Public and Private Data**

**Location:** Rooms 146A

**Organizer and Chair:** Andrew White (National Center for Education Statistics)

##### **Linking a Retail Gasoline Price Survey with Commercial Data**

Maura Bardos (Energy Information Administration), Amerine Woodyard (Energy Information Administration), Jeramiah Yeksavich (Energy Information Administration)

##### **Utility of Open Information Sources to Develop Estimates of Selected Crime and Justice**

##### **Indicators: Measuring Arrest-Related Deaths**

Duren Banks (RTI International), Michael Planty (RTI International), G. Lance Couzens (RTI International)

##### **Evaluation of Vendor School and Teacher Lists for the 2015-16 National Teacher and Principal Survey**

Maura Spiegelman (National Center for Education Statistics), David Sheppard (U.S. Census Bureau), Quentin Brummet (U.S. Census Bureau)

10:30 a.m. – 12:15 p.m.

### **Concurrent Session H-2**

#### **Innovative Uses of R in Federal Agencies**

**Location:** Rooms 146B

**Organizer and Session Chair:** Wendy Martinez (Bureau of Labor Statistics)

##### **FDA's Approach to R Shiny Standardized, Interactive Tools**

Jimmy Wong (U.S. FDA)

##### **NHANES Weights: Use of R for Data Visualization, QA/QC, and Calibration**

Zipf, George (National Center for Health Statistics)

##### **R: Innovating at the Bureau of Labor Statistics**

Arcenis Rojas (Bureau of Labor Statistics)

**Discussant:** Wes Chou (Bureau of Labor Statistics)

10:30 a.m. – 12:15 p.m.

### Concurrent Session H-3

#### Statistical Methods and Testing

**Location:** Rooms 146C

**Organizer:** Grace Kena (Bureau of Justice Statistics)

**Session Chair:** Rachel Dinkes (American Institutes for Research)

##### **The Effect of the Conservation Reserve Program on Rural Economies: Deriving a Statistical Verdict from a Null Finding**

Timothy Wojan (Economic Research Service), Jason Brown (Kansas City Federal Reserve Bank), Dayton Lambert (University of Tennessee)

##### **State-Level Design-Based Estimates for National Surveys**

Stephen Ash (U.S. Census Bureau), Brian Shaffer (U.S. Census Bureau)

##### **Multivariate Small Area Estimation Under Informative Sampling and Nonresponse**

Michael Sverchkov (Bureau of Labor Statistics), Danny Pfeffermann (Bureau of Labor Statistics)

##### **Semiparametric Panel Data Models Using Neural Networks**

Andrew Crane-Droesch (U.S. Department of Agriculture/Economic Research Service)

##### **Testing Significance Tests: A Simulation with Cliff's Delta, $t$ -tests, and Mann-Whitney $U$**

Tyler Barnes (Department of Veterans Affairs/Wright State University), Scott C. Moore (Department of Veterans Affairs), Katerine Osatuke (Department of Veterans Affairs)

10:30 a.m. – 12:15 p.m.

### Concurrent Session H-4

#### Administrative Records for Sampling Efficiency

**Location:** Rooms 145AB

**Organizer:** Denise Abreu (National Agricultural Statistics Service)

**Session Chair:** Wendy Barboza (National Agricultural Statistics Service)

##### **Sampling with Administrative Records in the National Survey of Children's Health**

Scott Albrecht (U.S. Census Bureau), Jason Fields (U.S. Census Bureau), Keith Finlay (U.S. Census Bureau)

##### **The Use of Administrative Records and the American Community Survey to Study the Characteristics of Undercounted Young Children in the 2010 Census**

Leticia Fernandez (U.S. Census Bureau), Rachel Shattuck (U.S. Census Bureau), James Noon (U.S. Census Bureau)

##### **Logistic Regression with Linked Data**

Abdelnasser Saïdi (Statistics Canada), Kenneth Chu (Statistics Canada), Abel Dasylva (Statistics Canada), Félix Labrecque-Synnott (Statistics Canada)

##### **A Method for Assigning Weight to Variable Matching in Record Linkage**

Salam Abdus (Agency for Healthcare Research and Quality), Steven C. Hill (Agency for Healthcare Research and Quality), Marc I. Roemer (Agency for Healthcare Research and Quality)

##### **Application of Jaro-Winkler String Comparator in Enhancing Veterans Administrative Records**

Hyo Park (Veterans Affairs), Eddie Thomas (Veterans Affairs), Pheakdey Lim (Veterans Affairs)

10:30 a.m. – 12:15 p.m.

**Concurrent Session H-5**

**Research Priorities for Enhancing Quality  
in the Integration of Multiple Data**

**Sources**

**Location:** Rooms 147AB

**Organizer:** John Eltinge (U.S. Census Bureau)

**Session Chair:** Alex Brown (Joint Program in Survey  
Methodology)

**Panelists:** To be determined

---

12:15 – 1:45 p.m.

**Lunch on Your Own**

---

1:45 – 3:30 p.m.

**Concurrent Session I-1**

**Re-Engineering and Modernizing the 2017  
Economic Census**

**Location:** Rooms 146A

**Organizer:** Diane K. Willimack (U.S. Census Bureau)

**Session Chair:** Ron Jarmin (U.S. Census Bureau)

**An Overview of the Improved 2017 Economic  
Census**

Kimberly Moore (U.S. Census Bureau), William  
Samples (U.S. Census Bureau)

**Executing a Multi-Year Multi-Method Electronic  
Data Collection Re-engineering: Experiences from  
2017 Economic Census Development**

Amy E. Anderson Riemer (U.S. Census Bureau)

**2017 Economic Census Contact Strategy: Using  
Data to Make Decisions**

Susanne Johnson (U.S. Census Bureau)

**From Research to Implementation of Product  
Estimation in the 2017 Economic Census: Hard,  
Harder, and Hardest**

William C. Davie Jr. (U.S. Census Bureau), Scot Dahl  
(U.S. Census Bureau), Katherine Jenny Thompson  
(U.S. Census Bureau)

**Discussant:** Adam Safir (Bureau of Labor Statistics)

1:45 – 3:30 p.m.

**Concurrent Session I-2**

**Record Linkage Techniques to Enhance  
Criminal Justice and Immigration Statistics**

**Location:** Rooms 146B

**Organizer:** Christopher Cutler (Abt Associates)

**Session Chair:** Lynn Langton (Bureau of Justice  
Statistics)

**Record Linkage of Bureau of Justice (BJS) Federal  
Criminal Case Processing Data**

Ryan Kling (Abt Associates), Christopher Cutler (Abt  
Associates), Mark Motivans (Bureau of Justice  
Statistics)

**Record Linkage Application in DHS USCIS Person  
Centric System**

Damian Kostiuik (U.S. Citizenship and Immigration  
Services)

**Evaluation of Data Matching in Immigration  
Enforcement Outcome Tool**

Hongwei Zhang, Katherine Shanahan (Department  
of Homeland Security, Office of Immigration  
Statistics)

**Data Matching Practice and Application in DHS  
CBP Arrival and Departure Information System  
(ADIS)**

Michael Gorman (U.S. Customs and Border  
Protection)

**Discussants:** Jiashen You (DHS, Office of  
Immigration Statistics)  
Ryan Kling (Abt Associates)



1:45 – 3:30 p.m.

**Concurrent Session I-3**

**New Advances in Applications of Geospatial Technology**

**Location:** Rooms 146C

**Organizer and Chair:** Darius Singpurwalla (National Science Foundation)

**Mapping Geographic and Temporal Variations in Select Mortality and Natality Outcomes with R-INLA in Small Areas**

Diba Khan (National Center for Health Statistics), Lauren M. Rossen (National Center for Health Statistics), Brady Hamilton (National Center for Health Statistics), Yulei He (National Center for Health Statistics), Holly Hedegaard (National Center for Health Statistics), Rong Wei (National Center for Health Statistics), Margaret Warner (National Center for Health Statistics)

**Crossing Boundaries: A Case Study on Building Accessible Tools to Combine Public-Use Data**

Matthew Graham (U.S. Census Bureau)

**Evaluation of Two Different Interviewing Protocols to Test a Mobile Mapping Instrument for the June Area Survey**

Denise Abreu (National Agricultural Statistics Service), Michael Hyman (National Agricultural Statistics Service), Linda A. Lawson (National Agricultural Statistics Service), Sonia Hickman (National Agricultural Statistics Service)

**The Time Use Data-based Measures of the Wellbeing Effect of Community Development: An Evaluative Approach.**

Włodzimierz Okrasa (Central Statistical Office, Poland). Dominik Rozkrut (Central Statistical Office, Poland).

**Sidestepping the Box: Designing a Supplemental Poverty Indicator for School Neighborhoods**

Doug Geverdt (National Center for Education Statistics), Laura Nixon (U.S. Census Bureau)

1:45 – 3:30 p.m.

**Concurrent Session I-4**

**Analyzing Nonresponse – New Research**

**Location:** Rooms 145AB

**Organizer and Chair:** Alice Yu (Bureau of Labor Statistics)

**Session Chair:** John Ahearn (Bureau Labor of Statistics)

**Correlates of Nonresponse in the 2012 and 2014 Medical Expenditure Panel Survey**

Frances Chevarley (Agency for Healthcare Research and Quality), William Mosher (Johns Hopkins Bloomberg School of Public Health)

**Subnational Geography and the Overseas U.S. Citizens Population**

Carl Turner (Fors Marsh Group), Colin MacFarlane (Fors Marsh Group), Krysha Gregorowicz (Fors Marsh Group), Jonathan Mendelson (Fors Marsh Group)

**Investigation into Responses for the Occupational Requirements Survey**

Alice Yu (Bureau of Labor Statistics), Erin McNulty (Bureau of Labor Statistics)

**Understanding Their Apprehension: A Look into the Federal Employee Viewpoint Survey**

Mark Gorsak (Office of Personnel Management), Taylor Lewis (Office of Personnel Management), Judah Frank (Office of Personnel Management)

**Discussant:** John S. Dixon (Bureau Labor of Statistics)

1:45 – 3:30 p.m.

**Concurrent Session I-5**

**Advances in Evaluating Labor Force Statistics**

**Location:** Rooms 147AB

**Organizer:** Robin Kaplan (Bureau of Labor Statistics)

**Session Chair:** Jeffrey Gonzalez (Bureau of Labor Statistics)

**Creating State Specific Occupational Replacement Rates**

Alex Roubinchtein (Washington State Employment Security Department)

**Employment Transitions in Washington State – An Examination of Job Flows Among Persons Employed in Growing and Declining Industries, 2014-2015**

Jonathan Adam Lind (Washington State Employment Security Department)

**Using Administrative Records for Survey Item Replacement: An Analysis of the LEHD-NSCG Annual Earnings Ratio**

Michaela Dillon (U.S. Census Bureau)

**Discussant:** Michael Wolf (Bureau of Labor Statistics)

---

3:30 – 3:45 p.m.

**Break**

(Concourse)

---

3:45 – 5:15 p.m.

**Concurrent Session J-1**

**Progress in Measuring Income Statistics**

**Location:** Rooms 146A

**Organizer and Chair:** George Sheldon (U.S. Department of Veterans Affairs)

**Measuring Trends in the Distribution of Annual Earnings**

David Pattison (Social Security Administration)

**Creating a Comprehensive Income Dataset**

Bruce D. Meyer (University of Chicago, AEI, NBER)

**What Does Consumer Heterogeneity Mean for Measuring Changes in the Cost of Living?**

Robert S. Martin (Bureau of Labor Statistics)

**Discussant:** George Sheldon (U.S. Department of Veterans Affairs)

3:45 – 5:15 p.m.

**Concurrent Session J-2**

**Forecasting Methods**

**Location:** Rooms 146B

**Organizer and Chair:** Darius Singpurwalla (National Science Foundation)

**An Assessment of Crime Forecasting Models**

Bruno Gasperini (IMPAQ International), Aaron Heuser (IMPAQ International), Minh Huynh (IMPAQ International), Hautahi Kingi (IMPAQ International), James Moore (IMPAQ International), Chriz Zhang (IMPAQ International)

**Jointly Predicting U.S. Recessions and Restaurant Downturns: Integrated Models using the ERS Food Expenditure Series**

Timothy Park (Economic Research Service), Howard Elitzak (Economic Research Service, USDA), Abigail Okrent (Economic Research Service, USDA)

**Clarifying the Confidence Levels of PPI and CPI Forecasts in the USDA's Food Price Outlook**

David Levin (USDA – Economic Research Service)

3:45 – 5:15 p.m.

**Concurrent Session J-3**

**Ensuring Data Quality and High Response Rates on Federally-Funded Establishment Data Collections**

**Location:** Rooms 146C

**Organizer and Chair:** Chris Ellis (RTI International)

**Ensuring Optimal Response Rates on Agency Data Collections Over Time**

Susan Brumbaugh (RTI International), Elizabeth Smith (RTI International), Chris Ellis (RTI)

International), Mary Cowhig (Bureau of Justice Statistics), Zhen Zeng (Bureau of Justice Statistics)

**Reasons for Late Response and Nonresponse in Surveys of Government Agencies**

Tim Smith (RTI International), Christian Genesky (RTI International), Danielle Kaebler (Bureau of Justice Statistics), Anthony Whyde (Bureau of Justice Statistics)

**Shifting Data Quality Follow-Up Methods for a Time-Series Collection of Local and State Agencies**

Chris Ellis (RTI International), Scott Ginder (RTI International), Megan Waggy (RTI International), Mary Cowhig (Bureau of Justice Statistics)

**Discussant:** Ann Carson (Bureau of Justice Statistics)

**3:45 – 5:15 p.m.**

**Concurrent Session J-4**

**Innovations in Survey Design**

**Location:** Rooms 145AB

**Organizer and Session Chair:** Jessica Graber (U.S. Census Bureau)

**Collecting Electronic Health Record Data for the National Ambulatory Medical Care Survey and the National Hospital Care Survey**

Carol DeFrances, Ph.D. (National Center for Health Statistics), Denys T. Lau, Ph.D. (National Center for Health Statistics)

**Data Integration Innovations to Enhance Analytical Capacity**

Steven B. Cohen (RTI International)

**National Health Interview Survey 2019 Content Redesign – Analytic Implications**

Sarah Lessem (National Center for Health Statistics), Renee Gindi (National Center for Health Statistics), Stephen Blumberg (National Center for Health Statistics), Aaron Maitland (National Center for Health Statistics), Ben Zablotzky (National Center for Health Statistics), Lindsey Black (National Center for Health Statistics), Emily Zammiti (National Center for Health Statistics)

**Measuring Program Knowledge Longitudinally: The Uses and Methodological Challenges for Federal Agencies**

Matt Messel (Social Security Administration), Mark Sarney (Social Security Administration), David Rogofsky (Social Security Administration), Laith Alattar (Social Security Administration)

**Balancing Cross-sectional and Longitudinal Design Objectives for the Survey of Doctorate Recipients**

Wan-Ying Chang (National Center for Science Engineering Statistics), Patricia Ruggles (National Center for Science Engineering Statistics)

## Plenary Session I

Wednesday, March 7, 4:00 p.m.

Room 146ABC

**Keynote:** Combining Data and Crime  
**Keynote speaker:** **Sharon Lohr**, Professor Emerita, Arizona State University

One hundred years ago, crime statistics were collected differently by each community---if collected at all. After a brief tour of the history of US crime statistics, I evaluate the strengths and shortcomings of today's data sources for homicide, violent and property crime, and fraud. Where do we have high quality data, and how might various sources be leveraged and combined to provide a better picture of crime? What are the error properties of combined estimates? What are some possible unintended consequences of combining data sources?

---

## Plenary Session II

Thursday, March 8, 4:00 p.m.

Room 146ABC

**Keynote Panel:** Initiatives to Modernize the Federal Statistical System  
**Panel Moderator:** **Nancy Potok**, Chief Statistician of the United States

Panelists Include:

**Mary Bohman** (Economic Research Service)

**Jeri Mulrow** (Bureau of Justice Statistics)

**Sally Thompson** (Bureau of Economic Analysis)

**William Wiatrowski** (Bureau of Labor Statistics)

In early 2017, the Interagency Council on Statistical Policy (ICSP) committed to focusing attention and resources on modernizing the Federal statistical system and identified six strategic priorities. These are: 1) researching methodology for assessing and assuring blended data quality; 2) improving data access for researchers through the Federal Statistical Research Data Centers (FSRDCs); 3) employee development; 4) improving intergovernmental relationships around data sharing; 5) improving survey response rates; and 6) communicating the value of Federal statistics.

This session will focus on four of the six priorities – data quality, employee development, FSRDCs, and communication. Statistical agency leaders will discuss modernization across the Federal statistical system; and next steps for ICSP member agencies and others within the Federal Government in order to achieve tangible successes. The ICSP leaders will participate in a panel discussion, led by the U.S. Chief Statistician, around the strategic priority initiatives.

# Concurrent Session A-1

## Advances in Web Scraping for Federal Statistics

### **Web-scraping & Similar Tools to Capture Information for BLS Uses**

David Friedman (Bureau of Labor Statistics, Associate Commissioner-Office of Price and Living 38 Conditions) & Bill Thompson (Bureau of Labor Statistics/Producer Price Index)

Web-scraping is the process of extracting data from websites, typically through the use of a software program that simulates human exploration of a website(s). Obviously, the amount and type of information on the internet offers ample opportunities for web-scraping, but as with most uses of alternative data sources it also raises many operational, logistical, and, yes, legal questions. While BLS is making strides towards leveraging the opportunities in research projects and in a couple of situations targeted toward production/estimation, there are concerns/issues that are challenging (troubling?) most if not all the efforts. This presentation will focus on three BLS experiences that illustrate the potential of web-scraping. The presentation will also identify obstacles including informed consent and other similar challenges presented by web-scraping. In addition, based on surveys of BLS offices and programs that have either used web-scraping or similar tools, are currently using them, or are considering using them or expanding use of them in the future, the presentation will identify recommendations on how to overcome or mitigate obstacles encountered with web-scraping.

### **SABLE: Tools for Web Crawling, Web Scraping, and Text Classification**

Brian Dumbacher (U.S. Census Bureau), Lisa Kaili Diamond (U.S. Census Bureau)

For economic surveys conducted by the U.S. Census Bureau, respondent data or equivalent-quality data can sometimes be found online on respondent websites and government websites. An automated process for finding useful data sources and then scraping the data is ideal but challenging to develop. Websites and the documents on them have different structures, so a long-term solution needs to be able to adapt to new situations.

To this end, Census Bureau researchers are developing a collection of tools for web crawling and web scraping known as SABLE, which stands for Scraping Assisted By LEarning. Elements of SABLE involve machine learning to perform text classification and autocoding. SABLE is based on two key pieces of open-source software: Apache Nutch and Python. This presentation gives an overview of SABLE and describes research to date, potential applications to economic surveys, efforts in moving to a production environment, and future work.

**Evaluating the Use of Web-Scraped List Frames to Assess Undercoverage in Surveys**

Linda J. Young (USDA National Agricultural Statistics Service)

For the census of agriculture, NASS has traditionally used the June Area Survey (JAS) sample, which is drawn from its area frame, to account for undercoverage of the NASS list frame. Beginning in 2012, capture-recapture methods were used for this purpose, due to the presence of misclassification in the JAS sample. The types of farms for which the coverage of the list frame is a concern are not well represented in the JAS sample. Thus, NASS has been evaluating the use of web-scraped list frames as a second frame from which a sample could be drawn to assess undercoverage for surveys. Capture-recapture methods based on samples from the NASS list frame and a web-scraped list frame were used for the 2015 Local Foods Marketing Survey. Here the assumptions underpinning these methods are evaluated. To the extent possible, data from the local foods survey are used to examine whether the assumptions are met. Simulation studies are used to assess the effect of violations of the assumptions on estimates. Alternative methods that are robust to departures from the assumptions are evaluated.

**Seizing Opportunities and Preserving Principles**

Bob Sivinski (OMB)

Web scraping and other methods of capturing wild data presents opportunities for increasing the scope, timeliness, and efficiency of federal statistical products. Of course, taking advantage of these new opportunities presents challenges with regard to methodology, interpretation, risk management, and statistical policy. This presentation considers some of the policy and ethical challenges inherent in the use of data that, unlike most survey data, isn't drawn probabilistically from a well-understood population.

The ethical and policy challenges around the use of web scraped data are the same challenges that statisticians have grappled with for many years; bias, representativeness, measurement and communication of uncertainty, informed consent, data quality, reproducibility, etc. are all well-trod ground for the producers of federal statistics. In order to responsibly take advantage of emerging opportunities presented by new data sources, producers of federal statistics need to give these issues the same level of attention that they have historically been given in the realm of sample surveys. This may require the development of new best practices and standards for transparency and validation, designed specifically for new sources of data.

**Museum Frame Development – A Universe is Comprised of Many Worlds**

Lisa M. Frehill (Institute of Museum and Library Services), Jason Enos (Institute of Museum and Library Services), Matthew Birnbaum (Institute of Museum and Library Services)

Under 20 USC Chapter 72 § 9108, the Institute of Museum and Library Services (IMLS) has the authority to “conduct policy research, analysis, and data collection to extend and improve the Nation’s museum, library, and information services.” While IMLS has two long-running statistical collections associated with the nation’s libraries, in order to develop a similar research collection for museums necessitates development of an appropriate survey frame. This paper builds on an IMLS legacy research project, which used multiple strategies to develop a universe frame for museums and related organizations, with the most recent update in FY 2015. As with other establishment frames, the museum frame needs to be updated and additional information gathered to differentiate if file entries conform to a standard

definition of museums, and if not, whether the entry represents a “museum-related organization,” or represents an out-of-scope establishment.

The paper will describe the ways in which the research team replicated and expanded upon web-scraping methods described by Rhodes, Kim and Loomis (2015), which generated a survey frame of retail establishments that provide electronic nicotine delivery systems. Our study focuses on generating the survey frame for children’s museums, in particular. Analyses to be presented assess the reliability and efficacy of various methods of frame generation for a universe that includes both non-profit and government-operated establishments. Metrics associated with methodological efficiency and effectiveness (e.g., number of accurate entities identified) are presented. These analyses also illustrate the limitations associated the use of IRS 990 data, which has become a popular “big data” method for studying non-profit institutions, such as museums. Finally, the paper concludes with guidelines for generating survey frames for non-profit and government-operated organizations.

## Concurrent Session A-2

### Does Survey Response Change Respondent Behavior?

#### **Speeders and Sloths: The Effects of Respondent Timing on Web Survey Data Quality**

Michael Planty (RTI International), Lynn Langton (Bureau of Justice Statistics), Marcus Berzofsky (RTI International), Christopher Krebs (RTI International), Christine Lindquist (RTI International)

In this paper we describe and assess the impact of response time on the quality of estimates for a web-based survey. Respondents who spend too little or too much time on the survey or survey sections introduce potential measurement error due to a lack of attention, or proper comprehension of the items and response categories (Greszki, 2015). Short or very long response times can be a potential indicator of low data quality either due to measurement error when there is a response given, or nonresponse bias and variance inflation when a response is not given. The Campus Climate Survey Validation Study (CCSVS) was a survey administered to 23,000 college students at nine U.S. institutions on the prevalence and characteristics of sexual violence. We first describe the distribution of respondent times, identifying short (speeders) and long (sloths) times. We examine the characteristics of these groups including the point during the field period when speeding is likely to occur, the time of day and day of the week speeders are likely to respond, whether mode (type of device) impacts the likelihood of speeding, and across-school variation in the prevalence of speeding. We then quantify the impact on point estimates and precision (relative standard errors) when these outliers are removed using various cut-points. In addition, the CCSVS randomly assigned monetary incentives to respondents. We assess the impact that varying incentive amounts have on respondent timing with the assumption that respondents are more compelled to provide better quality responses (i.e., are less likely to speed) when the incentive is higher. The paper concludes by discussing these implications and limitations associated with conducting self-administered, web-based surveys.

#### References:

GRESZKI, R; MEYER, M; SCHOEN, H. EXPLORING THE EFFECTS OF REMOVING "TOO FAST" RESPONSES AND RESPONDENTS FROM WEB SURVEYS. *Public Opinion Quarterly*. 79, 2, 471-503, 2015



**Response Burden and Bias in the First National Food Acquisition and Purchase Survey (FoodAPS-1): An Empirical Analysis Based on Respondent Feedback Survey**

Xingyou Zhang (Economic Research Service), John Kirlin (Economic Research Service), Mark Denbaly (Economic Research Service), Elizabeth Larimore (Economic Research Service)

The first National Household Food Acquisition and Purchase Survey (FoodAPS-1) collected a week-long food acquisition information from participating households. FoodAPS's burden of collecting information from all household members could have increased the household primary respondents' potential reporting errors measured by reporting frequency. FoodAPS-1 also had provided a base incentive of \$100 to each primary respondent and a gift card to each participating household member. The anticipation of receiving the incentive at the end of the collection might have altered or modified the household's normal food acquisition behavior/pattern activities, introducing bias. This study will address two questions: 1) whether reporting frequency is associated with response burden; and 2) whether survey participation altered normal food acquisition behaviors.

A very brief participant feedback survey containing four questions was conducted after the final FoodAPS-1 interview. The response rate for the feedback survey was 97.6% (4,712 out of 4,826 participated households). Our preliminary data analysis shows that 29% of all participants did not report completing their meals and snack form every day during FoodAPS one week long data collection period as expected, 21% of primary respondents had difficulties in engaging household members for participation, and 7% had difficulties in tracking their own food activities. Nearly 11% of responding households reported changes in food acquisition behaviors. Those with difficulties in engaging household members are nearly 60% more likely not to finish their meals and snacks forms every day (odds ratio 1.585, 95%CI 1.356-1.852), and those with difficulties in food tracking more than three times more likely not to finish their meals and snacks forms every day (odds ratio 4.566, 95%CI 3.549-5.875). Multilevel logistic models will be constructed to evaluate the demographic and socioeconomic factors that contribute to non-daily reporting and changes in food acquisition behaviors.

**Changes in Interview Length over a Data Collection Period: Interviewer Learning or Shifting Respondent Characteristics?**

James Dahlhamer (National Center for Health Statistics), Aaron Maitland (National Center for Health Statistics), Stephanie Coffey (U.S. Census Bureau)

Prior studies have shown that interview length tends to get shorter over the course of data collection. Two competing hypotheses have been posited to describe this effect, what Kirchner and Olson (2017) describe as the experience and response propensity hypotheses, respectively. The experience hypothesis states that interviewers learn from their interviews and alter their behavior in subsequent interviews. The response propensity hypothesis states that interviews get shorter because interviewers encounter different, more reluctant respondents as the data collection period progresses. Any changes in measurement are more attributable to respondents than interviewers.

Using 2016 National Health Interview Survey (NHIS) data, we examine the relationship between interviewing experience, response propensity, and the length of sample adult interviews within a monthly data collection period. We estimate a two-level, cross-classified random effects model with interview length cross-classified by interviewer and census tract. Consistent with earlier research, the model includes a within-data collection experience measure, respondent and case characteristics such

as age, sex, race/ethnicity, and education, and a set of interviewer characteristics, including sex, education, interviewing experience, and cooperation rates. In addition, we extend prior research by including richer paradata measures of participant reluctance (e.g., householder statements of privacy concerns and time constraints) as well as neighborhood characteristics (e.g., poverty level, graffiti in the block face) that may influence interviewer behavior. We discuss the implications of our results for interviewer training regimens, with an emphasis on minimizing the effects of interviewers on the measurement process.

### **Smart Math Saves Time and Improves Communication**

Laura Nielsen, Ph.D. (Environmental Protection Agency), Susan Day (Eastern Research Group)

Federal agencies collecting information from regulated communities have the responsibility under the Paperwork Reduction Act (PRA) to inform the public of the industry burden associated with the collection. To this end, burden estimates encompass, "...the total time, effort, of financial resources expended by persons to generate, maintain, retain, disclose, or provide information to or for a federal agency."

Through revision of a number of economic analyses as part of renewals to EPA Information Collection Requests (ICRs), the authors draw generalized lessons for improved estimation methods that also produce useful communication metrics. The techniques include computational simplification via modeling, scaling, and roll-up from activity-level to per-response level unit burdens, temporal pro-rating, and purposeful selection of the response unit itself. As an example of the first technique, EPA's TRI program uses ratio-based burden methodology (RBBM) to simplify calculations, ensure internal consistency, and improve access to per-response unit burdens. The revised structure consists of a system of four ratio models that all hinge on only one base number: Form R nominal burden. The ratio models provide useful metrics, such as for the A/R model at 0.615, conveying that the TRI Form A burden requirement is 61.5% of the TRI Form R requirement. In providing the RBBM simplifications, the TRI program realized improvements in estimate consistency and transparency as well as time savings and improved communications. Additional examples for the Toxic Substances Control Act (TSCA) ICRs show that key techniques can be reapplied to get similar results.

## **Concurrent Session A-3**

### **New Findings on Wealth of U.S. Households**

#### **Economic Experiences of Disadvantaged Families in the Great Recession and its Aftermath**

Lisa Dettling (Federal Reserve Board), Joanne Hsu (Federal Reserve Board)

The timing and magnitude of the Great Recession and the recovery have varied substantially across groups in the United States. A persistent popular narrative describes families and areas that have been "left behind" by the modern economy. This paper analyzes trends in income, wealth, economic expectations, and financial distress across groups, narrowing in on the extent and size of divergent trends in the Great Recession and its aftermath for lower income, rural, and less educated families, families living in poverty, and families living in former industrial, manufacturing, or mining centers, using data from the 2007-2016 Survey of Consumer Finances (SCF). While the existence of stagnant wages and poor employment outcomes for these groups has been well documented, the SCF provides a unique

opportunity to study a complete picture of household financial circumstances, including more complete measures of income (including, importantly, transfer income like disability insurance), income volatility, job histories, savings, and wealth accumulation. For households in fragile economic situations, measures of components of debt, debt payment burdens and behavior, usage of high-cost credit like payday loans, and overall credit market experiences are of particular relevance. Finally, the data also permit the examination of local area characteristics – such as house prices, employment growth, industry structure and the depth of the recession -- in order to comprehensively examine how the macroeconomy affects family finances and economic expectations for the most vulnerable households.

### **A Better Asset and Debt Data Infrastructure for an Ethnically Plural America**

William Darity Jr (Duke University), Darrick Hamilton (The New School)

We will characterize the wealth position of various ethnic/racial groups based on surveys in strategically targeted metropolitan areas. We will produce the asset and debt position of ethnic and racial groups for which existing sources of data reveal little. Ethnic and racial categories will be defined based on specific ancestral origin, so as to address the heterogeneity of asset experiences resulting from grouping specific groups in catchall categories of like Asian American and Latino American. In addition, the broad ethnic and racial categories Native American and black American will be disaggregated based on tribal and ancestral origins as well. Little is known about the asset positioning of these sub-groupings described above, particularly those with Native American and Asian heritage. An intent of these more refined ethnic/racial categories is to emphasize the more specific asset positioning of these subpopulations rather than hiding them away in broad categorizations such as “non-white.”

Surveys in the Baltimore, Boston, Los Angeles, Miami, Tulsa and Washington, DC metropolitan areas were conducted to ascertain conventional and unconventional financial information related asset and debt accumulation. Examples will include respondent’s use of payday lending, check cashing institutions, and international receipts and remittances, as well as the roles of individual or family member incarceration in wealth accumulation. This information is not commonly found in existing surveys.

Finally, we will offer proposals of how best to integrate this sampling scheme into our national data infrastructure, so as to better understand one of the most key indicators of wellbeing, wealth, for groups disaggregated based on ancestral origin, which is crucial as American ethnic plurality increased overtime.

### **Retirement Adequacy and Wealth Distribution Among Early Savers**

Alice Henriques (Federal Reserve Board), Lindsay Jacobs (Federal Reserve Board), Elizabeth Llanes (Federal Reserve Board), Kevin Moore (Federal Reserve Board), Jeffrey Thompson (Federal Reserve Board)

The key financial resource that many households rely on in retirement are pension plans. These plans have drastically transformed over the past three decades and a large literature documents these changes and the possible effects on households’ retirement wealth. While we have good information about what types of pensions households rely upon as they transition into retirement, we know less about how they arrive at that life event, how this path has changed over time as plan offerings and availability have evolved, and the implications for retirement preparedness looking ahead. In our research, we use the Survey of Consumer Finances, including the most recent 2016 data, to study the

evolution of retirement wealth for cohorts that have begun their retirement planning (around age 40) – and follow them until they reach the ‘doorstep’ of retirement. We build a measure of comprehensive retirement wealth to include wealth from DB plans and expected Social Security income relying on individual-level estimates of earnings histories. This exercise suggests younger cohorts initially appear more prepared for retirement (in their early 40s) but accumulate assets less rapidly through their 40s and 50s. We then examine how the level and composition of these resources vary across income, wealth, race, education groups, and expectations about retirement age. For some groups, the addition of Social Security wealth has a small effect on retirement wealth and preparedness, while for many other groups, Social Security may be the primary source of income in old-age – leading to little accumulation of any other retirement wealth.

### **Evaluating Wealth Data in the Redesigned 2014 Survey of Income and Program Participation**

Jonathan S. Eggleston (U.S. Census Bureau), Michael Gideon (Amazon)

The U.S. Census Bureau redesigned the 2014 panel of the Survey of Income and Program Participation (SIPP). As part of the redesign, several changes were made to the wealth module of the survey to fill gaps in question content and improve clarity of existing questions. To evaluate the effects of these changes, we compare SIPP wealth data with Survey of Consumer Finances (SCF) wealth data to investigate how the match between SIPP and SCF changed after the questionnaire redesign. We find that the match between SIPP and SCF has improved for many wealth estimates, although numerous discrepancies remain. We offer potential explanations for why some estimates have changed and why the difference between SIPP and SCF is large for certain estimates.

## **Concurrent Session A-4**

### **Non-Probability Sampling and Estimation**

#### **Redirected Inbound Call Sampling (RICS) – An Example of Fit for Purpose Non-Probability Sample Design**

Burton Levine (RTI International), Karol Krotki (RTI International)

Redirected Inbound Call Sampling (RICS) is a new nonprobability sampling methodology that redirects telephone calls which fail to connect to their intended target to a survey. Pairing RICS with an interactive voice response system to collect data dramatically reduces data collection costs compared to using live interviewers. In addition, the methodology offers researchers very fast turnaround and large sample sizes. We are encouraged by how closely the unweighted demographic distribution of the respondents match known population benchmarks. This in turn leads us to be optimistic about using this new methodology to produce useful population estimates. The two main challenges of this methodology are minimizing measurement error from the interactive voice recognition (IVR) and quantifying bias. In this talk we will (1) describe RICS methodology, (2) present results from experiments to evaluate measurement error, and, (3) present results from experiments designed to evaluate bias. In summary, the evidence suggests that RICS could be a viable survey research tool, producing results that are comparable to those obtained using much more expensive and time-consuming survey methods.

**Benchmark Assessment of Respondent Driven Sampling Data for Foreign-Born Korean Americans**

Sunghee Lee (University of Michigan), Sooin Lee (Mathematica Policy Research), Michael Elliott (University of Michigan), Z. Tuba Suzer Gurtekin (University of Michigan)

Respondent driven sampling (RDS) is a relatively new sampling method specifically proposed for sampling rare or hidden populations. RDS starts with the members of the target population and traces their social networks as well as the networks of those who are connected. Through chain referrals stimulated by incentivized recruitment coupons, RDS exploits the social networks for sampling purposes without screening.

With the pressing needs for studying rare/hidden populations despite the increasing expense and difficulty in sampling for rare or hidden populations, RDS has gained popularity rapidly. However, in contrast to the increasingly large volume of research using RDS data, its methodological assessments are very limited. Statistical inference for data collected under RDS designs are under-developed as their formal base sits on strong assumptions that can be easily violated in practice, such as respondents accurately reporting the number of rare/hidden population persons in their network and randomly choosing among them. Scarcity, if not near absence, of publicly available RDS data makes objective methodological assessments even more challenging, leaving the level of population representation through RDS under-scrutinized.

This study aims to provide the beginning of an already-overdue empirical investigation into the realities of RDS data collection through the total survey error (TSE) framework. Targeting foreign-born Koreans in the U.S. a rare group accounting for 0.3% of the general population, we implemented RDS in Los Angeles County and State of Michigan for a Web survey on general socio-demographic, social and health topics. Respondents were followed up after the main survey regarding their recruitment experiences. This study examines: 1) sampling productivity, 2) the nature of recruitment and recruitment chains, and 3) potential biases by comparing to external data using probability samples, such as the American Community Survey. This study warrants further research about innovative estimation methods.

**Comparing Alternative for Estimation from Nonprobability Samples**

Richard Valliant (Universities of Michigan & Maryland)

Two approaches to estimation from nonprobability samples are quasi-randomization and superpopulation modeling. In the former, the sample is treated as if it was obtained via a probability mechanism but, unlike in probability sampling, that mechanism is unknown. Pseudo selection probabilities of being in the sample are estimated by using the sample in combination with some external data set that covers the desired population. In the superpopulation approach, observed values of analysis variables are treated as if they had been generated by some model. The model is estimated from the sample and, along with external population control data, is used to project the sample to the population. The specific techniques are the same or similar to ones commonly employed for estimation from probability samples and include binary regression, regression trees, and calibration. This paper reviews some of the options that can be used in the two approaches.

**Combining Probability and Nonprobability Samples to form Efficient Hybrid Estimates**

Jill A. Dever (RTI International)

Nonprobability surveys, those without a defined random sampling scheme or sampling frame, are becoming more prevalent. These studies can offer faster results at less cost than many probability surveys, especially for targeting important subpopulations. This can be an attractive option given the continual challenge of doing more with less, as survey costs continue to rise and response rates to plummet. Nonprobability surveys alone, however, may not fit the needs (purpose) of Federal statistical agencies where population inference is critical.

Nonprobability samples may best serve to enhance the representativeness of certain domains within probability samples. For example, if locating and interviewing a required number of subpopulation members is resource prohibitive, data from a targeted nonprobability survey may lower coverage bias exhibited in a probability survey. In this situation, the question is how to best combine information from both sources.

This research attempts to answer this question through an evaluation of hybrid estimation methods currently in use that combine probability and nonprobability data. Methods that employ generalized analysis weights are the focus because they enable other survey researchers and policy makers to analyze the data. The goal is to identify procedures that maximize the strength of each data source to produce hybrid estimates with the low mean square error. Empirical results are shown using simulated and real-world data to illustrate the hybrid estimators.

## Concurrent Session A-5

### Transparent Quality Reporting in the Integration of Multiple Data Sources

Session abstract: Over the past year, a working group of the Federal Committee on Statistical Methodology has carried out an in-depth study of some data-quality issues that arise in the integration of multiple data sources. Building on previous work by a consensus panel of the Committee on National Statistics, three FCSM-WSS workshops, and related methodological literature, the FCSM working group focused principal attention on criteria for transparent reporting on multiple dimensions of data quality.

The presentations in this session will synthesize some of the fundamental concepts and practical recommendations for transparent quality reporting that were identified in the abovementioned work. The first presentation will review some previously developed conceptual frameworks for assessing the quality of statistical data products and services that are based on the integration of multiple data sources; and will also review some practical definitions of "transparent quality reporting" for specified stakeholder groups. The second, third and fourth presentations will discuss some essential findings and recommendations in the areas of, respectively, input data quality, processing quality, and output data quality. In the subsequent floor discussion, audience members will be invited to comment on the issues in quality assessment and reporting that they have found to be especially important as they have considered integration of multiple data sources for specific cases as data producers or consumers.

## Concurrent Session B-1

### Advances in Web Probing

#### **An Automated Refusal Conversion Strategy for Web Surveys**

Taylor Lewis (U.S. Office of Personnel Management), Mark Gorsak (U.S. Office of Personnel Management), Naomi Yount (Westat)

Survey research organizations typically attempt to convert sampled individuals who initially refuse to participate in a survey. This is especially true for interviewer-administered surveys. Knowing the reason(s) for the refusal can help the interviewer craft an appeal. However, the use of such conversion attempts in self-administered, Web-based surveys is not as straightforward to implement.

This paper presents results from an experiment conducted during the 2017 Federal Employee Viewpoint Survey (FEVS) to assess an automated refusal conversion process whereby a portion of sampled individuals was given the opportunity to opt out from the survey. In particular, approximately 10% of the overall FEVS sample—about 112,000 full- and part-time federal employees—was offered the opportunity to stop receiving additional email solicitations. Before effectively adding the individual to the “unsubscribe” list, however, he or she was asked to cite the primary reason for choosing not to take the FEVS (e.g., too busy, confidentiality concerns, survey results not being shared with employees). A randomly assigned subset of the experimental cohort was given an appeal, tailored to the reason provided, in the form of a concise list of assurances and survey facts about which the individual may not have been aware. At that point, the individual could either confirm intentions to opt out or click a link to navigate back to the start of the survey. Because the complementary portion was given no such appeal, we are able to investigate if, and under what circumstances, the strategy is effective in convincing individuals who may not have been initially inclined to participate to do so.

#### **Using Close-Ended Web Probes to Inform the Redesign of the National Health**

Paul Scanlon (National Center for Health Statistics)

Start in 2015 the Collaborating Center for Questionnaire Design and Evaluation Research (CCQDER, formally the QDRL) at the National Center for Health Statistics began to explore the potential for using web probing to supplement its questionnaire evaluation studies. In comparison to cognitive interviewing, web probing has the potential to examine the cognitive processes and question-response behaviors of a larger population, which can be particularly useful when evaluating a questionnaire for a national survey. However, up until that point, almost all work on web probes had explored how they could be administered as open-ended questions—attempting, in other words to replicate the data obtained from qualitative, face-to-face methods such as cognitive interviews. On the other hand, in an effort to reduce respondent burden and to focus the information obtained via the web probes, CCQDER developed a mixed-methods approach which took the information gleaned from cognitive interviews to develop close-ended probes that could be administered to a wider sample using a web panel.

This presentation will first provide an example of the use of close-ended web probes in the questionnaire design and evaluation process, focusing on the proposed chronic lung condition variable for the redesigned National Health Interview Survey (NHIS). This proposed question was evaluated

alongside the existing questions both in cognitive interviewing and through web probing on NCHS' Research and Development Survey (RANDS). The results of this evaluation informed the decision to consolidate three existing questions into one on the redesigned NHIS. Finally, some methodological considerations on the use of close-ended web probes will be presented, including how respondents interacted with these web probes and the impacts their presence had on the response to other survey items.

### **The Roles of Typicality, Specificity, and Set Size in Providing Examples for Survey Questions: Evidence from an Online Study Using Self-Administered Comprehension Probes**

Erica Yu (Bureau of Labor Statistics), Amy Swallow (Bureau of Labor Statistics)

Research has shown that providing a list of examples may help respondents to identify the intended target category or help to cue retrieval of relevant items from memory. However, examples may instead inhibit retrieval (part-set cueing effect). Currently, the Consumer Expenditure Survey (CE) provides examples of eligible items for most questions but these lists have not been tested for their effect on respondent comprehension of the target category or retrieval from memory.

In an online study of 1400 participants, we presented lists of examples that varied in their typicality, specificity, and size and qualitatively evaluated their impact on expenditure reporting and category comprehension. For comparison, we included two control groups, in which participants were presented lists that matched the current production lists or saw no lists at all (these participants saw only the target question). Participants were asked to report their expenditures for the target item category and report any expenses they considered but ultimately excluded from reporting. Self-administered comprehension probes included a task to name the first five eligible expenses that come to mind and a category judgment task.

Researchers will learn how lists of examples affect respondent comprehension and reporting. Conclusions will include recommendations for designing lists to reduce error, as well as a discussion of using close-ended comprehension probes in an online self-administered survey. The findings from this study should be applicable to other household surveys.

### **Benefits and Drawbacks of Using Crowdsourcing Techniques for Cognitive Testing**

Sarah Cook (RTI International), Rachel E. Morgan (Bureau of Justice Statistics), Christopher Krebs (RTI International), Lynn Langton (Bureau of Justice Statistics)

The Bureau of Justice Statistics (BJS) and RTI International (RTI) successfully used crowdsourcing techniques as a method of cognitive testing to recruit respondents and test survey questions as part of the National Crime Victimization Survey Redesign Research (NCVS-RR) program. Crowdsourcing is a technique used to obtain information from a large number of people, which are compensated or uncompensated, typically using the internet. As a cognitive testing method, crowdsourcing provides useful improvements to and clarifications of survey questions and is an important part of survey development. Researchers are able to gain insights on survey questions by targeting specific demographic characteristics, such as race, Hispanic origin, age, and gender. Benefits of crowdsourcing include the speed and low cost with which researchers can obtain information. This presentation includes an overview of crowdsourcing, including a comparison of crowdsourcing and traditional in-person cognitive interviewing, and how crowdsourcing can be an effective method of cognitive testing.



Also included in this presentation is a discussion on the types of topics that may be better suited for crowdsourcing compared to other methods of cognitive testing.

### **Impact of Dependent Interviewing on Consistency of Answers in the American Housing Survey**

Katie Gustafson (U.S. Census Bureau), Evan Brassell (U.S. Census Bureau)

Dependent interviewing is the use of a respondent's answers from a previous wave in a longitudinal survey to modify question phrasing or skip items, with the goal of reducing respondent burden and improving data quality. The American Housing Survey has used dependent interviewing in every survey having returning respondents until a new sample was drawn in 2015. For some dependent interviewing questions, notable differences were found between 2015 estimates and estimates from previous waves. Examination across waves of responses to dependent interviewing questions indicated that dependent interviewing may exert an influence by reducing changes in response and reducing uninformative responses, two known effects of dependent interviewing.

## **Concurrent Session B-2**

### **Topics in Editing and Imputation: Automated Systems, Machine Learning, Hot Deck, Response Propensities, and Edit Reduction**

#### **Innovative Approaches to Generating National and Regional Estimates of Crime and Arrest with Incomplete Data**

KiDeuk Kim (Urban Institute), Ashlin Oglesby-Neal (Urban Institute), Dean Obermark (Urban Institute)

How to deal with missing data is a widespread statistical problem for most real-life datasets. The Department of Justice measures the magnitude, nature, and impact of crime across the country through multiple data collection programs. However, not all law enforcement agencies report to the Federal Bureau of Investigation (FBI), and as a result, crime statistics are often incomplete, making it difficult to interpret crime data. Using the National Incident-Based Reporting System data compiled by the FBI, this presentation will introduce a machine-learning approach to missing data imputation and discuss its performance, compared to some of the traditional approaches (e.g., mean substitution, hot deck imputation, regression-based imputation). This presentation will also demonstrate the development of national and regional estimates of crime with incomplete administrative data and review the design and results of validation tests for the imputed data on crime and arrest around the country.

#### **Evaluating Hot Deck with Propensity Matching for the Advance Monthly Retail Trade Survey (MARTS)**

Katherine Jenny Thompson (U.S. Census Bureau), Laura Bechtel (U.S. Census Bureau), Nicole Czaplicki (U.S. Census Bureau)

Propensity score matching is frequently used in observational or experimental studies. The objective is to "balance" participants in the treatment and control panels on selected characteristics to eliminate or reduce confounding. This is accomplished by finding similar pairs of units on these characteristics using a predetermined distance measure, then splitting the pairs into separate treatment and control groups

so that the study approximates a random experiment. In general, the matching is performed using a single propensity score function constructed from the full considered set of characteristics for matching, where the modeled propensity score represents the unit's expected response to the treatment. Propensity score matching can also be used in hot deck procedures to match donors to recipients for imputation, with the "treatment" being the missing data replacement procedure. If there is evidence that covariates predictive of nonresponse are likewise predictive of outcome, then an estimated response propensity score could be used for propensity matching. For example, the size of a business may be predictive of both the response propensity and the reported value of sales. Note that a single propensity score is certainly convenient but is not necessary for implementation: Bergstralh and Kosanke at the Mayo Clinic have developed SAS code that directly implements greedy and optimal matching defined by weighted or unweighted combinations of continuous covariates, where the larger weight indicates more importance in the matching criteria.

The Advance Monthly Retail Trade Survey (MARTS) is investigating alternative imputation methods, including hot deck with propensity matching. The MARTS is a probability sample of companies, subsampled from the Monthly Retail Trade Survey (MRTS), whose respondents provide "early" estimates of monthly sales in retail trade industries; the more reliable monthly estimate is provided a few weeks later in the MRTS publication. Here, we focus on determining an effective propensity scoring matching application for the MARTS nonrespondents, given a predetermined set of imputation cells. We evaluate the proposed propensity models and matching algorithms (greedy or optimal) on empirical data from several months of the MARTS and MRTS collections.

### **Assessing the Automated Imputation of Missing and Erroneous Survey Data: A Simulation-Based Approach**

Larkin Terrie (Bureau of Economic Analysis)

To improve the efficiency and cost effectiveness of survey editing, the Bureau of Economic Analysis (BEA) is developing an automated data editing and imputation system for use with its surveys of multinational enterprises. Previous work has tested this auto-editing system by using it to edit survey forms that have already been manually edited and comparing its results to those produced by manual editing. This paper presents a more rigorous means of assessing the accuracy of imputations for missing and erroneously reported survey items that does not rely on comparisons with the results of manual editing. In doing so, it contributes to the general literature on measuring the quality and accuracy of auto-editing systems used in generating official government statistics. The proposed approach involves simulating the presence of missing and erroneous data in survey forms with neither missing responses nor responses identified as erroneous by BEA's validity checks. From these forms, data items are randomly selected for imputation in a way that mimics the distribution of erroneous and missing items among all received forms. Imputations for these items are then produced and compared to the original, and presumably correct, reported values. By repeating this process a large number of times, it is possible to draw conclusions about the average accuracy of the auto-editing system's imputation procedures and to use these conclusions to improve the procedures. In particular, this paper presents the results of applying this technique to data from one of BEA's surveys of foreign direct investment in the United States, the BE-15C. This technique helps to improve the quality of imputations for BE-15C data by providing an objective means of comparing the imputations produced by alternative imputation procedures.

**Edit Reduction Research in the U.S. Census Bureau's Economic Directorate**

Lisa Kaili Diamond (U.S. Census Bureau), Brian Dumbacher (U.S. Census Bureau)

The usefulness of economic estimates often depends on two characteristics: accuracy and timeliness. Editing methodologies have a direct impact on both of these attributes. For decades, statisticians have researched the implications of over-editing, and it is widely agreed that the over-editing of survey data can deplete resources, delay information releases, and introduce bias. In the U.S. Census Bureau's Economic Directorate, a team called EconEdit was formed to investigate current editing practices, increase the efficiency of the editing process, reduce costs, and improve the timeliness of data releases without diminishing data quality. In this presentation, we highlight findings from the EconEdit's study of the Annual Capital Expenditures Survey and the Quarterly Services Survey. Focus is given to developing adaptive editing strategies such as identifying editing stopping points.

## Concurrent Session B-3

### Measuring and Imputing Incomes

**Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions**

Jeff Larrimore (Federal Reserve Board), Jacob Mortenson (Joint Committee on Taxation), David Splinter (Joint Committee on Taxation)

Tax return data are increasingly the standard for tracking income statistics in the United States. However, these data have traditionally been limited by their inability to capture non-filers and to identify members of separate tax units living in the same household. We overcome these obstacles and create household records directly in the tax data using mailing address information included on tax forms. We then present the first set of tax-based household income and inequality measures for the entire income distribution. When comparing household income inequality results in the tax data to those using the March CPS, we confirm previous findings that the March CPS understates the inequality of household income. However, we also find that the previous approach of using tax units in the IRS data to proxy for households leads to an overstatement of household income inequality. Finally, using households in the IRS tax records, we illustrate how focusing on tax units rather than households alters the observed distribution of tax programs such as the Earned Income Tax Credit.

**Calculating a Supplemental Poverty Measure in the Survey of Income and Program Participation**

Ashley Edwards (U.S. Census Bureau), Liana Fox (U.S. Census Bureau), Lewis Warren (U.S. Census Bureau)

Since the first estimates of the Supplemental Poverty Measure (SPM) were produced by the Census Bureau in 2010, estimates have been derived from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC). However, original National Academy of Science recommendations on the implementation of a SPM called for calculating the measure from an alternate data source, the Survey of Income and Program Participation (SIPP). Those recommendations were based on the fact that the SIPP has traditionally done a better job than the CPS ASEC in collecting income data, particularly among low-income families. However, because much of the detailed content necessary for calculating

the SPM is collected through the SIPP in topical modules, limitations related to the timing and consistency of those modules throughout the panel have created formidable obstacles to calculating consistent annual SPM estimates. This research takes advantage of the recent redesign of the 2014 SIPP Panel, which provides more consistent collection of the detailed data on family relationships, income and transfer program receipt, and medical, work-related, and tax expenses necessary to calculate the SPM. We calculate SPM estimates for 2013 as derived from the SIPP, and evaluate differences in the SPM rate— as well as differences in the incremental impact of income sources and expenses— across the SIPP and CPS ASEC for 2013.

### **Improving Earnings Imputations in the Survey of Income and Program Participation**

Gary Benedetto (U.S. Census Bureau), Rebecca Chenevert (U.S. Census Bureau), Jonathan Eggleston (U.S. Census Bureau), Mark Klee (U.S. Census Bureau), Joanna Motro (U.S. Census Bureau), Robert Munk (U.S. Census Bureau)

Researchers and policy analysts are expressing increased concern about the quality of earnings data across various nationally representative household surveys. A primary reason for this increased concern is the widespread rise in both item-level and unit-level nonresponse. Nonrespondents' missing earnings are typically imputed via hot deck. However, research has demonstrated that hot deck imputed earnings are less reliable than reported earnings. In an effort to improve the quality of earnings imputations, we demonstrate how model-based imputation can be incorporated into production data from the 2014 panel of the Survey of Income and Program Participation (SIPP). Model-based imputation techniques can mitigate match bias by controlling for a broader array of earnings determinants than standard hot deck imputation techniques allow. Especially noteworthy is our ability to control for administrative earnings records when predicting earnings through model-based imputation. In this paper, we assess the quality of our model-based imputations relative to standard hot deck imputations by comparison to administrative records. We also evaluate the effect of both types of imputations on some key coefficients of earnings regressions.

### **Are Bracket Responses Accurate? Using IRS W-2 Records to Design Earnings Edits for the Redesigned CPS ASEC**

C. Adam Bee (U.S. Census Bureau)

The newly redesigned income section of the Current Population Survey Annual Social and Economic Supplement (CPS ASEC, also known as the March CPS) contains new questions presented to respondents who decline who do not respond to the earnings amount item. These new questions present the respondent with a series of ranges of amounts to choose among. These bracket responses are intended to obtain useable information without much additional respondent burden. This paper assesses the degree to which these bracket responses are accurate, with an eye toward informing the design of new earnings edits that will make use of these bracket responses. I compare joint distributions of CPS ASEC earnings amount responses and earnings bracket responses and individually linked IRS W-2 records. I find that bracket responses may represent a useful follow-up method to obtain useful information from reluctant survey respondents, and I present detailed recommendations for the implementation of new CPS ASEC earnings edits.

## Concurrent Session B-4

### Innovations in Sample Design

#### **Sampling Design and Variance Estimation for the National Survey of WIC Participants, Wave 3**

Stanislav Kolenikov (Abt Associates), David Judkins (Abt Associates)

The third wave of the National Survey of WIC Participants (NSWP-III) plans to collect data on five categories of WIC participants (pregnant women, breastfeeding women, postpartum women, infants, and children up to their 5th birthday) regarding their use of the program, and their experiences with it. One of the most challenging components of NSWP-III data collection and reporting is validation of eligibility for the program and estimation of the certification error rates. The face-to-face interviews collecting detailed income data are conducted by the field staff with participants who have been certified in the period of up to 6 weeks prior to the interview. To obtain the sample of the recently certified WIC participants, we plan to employ a multistage sampling design, in which the PSUs are geographically contained groups of local WIC agencies, stratified by region. This presentation describes the two statistical innovations used in drawing the sample and estimating the sampling variances. First, the sample is drawn using a high entropy conditional Poisson sampling method. Second, variances are estimated using the balanced bootstrap (or generalized BRR) replicate variance estimation method. We demonstrate the challenges we faced in implementing these methods, and the solutions we developed.

#### **Implications of Repeated Sampling in a Crime Survey**

Mike Brick (Westat), Grace Kena (Bureau of Justice Statistics), Pam Broene (Westat)

One of the reasons for conducting a longitudinal or panel survey rather than a repeated cross-sectional survey is to reduce the sampling error for estimates of change over time. The National Crime Victimization Survey (NCVS) uses a rotating panel design that retains sampled addresses for 7 interviews. Repeatedly sampling the same addresses has both operational and statistical effects. In particular, repeated sampling can affect response rates and reported outcomes (conditioning effects). This research examines the advantages and disadvantages of overlapping the sample for estimating crime victimization and other related characteristics.

The Bureau of Justice Statistics is exploring the feasibility of a lower-cost mail survey that local jurisdictions could field to produce reliable information on victimizations and community attitudes about safety and local police agencies. The standardized format of the instruments permits comparisons to other areas and over time. The prototype of this National Crime Victimization Survey – Companion Survey (NCVS-CS) was tested in 2015 and again in 2016 in the 40 largest metropolitan areas in the U.S. In each year, a mail survey was sent to a probability sample of more than 200,000 addresses. One-quarter of the addresses sampled in 2015, selected randomly, were retained in the sample for 2016. This experiment allowed investigation of the advantages and disadvantages of including the same addresses in the sample in both years.

We examine the effects of retaining the same addresses on response rates, estimates of level and change for touched-by-crime rates and community attitudes, and the year-to-year correlations for these estimates. We give recommendations for designing the NCVS-CS and similar surveys, for computing both level and change estimates.

**Assessing How Well Students in Economics Experiments Model Decision-Making for Non-Student Populations**

Stephanie Rosch (Economic Research Service), Jacob Fooks (Economic Research Service), Daniel Hellerstein (Economic Research Service), Lori Lynch (Economic Research Service), Kent Messer (University of Delaware), Sharon Raszap Skorbiansky (Economic Research Service), Collin Weigel (Johns Hopkins University), Eliana Zeballos (Economic Research Service)

According to the rules set forth by Vernon Smith (1982), economics experiments need to provide payments to motivate subjects to care about their performance within the experiment. These payments must be salient to subjects and dominate any other considerations which could affect experimental performance (e.g. a desire to please the experimenter, innate enjoyment of the experimental activities, overcome boredom). Improperly calibrated incentives increase the noisiness of subject responses, reducing the power of the experiment to detect a treatment effect. However, exactly what constitutes an appropriate menu of incentive payments for a given experiment can be hard to judge. We review the state of knowledge about how the amount and structure of payments impact participant performance for a broad range of experimental methods including market experiments, single- and multi-player games, auctions, risk elicitations, and choice experiments. For each method, we document (1) norms for the levels of payments used and (2) the effect of different payment scales on participant's decision-making and experimental outcomes. We also discuss common response patterns to payment scales across methods, contrast the findings from the experimental economics and experimental psychology literatures, and highlight critical gaps in the literature.

**Using Social Media for a Probability Sample: Is It Possible?**

Marcus Berzofsky (RTI International), Tasseli McKay (RTI International), Patrick Hsieh (RTI International), Rob Chew (RTI International), Katie Grimes (RTI International), Amanda Smith (RTI International), Natasha Latzman (RTI International), Marni Kan (RTI International), Mindy Stahl (RTI International)

Hard to reach populations, such as LGBTQ+ individuals, often have a small number of respondents in nationally representative surveys of youth. For topics in which intervention methods may differ across sub-population, such as suicide prevention, the limited number of sub-population respondents prevents researchers from being able to compare general and sub-population prevalence or identify sub-population specific risk and protective factors, both of which are critical to developing sub-population specific recommendations.

Alternative frames, such as social media platforms, often require non-probability based methods to obtain sample members in the subdomain of interest. Our approach attempts to develop a frame of youth from social media platforms which is representative of youth on the social media source and stratify the frame using public information about the frame members. The frame is developed by randomly selecting a large set of users and identifying those who meet our age criteria. To develop strata, we use information from focus groups to help inform predictive models. We will discuss our oversample of the subdomain of interest (i.e., LGBTQ+ youth), the creation of design-based weights based on known probabilities of selection, and the recruitment a sample of users from social media. We will describe the challenges associated with and our solutions for developing design-based probabilities of selection, recruiting the sampled users, and adjusting for potential coverage and nonresponse bias.

We present an evaluation of our ability to stratify the population through our predictive models and methods for improving our models using the pilot data. Coverage bias is measured by assessing the coverage penetration of the social media platform among youth and the characteristics of the youth. Moreover, we discuss sources for population control totals and methods using federally-sponsored national surveys of youth.

### **Comparing Random and Nonrandom Samples Using Model-Implied Randomization**

Vladislav Beresovsky (National Center for Health Statistics)

The current reality in survey statistic is characterized by multiple problems with traditional randomized surveys and proliferation of nonrandom samples from web surveys and administrative datasets. However, applicability of nonrandom samples for producing national estimates is an open question. The National Center for Health Statistics (NCHS) has sponsored a web survey collecting a subset of variables of the regular National Health Interview Survey (NHIS) with the purpose to validate estimates from the web sample by comparing them with estimates from the random NHIS sample. Making a definitive conclusion from such a comparison is difficult due to variability of estimates and the fact that estimates from the nonrandom sample utilize covariate distribution from the random sample. Applying techniques developed in biostatistics for testing for treatment effect in observational studies, we use models for conditional randomization of the nonrandom sample to test hypothesis about identity of estimates from both samples. Proper accounting for survey weights introduces additional complications. Validity of the proposed testing methods is illustrated in simulations and applied to real random and nonrandom NHIS samples.

## Concurrent Session C-1

### Contact Strategies

#### **Effect of Contact Strategy on Response Mode Selection**

Yuliya Romanyuk (Statistics Canada), Cilanne Boulet (Statistics Canada)

The issue of declining response rates is a major challenge for Statistics Canada and many other statistical agencies. One way to address this issue is offering respondents several ways of completing questionnaires (for instance, online or using paper) with the hope of increasing response rates. Since electronic questionnaires (EQ) typically result in better-quality data and are cheaper and faster to process than paper questionnaires (PQ), Statistics Canada has a vested interest in promoting self-response online while keeping other response channels available to achieve higher response rates. In order to evaluate appropriately the costs and the benefits of a multimode collection strategy, it is important to determine which factors affect the choice of a response mode and to understand how much statistical agencies can expect to sway respondents to choose EQ over PQ by encouraging them to fill out questionnaires online.

The contact strategy is a key tool with which a statistical agency can hope to direct respondents to choose a particular response mode. The extent to which the method of contact affects the choice of a response mode can be evaluated using the 2014 Canadian Census Test. A sample of dwellings was taken and randomly assigned to different panels; some panels were sent an invitation letter to complete a questionnaire online and others were sent a paper questionnaire directly. The response mode choice from the Census Test can then be compared with the choice made during the 2011 Census, while controlling for the method of contact. Census Test response rates in different groups, based on the 2011 response mode choice, show that the contact strategy is a very strong factor in determining the choice of a response mode and allow us to quantify its effect.

#### **Topic Interest in Voting Surveys: Experimental Evidence**

Colin MacFarlane (Fors Marsh Group), Carl Turner (Fors Marsh Group), Jonathan Mendelson (Fors Marsh Group), Krysha Gregorowicz (Fors Marsh Group)

The academic literature on survey-based estimates of election participation identifies two mechanisms that explain differences in estimates across surveys and differences in survey estimates and administrative voting data. First, respondents are more likely to respond to surveys on topics of interest to them, and specifically those who have a high propensity to vote are more likely to respond to voting surveys. Second, there might be a social desirability bias, whereby the wording of the survey invitations results in the respondents feeling pressure to report having voted.

To test issues of nonresponse and item response bias, the 2016 Post-Election Voting Survey of the Active Duty Military (PEVS-ADM) was partitioned into “control” and “experimental” samples. The control sample received communications branded normally as the 2016 PEVS-ADM, which used the same language in previous administrations, highlighting how the survey will ask voting-related questions. The experimental treatment was administered to a random subset of 15 percent of sampled individuals. They were sent invitation and reminder letters that characterized the survey as a “QuickCompass”



survey about general active duty experiences, rather than a post-election voting survey. This treatment tests if removing references to voting from the survey invitations can both increase response rates and decrease response bias, presumably by attracting non-voters who may not perceive themselves to be part of the population of interest under the old survey and thus were less likely to respond.

### **Adaptive Designs - Tailoring of Contact Materials to Increase Survey Participation**

Rebecca J. Powell (RTI International), Antje Kirchner (RTI International), Emilia Peytcheva (RTI International)

Contact materials sent to respondents to inform them of an upcoming survey request have the potential to influence sample members' participation decision. Sending advance letters has long been the practice, informed by experiments demonstrating significant increase in response rates across modes.

The leverage saliency theory (Groves et al., 2000) argues that different respondents are persuaded to take part in a survey by different design attributes. We manipulate two such design features across sample members to increase survey participation – we tailor the message in the contact materials to increase the perceived relevance of the survey request, and manipulate the emphasis of the survey sponsor (long recognized as a major factor driving participatory decisions) in all electronic communications.

The 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B: 08/18) is the third and final follow-up of sample members from the 2007-08 National Postsecondary Student Aid Study, providing a unique set of frame variables available for both respondents and nonrespondents. Using information from prior data collections, we randomly assign respondents to either a generic version of communication materials, emphasizing their experience in gaining a bachelor's degree, or a tailored version, emphasizing their experience in gaining a bachelor's degree in their specific field. These sample members were further randomly assigned to e-mail messages that were either sent from the data collection contractor or the National Center for Education Statistics (NCES).

Preliminary results suggest tailoring contact materials to include sample member's major slightly increases response rates for sample members who had previously been a nonrespondents, as well as data collection efficiency—measured in days till survey completion—for all respondents. We also see a slight increase in efficiency when we send emails with the NCES sponsorship compared to those with the data collection contractor sponsorship. Both experiments improve sample representativeness.

### **If We Boost It, Will They Come?: Evaluating Efforts to Promote Web Completion in a Multi-Mode National Establishment Survey**

Lauren Harris-Kojetin (National Center for Health Statistics), Christine Caffrey (National Center for Health Statistics), Melissa Hobbs (RTI International), Angela Greene (RTI International), Manisha Sengupta (National Center for Health Statistics)

The National Study of Long-Term Care Providers, conducted by the National Center for Health Statistics (NCHS), includes biennial establishment surveys in two sectors (adult day centers and residential care communities). The protocol offers web and mail simultaneously then computer-assisted telephone interviewing (CATI) for those not responding by mail or web. NCHS often does not have email addresses for cases, so uses mail contact to encourage web completion. NCHS seeks to maximize completion by

web over mail or CATI. However, in both sectors completions by web as a percentage of all completions decreased by 25% between the first (2012) and second (2014) survey waves.

For the third (2016) wave, NCHS made protocol changes intended to increase web completions. These protocol changes included: (1) the questionnaire packet cover letter mentioned the web completion option before the mail completion option; (2) a respondent-tailored insert was added to the packet with the web survey login credentials; (3) respondents calling the survey help desk to ask for another questionnaire were encouraged to complete by web; (4) adult day cases that provided email addresses in 2014 (46% of all adult day cases fielded in 2016) were sent two email reminders with web survey login credentials during the mailing phase; and, (5) during the CATI phase, 19% of adult day cases and 12% of residential care cases requested (and were emailed) the web login information rather than complete by CATI.

Completions by web as a percentage of all completions changed from 24% (2012) to 18% (2014) to 38% (2016) for residential care communities, and 33% to 25% to 49% for adult day centers, representing a doubling of the use of web as a completion mode between 2014 and 2016 in both sectors. Additional findings by subgroup and selected protocol changes will be shared.

#### **Reducing Collection Effort While Maintaining Data Quality in Business Surveys**

Jessica Andrews (Statistics Canada), Pierre Daoust (Statistics Canada), Matei Mireuta (Statistics Canada)

In 2016 Statistics Canada implemented a dynamic collection methodology for prioritizing collection for a number of business surveys. The Quality Indicator/Measure of Impact (QIMI) methodology reprioritizes units during active collection in order to concentrate efforts on domains where the quality of estimates is insufficient, where quality can be measured in terms of CV or a variety of other indicators. This prioritization leads to collection effort being directed towards the respondents whose information will have the most impact on the survey and reduces the number of contacts for respondents with less impact. In the future the addition to QIMI of information from the paradata on the progress of respondents through the surveys' electronic questionnaires will improve the targeting of collection effort.

## **Concurrent Session C-2**

### **Applications of Artificial Intelligence and Text Analysis in Federal Statistics**

#### **A Mixed-Methods Impact Analysis of Financial Regulations on the Banking Industry**

Ricky Rambharat (Office of the Comptroller of the Currency), Douglas Robertson (Office of the Comptroller of the Currency), Roger Tufts (Office of the Comptroller of the Currency), Desiree Schaan (Office of the Comptroller of the Currency), Eugene Floyd (Office of the Comptroller of the Currency), Bianca Werner (Office of the Comptroller of the Currency), Richard Moylan (Office of the Comptroller of the Currency)

As a response to the recent Great Recession, banking supervisors in the United States (and worldwide) have implemented varied new capital, liquidity, and operational regulations with the objective of ensuring that banks can robustly react to stressed financial environments. For example, in the U.S., the Dodd-Frank Act Stress Test (DFAST) and Comprehensive Capital Analysis and Review (CCAR) are new regulatory exercises conducted by the Federal Reserve Board to ascertain the quality of the capital planning processes of large financial institutions, and to determine whether these institutions have sufficient capital to absorb losses during “stressed” economic conditions. In this study, we adopt a Mixed-Methods approach to analyze the impacts of these regulations on banking institutions. Our focus is on larger institutions, as many of the new regulatory directives apply only to them. Our study combines analyses of both qualitative and quantitative data to extract key themes that are of primary concern to bank management, and thus bank supervisory agencies. As part of the qualitative facet of our study, we designed and executed a survey instrument as a pilot on the largest banks in the U.S. Both common and idiosyncratic themes emerged from our qualitative study. A subsequent phase of our work involves a textual and sentiment analysis of annual and investor reports from the largest banking institutions in order to quantifiably gauge the key impacts of the major post-recession regulations. Applying Blei et. al (2003) [in J. of Machine Learning Research], we explore a topic modeling approach based on a latent Dirichlet allocation (LDA) to identify key sentiments that inform policy stakeholders tasked with ensuring safety and soundness of financial institutions, particularly as regards their reaction to pivotal regulations.

**Coding Verbatim Responses Using an Auto-coding Program Based on a Two-step Matching Process: National Hospital Ambulatory Medical Care Survey Emergency Department Data, 2015**

Akintunde Akinseye (National Center for Health Statistics), Brian W. Ward (National Center for Health Statistics)

The National Center for Health Statistics annually conducts the National Hospital Ambulatory Medical Care Survey (NHAMCS), which monitors ambulatory health care in U.S. hospitals. NHAMCS contains variables with verbatim responses that need to be coded according to certain specifications, prior to creating analytic datasets. Traditionally, this coding process required certified medical coders (CMC) to manually perform this task. We developed an auto-coding program to automate the code assignment for NHAMCS emergency department (ED) patient’s reason for visit (RFV) verbatim entries.

The program we developed uses a two-step code assignment process: the first stage utilizes a SAS macro program to perform a word-matching algorithm that assigns best-fit codes to RFV verbatim entries. Matching was set at a  $\geq 75\%$  word-matching threshold, using the RFV dictionary as a reference. The second stage assigned codes based on an exact matching to verbatim text of previously coded data by CMC from 2 years prior.

Verbatim data from the 2015 NHAMCS ED visit file were used for this research. Of the 43,565 verbatim entries distributed across the 5 RFV variables, a total of 26,217 (60% coverage) were coded using this program. Accuracy was tested by comparing CMC coded data from 2015 NHAMCS ED with the program-assigned codes, which resulted in 24,566 matching exactly (94% accuracy). To improve coverage and accuracy, the next step will be to examine non-coded verbatim entries for abbreviations, misspellings, and “junk” words.

The underlying logic used in our program can be applied to other settings where verbatim data are coded.

### **Comparison of Machine Learning Algorithms to Build a Predictive Model for Classification of Survey Write-in Responses**

Andrea Roberson (U.S. Census Bureau), Justin Nguyen (U.S. Census Bureau)

The Annual Capital Expenditures Survey (ACES) provides detailed and timely information on capital investment in structures and equipment by nonfarm businesses during the year. The data are used to improve the quality of current economic indicators of business investments, as well as estimates of gross domestic product. Studies conducted by the U.S. Census Bureau have assessed Economic Directorate survey processing procedures and targeted areas for improvement. The resulting initiatives for the ACES questionnaire sought to reduce the workload of analysts who review and edit write-in responses. The form allows respondents to write-in a capital expenditure category as an “Other” category, when no other classification can be determined. This paper aims to review the use of machine learning to develop and validate predictive models to separate the write-ins in the “Other” category into multiple classes. We will examine how the Census SABLE (Scraping Assisted By LEarning) tool is applied to classify the “Other” category into multi-label descriptions: Structures, Equipment and Not Applicable. Finally, we examine the increases in efficiency and reductions in time cost and burden on the editors made possible by deploying SABLE into the data editing process.

### **Using Machine Learning Techniques to Interpret Open-ended Responses in Web Surveys**

Laura Wronski (SurveyMonkey)

Web surveys make it possible to collect open-ended responses in a format that isn’t possible in traditional (telephone or paper and pencil) surveys. Past survey research has shown that survey respondents are more open and honest, and less susceptible to social desirability bias, when responding to surveys conducted on the web compared to surveys conducted in-person or over the phone. Rather than relying on interviewers to transcribe and code, often by hand, every single response, web surveys can make use of new machine learning techniques to analyze data that respondents provide in their own words.

As the use of web surveys increases, it is becoming more important to understand the differences in data quality between responses to closed-ended (typically multiple choice) questions from which so many survey statistics are derived and open-ended questions (text boxes and comment boxes) whose potential is just beginning to be tapped.

In this study, we examine data from two different open-ended questions asked on different surveys sent using SurveyMonkey, an online survey platform. We examine any differences in item nonresponse for these open-ended questions compared to closed-ended questions, to see whether certain demographic or other characteristics influence respondents’ likelihood of responding to open-ended questions. Next, we look at other data quality measures—such as length of response or literacy level of response—to determine which are most suitable for use with open-ended text analysis. Finally, we compare the results from different machine learning techniques to learn about the types of open-ended responses we receive, and whether their sentiment varies by different demographic or other characteristics of interest.

Our goal in this research is a first step in providing recommendations for the optimal use of open-ended questions in all types of surveys.

## Concurrent Session C-3

### Supplemental Poverty Measures

#### **Moving to the Median: The Case for Changing the Range of Expenditures Underlying SPM Thresholds**

Liana Fox (U.S. Census Bureau), Thesia I. Garner (Bureau of Labor Statistics)

The Supplemental Poverty Measure (SPM) was developed in 2010 as an improvement to the official poverty measure. These improvements included, among others, adjustments to resources and the creation of new poverty thresholds. Three thresholds are produced: ones for renters, owners with mortgages, and owners without mortgages. Next these thresholds are further adjusted to account for unit composition and geographic price differences. Each year's thresholds are based on the 30th-36th percentile-range of a five-year rolling average of food, clothing, shelter and utilities (FCSU) expenditures for consumer units with two children. The SPM thresholds were developed in accordance with recommendations from the Interagency Technical Working Group on Developing a Supplemental Poverty Measure.

Setting the threshold at the lower end of the FCSU expenditure distribution has long been a topic of debate. There is a question as to the efficacy of using the 30th-36th percentiles of the expenditure distribution as opposed to examining some percentage of the median, such as 78% to 83% of the median as was used for the National Academy of Sciences thresholds. Moving to the median has some methodological advantages. First, fewer consumer units at the median receive in-kind benefits, so moving to the median would reduce the need to impute some of these noncash benefits, as well as the impact of these imputations on the thresholds. Second, if one were to add medical care to the bundle of goods and services, the median would more adequately account for health care spending needs than the range of expenditures around the 33rd percentile.

This paper will detail the conceptual differences between setting the threshold around the 33rd percentile versus a percentage of the median. This paper will also implement a series of alternative thresholds and explore the impact of these on SPM poverty rates from 2011-2016.

#### **Controlling for Prices before Estimating SPM Thresholds and the Impact on SPM Poverty Statistics**

Thesia I. Garner (Bureau of Labor Statistics), Juan Munoz (Bureau of Labor Statistics)

An advantage of the Supplemental Poverty Measure (SPM), over the official measure, is that the thresholds are produced for individual geographic areas; the official poverty measure does not allow for such an adjustment. SPM reference unit (two adults with two children) thresholds for owners with mortgages, owners without mortgages, and renters are produced at the Bureau of Labor Statistics (BLS) and are based on out-of-pocket spending for food, clothing, shelter, utilities, and a multiplier for other basic goods and services. Thresholds are estimated from data collected in the U.S. Consumer Expenditure Interview Survey (CE), a product of BLS. Reference unit SPM thresholds are sent to the

Census Bureau for further adjustment and to produce poverty statistics. Adjustments to the reference unit thresholds include accounting for differences in compositions (adults and children) and for differences in prices across areas. The purpose of the second adjustment is to convert the “national average” SPM thresholds into thresholds that reflect prices faced by consumers in individual geographic areas. However, as recently noted by Bishop et al. (2017), implicit in the originally BLS-produced SPM thresholds are differences in prices across areas that are not considered. The focus of the current research is to produce geographic price adjustments using the CE data and then to apply these to the consumer unit level shelter and utility expenditures before threshold construction, and to examine the impact on SPM poverty statistics. Quality-adjusted normalized prices for housing are produced separately for owners and for renters. After adjusting for spatial differences in housing and prices across geographic areas, “new national average” reference unit thresholds are produced. These new thresholds are then sent to the Census Bureau for a second geographic adjustment; this step is followed to convert the new national average thresholds into local area SPM thresholds. This second geographic adjustment is applied to the housing share of the reference unit SPM thresholds only, unlike the CE-based adjustment which is applied at the CU level before threshold construction. Introducing this new adjustment to housing expenditures before producing the SPM thresholds results in overall SPM poverty increasing, but does not impact poverty rates for individuals living in the Northwest or West, nor does it change the rate of poverty for owners with mortgages. Other groups considered experience increases in SPM poverty.

### **Enhancing the Supplemental Poverty Measure to Estimate the Impact of Health Insurance Benefits on Poverty**

Dahlia K. Remler (Baruch College, CUNY), Sanders Korenman (Baruch College, CUNY), Rosemary Hyson (Baruch College, CUNY)

What is the impact of expanded health insurance coverage under the Affordable Care Act, or of any health insurance benefit, on poverty? Answering this question requires a valid measure of poverty that both incorporates a “need” for health care or insurance in the poverty threshold and values health benefits as resources available to meet that need. Researchers long struggled to develop such a poverty measure, but could find no valid approach (National Academy of Sciences 1995). The Supplemental Poverty Measure (SPM), based on 1995 NAS recommendations, improves on the official federal poverty measure by including many in-kind benefits as resources, but excludes health insurance benefits. Instead, it subtracts medical out-of-pocket expenditures from resources. Consequently, research using the SPM can assess only indirect impacts of insurance on poverty—impacts on out-of-pocket payments—and not direct impacts of health insurance on meeting health care needs (Sommers & Ollereich 2013).

Recent innovations in poverty measurement, building on the SPM, allow construction of a Health-Inclusive Poverty Measure or “HIPM” (Korenman and Remler 2016). A valid HIPM was made possible by two provisions of the ACA: guaranteed issue and community rating. With these, researchers can determine dollar amounts to add to the poverty threshold to reflect a basic health insurance “need” from information on the number and ages of household members, irrespective of their health status.

We use CPS data to construct a HIPM for the US 2014 and estimate impacts on poverty of health insurance benefits, other means-tested programs, refundable tax credits, and social insurance programs. HIPM estimates show that health insurance benefits are among the most important anti-poverty

programs: Medicaid has a large impact on child poverty; ACA premium subsidies greatly reduce poverty among the individually-insured; and health insurance benefits provide the bulk of poverty relief among low-income non-elderly adults without children.

**Supplemental Poverty Measure: A Comparison of Geographic Adjustments with Regional Price Parities vs. Median Rents from the American Community Survey: An Update**

Trudi Renwick (U.S. Census Bureau), Bettina Aten (Bureau of Economic Analysis), Eric Figueroa (Bureau of Economic Analysis)

One of the innovations of the Supplemental Poverty Measure is to make adjustments in the official poverty threshold to account for geographic price level differences, particularly for differences in the cost of shelter as measured by rents. A more recent initiative is to estimate thresholds that include price differences for goods and services other than rents.

In a 2014 paper (Renwick et al., 2014), two types of geographic adjustments were compared: one based on the ACS median rent index (MRI), and one based on the published set of state and metropolitan regional price parities (RPPs). This paper looks at RPPs that are more narrowly focused on just food, clothing and rents and updates the poverty estimates from the 2014 paper.

This paper highlights the results of these two different adjustments to the national poverty thresholds. The first is based solely on differences in tenant rents while the second is a composite index that includes differences in the price levels of food, apparel and rents. Adjustment using the first index increases the national poverty rate, while the second decreases the rate. The composite index adjustments are larger at the national level, across four broad regions, and across most states.

## **Concurrent Session C-4**

### **Small Area Estimation: Applications and Practical Demonstrations**

**Designing the 2017 Programme for the International Assessment of Adult Competencies to Produce Direct and Indirect Estimates**

Leyla Mohadjer (Westat), Tom Krenzke (Westat), Jianzhu Li (Westat), Wendy Van de Kerckhove (Westat), Lin Li (Westat)

The sample for the 2012 Programme for the International Assessment of Adult Competencies (PIAAC) was designed to produce internationally comparable and nationally representative direct estimates (based solely on survey data) with adequate levels of precision for the nation as a whole and for major population subgroups. As a result, the U.S. PIAAC, sponsored by the National Center for Education Statistics (NCES), conducted a supplemental data collection in 2014 with a focus on unemployed and young adults. In addition, NCES planned for a third round of data collection in 2017 with the goal of updating the direct estimates as well as producing county level estimates using the combined data. This paper presents the research conducted to arrive at a design for a nationally representative sample for 2017 that best supports the objective of producing indirect estimates for US counties.

Initial simulations indicated that the PIAAC 2012/14 combined sample did not support the production of county-level estimates. A second analysis considered selecting a new sample for 2017 based on the representativeness of a combined 2012/14/17 PIAAC county sample as compared to the entire frame of counties in the U.S. To accomplish this, we used an overlap control sampling procedure that is applicable to probability proportionate to size (PPS) samples with one unit selected per stratum. To further improve the coverage, Westat considered increasing the measure of size (MOS) for counties and states under-represented in the 2012/2014 sample. In addition, we explored various approaches to stratifying the PSUs to increase the likelihood of sampling the small counties. The evaluation was based on a number of simulations that examined the potential improvement on the coverage rates based on various MOS assignments and stratification schemes. This presentation will report on the findings of the research.

### **Benchmarking Options for Model-Based County-Level Estimation of Agricultural Cash Rental Rates**

Michael Bellow (USDA National Agricultural Statistics Service), Nathan Cruze (USDA National Agricultural Statistics Service), Andreea L. Erciulescu (National Institute of Statistical Sciences and USDA National Agricultural Statistics Service)

Estimates of agricultural cash rental rates at different geographic levels have taken on increased importance over recent years due to their application in rental agreement formulation, farm program administration and related activities. The USDA's National Agricultural Statistics Service (NASS) has developed a model-based approach to estimation of rental rates at the county level that uses two years (not necessarily consecutive) of data from its Cash Rent Survey (CRS) in conjunction with auxiliary information on crop yield, total value of production and productivity indexes. This methodology (which has been incorporated into the agency's operational estimation program) involves separate univariate area-level models for the average and difference of cash rental rates between the two survey years in a Fay-Herriot type formulation with closed-form expressions available for estimating the rental rates and their mean-squared errors. An important issue is efficient benchmarking of county-level cash rents estimates to ensure consistency with 1) official state-level figures, and 2) figures for subdivisions of states known as agricultural statistics districts. The procedure currently employed by NASS is based on methodology developed by Ghosh and Steorts (2013) that involves two-stage benchmarking using a single weighted squared error loss function combining county-level and agricultural statistics district-level loss without making any specific distributional assumptions. In this paper, we discuss results of an empirical study comparing the current approach with two alternative ratio-based benchmarking methods using CRS data, model-based estimates and official figures from 2013-14.

### **Small Area Estimation in the Annual Survey of Public Employment and Payroll (U.S. Census Bureau)**

Bac Tran (U.S. Census Bureau)

The Annual Survey of Public Employment & Payroll (ASPEP), conducted by the U.S. Census Bureau, provides statistics on the number of federal, state, and local government civilian employees and their gross payrolls. The universe of ASPEP is about 90,000+ state and local government units. Every five years ESMD conducts a Census of Governments: Employment. Between two consecutive censuses, ESMD conducts the ASPEP, a nationwide sample survey covering all state and local governments in the United States. The ASPEP survey is designed to produce reliable estimates, for example, the number of full-time and part-time employees and payroll at the national level for large domains. However, it is also required to estimate the parameters for individual function codes within each state. This requirement prompted



us to develop a methodology that employs Small Area Estimation (SAE) methods. The outlier treatments (Trinh & Tran, JSM 2017) will be discussed briefly in this research to improve the quality of the estimates.

#### **County-Level Estimates of Mortality and Natality Indicators from the National Vital Statistics System**

Lauren Rossen (National Center for Health Statistics), Diba Khan (National Center for Health Statistics)

Data from the National Vital Statistics System (NVSS) are often used to examine sub-state variation in various indicators of population health (e.g., birth and/or mortality outcomes). However, there are challenges to examining sub-state variation using data from the NVSS. Birth or death rates at the sub-state level are often unstable, and rates are typically suppressed when based on a numerator of fewer than 20 events. As a result, data are often aggregated over several years or over larger geographic regions to generate stable estimates of birth or death rates. This aggregation may mask important spatiotemporal patterns and trends.

Small area estimation methods can be used with birth and death data to generate smoothed sub-state (i.e., county-level) estimates of important population health indicators in order to explore spatiotemporal variation. We implemented hierarchical Bayesian space-time interaction models using Integrated Nested Laplace Approximation (INLA) in R to generate stable county-level estimates of selected birth and death outcomes from the National Vital Statistics System. The advantages and limitations of these methods relative to other approaches will be discussed.

#### **Bayesian State-level Estimates of Diabetes Prevalence in the United States, 2006-2015**

Diba Khan (National Center for Health Statistics), Rong Wei (National Center for Health Statistics), Yulei He (National Center for Health Statistics), Hee-Choo Shin (National Center for Health Statistics), Donald J Malec (National Center for Health Statistics)

Over the past several years, the prevalence of diabetes has been a major source of concern in the U.S. Diabetes is a costly disease and increases the risk of heart disease, kidney failure, hearing and vision loss. Recently, geographic variation in diabetes prevalence at the county level has been explored using data from the Behavioral Risk Factor Surveillance System (BRFSS). We use National Health Interview (NHIS) survey data from 2006-2015 to explore geographic variation in the diabetes prevalence at the state level. Small area estimation methods are used to predict diabetes prevalence estimates at the state level. Results may help identify states with significantly high or low diabetes prevalence. This method can be applied to analyze a variety of outcomes from the NHIS.

## Concurrent Session C-5

### New Methods to Evaluate Response Error

#### **Measuring Systematic Income Misreporting Heterogeneity Using a Novel Dataset**

Christian Imboden (University of Oregon), John Voorheis (U.S. Census Bureau), Caroline Weber (University of Oregon)

We compare administrative income records from the Internal Revenue Service (IRS) and Social Security Administration (SSA) with responses to the Current Population Survey Annual Social and Economic Supplement (CPS ASEC) from 2001 through 2013. We compare the earnings distributions across these two data sources in a regression framework that allows us to isolate household characteristics, including demographics and primary earnings source, that are related to earnings misreporting. We find that predominantly self-employed households report substantially more earned income on the CPS ASEC than in the IRS or SSA administrative records, whereas wage-earners report much more similar amounts, on average, in both datasets. This large discrepancy between the sets of income reports for self-employed households may be partially due to the lack of employer-generated income statements for these units, and is consistent with the findings of random IRS audits (IRS Data Book, 2015). Separately, we find that the groups most likely to underreport wage incomes on the CPS ASEC are historically disadvantaged demographic groups. These findings add to a growing body of knowledge highlighting the potential for administrative records to improve survey measurement, especially for groups that appear to systematically misreport earnings. These findings are also important when those misreporting groups factor into the estimation of key labor phenomena, such as the black-white wage gap.

#### **Respondents' Reporting Accuracy of Social Security Benefits and its Implications in the Health and Retirement Study**

Irena Dushi (Social Security Administration), Howard Iams (Social Security Administration)

The main goal of this paper is to assess whether and to what extent is the amount of Social Security benefits reported accurately among older respondents in the Health and Retirement Study (HRS). This is an important issue because using survey reports may lead to biased estimates of the Social Security benefits and of the share of income received from Social Security among the elderly beneficiaries. Furthermore, underreporting of Social Security benefits may affect estimates of well-being of the elderly such as the proportion of beneficiaries classified as poor or near poverty.

#### **Using Administrative Records to Evaluate Absolute and Relative Reporting Accuracy in Surveys**

Joanne Pascale (U.S. Census Bureau), Kathleen Call (State Health Access Data Assistance Center), Angela Fertig (University of Minnesota), Don Oellerich (Health and Human Services)

Before and after implementation of the Affordable Care Act in 2014, the research community studied ways to measure whether and how individuals obtain coverage. Decades of research on measurement error pre-dating health reform indicate that reports of specific coverage type are problematic. Examples include respondents who report the same coverage twice but as two different types of coverage; report the wrong type of public coverage due to the similarity of program names (e.g., Medicare and Medicaid); or report public coverage as private and vice versa. Rigorous validation studies – where

survey reports are matched with some kind of outside “truth” source on coverage type – are relatively rare. One type of public program (Medicaid primarily for low-income) has been studied extensively (in part due to a centralized, national database of enrollees), but other coverage types lack this kind of database. Thus, little is known about how misreporting of one type of coverage affects other types of coverage.

**Understanding Differences in the Disability Prevalence Across Federal Surveys: Why the 2014 Survey of Income and Program Participation Stands Out**

Heide Jackson (U.S. Census Bureau), Danielle Taylor (U.S. Census Bureau)

Although many federal surveys use a common set of questions to measure disability, Wave 1 of the 2014 Survey of Income and Program Participation (SIPP) had a disability prevalence substantially higher than other contemporaneous federal surveys. This paper will seek to explain why the 2014 SIPP had a higher disability prevalence than other federal data sources. We draw upon survey and collection data from the 2014 SIPP Panel and compare the disability prevalence in the 2014 SIPP Panel with the disability prevalence in the 2008 SIPP, the 2010-2014 American Community Survey, and the 2010-2014 National Health Interview Survey. We will use decomposition techniques to quantify how much of the difference in the disability prevalence is related to differences in the characteristics of the samples recruited in the 2008 and 2014 SIPP panels. Regression and descriptive analyses will be used to test whether errors in data collection may have resulted in high disability prevalence. Preliminary findings suggest the high disability prevalence in the 2014 SIPP is a function of how disability status was reported and not of: weighting, changing sample composition, or data collection error. While the high disability prevalence in the 2014 SIPP is unexpected, SIPP’s rich information and relatively large sample of people with disabilities may make it a useful resource for researchers studying disability.

## Concurrent Session D-1

### Hierarchical Bayes Small Area Estimation for Domains Defined by Demography, Geography or Industry

#### **Small Domain Estimation Using Probability and Non-Probability Survey Data**

Adrijo Chakraborty (NORC at the University of Chicago), Nada Ganesh (NORC at the University of Chicago)

Researchers are often interested in combining information from probability samples with larger non-probability surveys, in order to increase overall sample size for domains of interest. Probability surveys can be designed to produce unbiased estimates, whereas non-probability survey estimates are typically biased. In this talk, we discuss methods for combining information from probability and non-probability surveys based on well-known models used in small area estimation. This approach allows us to estimate the bias associated with the non-probability survey estimates. We consider a hierarchical Bayesian approach for implementing the small area models. We illustrate our method using a real complex survey data.

#### **On Increasing the Number of County-Level Crop Estimates**

Andreea L. Erciulescu (National Institute of Statistical Sciences and USDA National Agricultural Statistics Service), Nathan Cruze (USDA National Agricultural Statistics Service), Habtamu Benecha (USDA National Agricultural Statistics Service), Valbona Bejleri (USDA National Agricultural Statistics Service), Balgobin Nandram (Worcester Polytechnic Institute)

Area-level and subarea-level models are excellent reproducible tools that combine survey data and ancillary data to produce reliable estimates for areas where survey estimates are available. In this paper, we illustrate methods of constructing estimates for areas where the survey estimates are not available. A state-and-crop-specific subarea-level model is applied to county-level crop acreage, incorporating data from a probability sample collected by the United States Department of Agriculture's (USDA's) National Agricultural Statistics Service and ancillary data from other sources. Crop acreage estimates are constructed for the in-sample and the out-of-sample counties in the state. Implications on benchmarking county-level estimates to state-level targets and implications on the increase in the number of published county-level estimates are discussed.

#### **Small Area Estimation for Measures Related to Tobacco Use and Policies Using the Tobacco Use Supplement to the Current Population Survey**

Benmei Liu (National Cancer Institute), Isaac Dompok (Census Bureau)

To evaluate tobacco control programs, monitor progress in the control of tobacco use, and conduct tobacco-related research, policy makers, cancer control planners and researchers often need county level data for tobacco related measures. However, government-sponsored large scale surveys such as the Tobacco Use Supplement to the Current Population Survey (TUS-CPS) is mainly designed to produce reliable estimates at the national and state levels. The survey doesn't have sufficient sample at county level to support estimates with adequate precision.

Model-based small area estimation techniques are thus used to derive estimates for all counties. The key idea is to combine information from a variety of relevant sources to form indirect estimators that generally increase the effective sample size and therefore increase precision, and enable predictions in the absence of sample. This paper describes the methodology used and presents some of the results.

Across all the different outcomes, the modeled estimates have shown consistency with direct estimates in the aggregate, and reduce variance for each county in a general sense. Extensive model selection and diagnosis procedures are necessary to validate the small area model chosen because all the inferences rely on the model. Such methods can be readily applicable to other applications with similar goals.

### **Small Area Co-Modeling of Point Estimates and Their Variances for Domains in the Current Employment Statistics Survey**

Julie Gershunskaya (Bureau of Labor Statistics), Terrance Savitsky (Bureau of Labor Statistics)

Every month, the Bureau of Labor Statistics publishes estimates of employment from the Current Employment Statistics survey at the state and national total levels, as well as at various detailed levels by industry and geography. For smaller domains, where the direct sample-based estimates are not reliable, estimates are produced using models. In this paper, we consider the area level Fay-Herriot model and explore several alternatives, such as: (i) co-modeling of the variances of the direct estimators instead of adhering to the traditional assumption of the “fixed and known” variances; (ii) accounting for possible deviations from the normality assumption of the random effects by assuming the mixture of the normal distributions.

We adopt the hierarchical Bayesian paradigm for development of the models. The code is implemented in the Stan modeling language (Gelman et al., 2015) using a Variational Bayes algorithm, making the computation fast and feasible for use in the monthly production environment.

We compare the performances of alternative models on both synthetic data and in application to estimates from several years of the Current Employment Statistics survey.

## **Concurrent Session D-2**

### **Probing Nonresponse Bias in Surveys – Measures and Consequences**

#### **Lessons from Nonresponse Bias Studies Involving Federal Surveys**

Elise Christopher (National Center for Education Statistics), Morgan Earp (Bureau of Labor Statistics), Tala Fakhouri (National Center for Health Statistics), Steven Frenk (National Center for Health Statistics), Kathryn Piscopo (Substance Abuse and Mental Health Services Administration), Peter Miller (U.S. Census Bureau)

The growth in unit nonresponse and the costs involved in lowering its impact have become threats to the Federal survey enterprise. Unit nonresponse is a problem because it undermines confidence in

inferences from probability samples. Nonresponse may produce bias in survey estimates and reduce the size of samples available for analysis. The relationship between unit nonresponse and bias in survey estimates is complex. Lower unit nonresponse does not automatically produce different survey estimates, as some estimates may show bias while others do not.

We need a better overall understanding of the relationship between unit nonresponse and nonresponse bias. This paper presents a systematic review of several dozen nonresponse bias studies involving federal surveys published in peer reviewed outlets after OMB issued guidelines for nonresponse bias studies in 2006. Through a systematic coding of characteristics of these studies, we will document the methods employed to analyze this relationship, assessing their commonalities and their strengths and weaknesses. The methods to be reviewed include how information about nonrespondents is obtained, how bias is measured, and how the researchers may or may not take into account the possible mitigating effect of post-survey nonresponse bias adjustments in their conclusions about the extent of nonresponse bias. The aim of the paper is to summarize current knowledge about the effect of unit nonresponse on estimates in federal surveys.

### **Nonresponse Bias Indicators and Adjustments Employed in Federal Surveys**

Joseph Schafer (U.S. Census Bureau), Stephanie Coffey (U.S. Census Bureau), James Dahlhamer (National Center for Health Statistics), Jeffery Gonzalez (Bureau of Labor Statistics and Peter Miller (U.S. Census Bureau)

Measuring and mitigating nonresponse bias are increasingly significant activities in Federal surveys. This paper reports lessons from case studies examining nonresponse bias indicators and weighting adjustments employed in several Federal surveys. The surveys examined present varied circumstances – cross sectional vs. longitudinal, establishment vs. household -- with differing auxiliary data available for use in constructing bias indicators or making adjustments. In the case studies, our aim is to catalog and assess current methods for estimating nonresponse bias and adjustment and to suggest alternative practices. How are indicators of bias quantified across surveys with different designs and different kinds of auxiliary information? What do indicators of bias show with and without weighting adjustments? How do “one-size-fits-all” weighting adjustments perform for different estimands (e.g. means and totals vs. ratios and regression coefficients). Can we improve on these adjustments with custom calibrated weights for specific analyses? Responses to these questions, based on the case studies will offer some insight into current practice and a template for an expanded research effort.

### **Alternative Indicators of the Risk for Nonresponse Bias**

Raphael Nishimura (Abt Associates), James Wagner (University of Michigan), Michael Elliott (University of Michigan)

The growth of non-response rates for social science surveys has led to increased concern about the risk of non-response bias. Unfortunately, the non-response rate is a poor indicator of when non-response bias is likely to occur. Beside response rate, we consider a set of indicators that have recently been introduced in the literature as alternative measures of non-response bias, including three measures that use auxiliary variables only: variability of non-response weights/R-indicators, coefficient of variation of subgroup response rates, AUC/pseudo-R<sup>2</sup>; and two measures that focus on auxiliary and survey variables together: fraction of missing information (FMI), and correlation between non-response weights and survey variables. A large-scale simulation study is used to explore how each of these indicators

performs in a variety of circumstances. Although, as expected, none of the indicators fully depict the impact of non-response in survey estimates, we discuss how they can be used when creating a plausible account of the risks for non-response bias for a survey. We also describe an interesting characteristic of the fraction of missing information that may be helpful in diagnosing not-missing-at-random mechanisms in certain situations.

## **Concurrent Session D-3**

### **Novel Approaches to Using Federal and Local Data**

#### **Examining the Utility of Educational Administrative Records for Research and Improving Survey Operations: A Pilot Project at the U.S. Census Bureau**

Nikolas Pharris-Ciurej (U.S. Census Bureau), Quentin Brummet (U.S. Census Bureau), Thurston Domina (University of North Carolina, Chapel Hill), Andrew Penner (University of California, Irvine), Emily Penner (University of California, Irvine), and Sonya Porter (U.S. Census Bureau)

This presentation discusses a pilot project in the Center for Administrative Records Research and Applications at the Census Bureau to assess the feasibility of obtaining educational administrative records and to evaluate their utility for research that informs survey and decennial census operations. Using educational administrative records in survey and decennial census operations has the potential to reduce respondent burden, lower costs, and improve data quality, but much research is needed in order to understand the potential of these data. In particular, we will present a specific application of these data and then discuss broadly the promise these data hold for improving data quality.

We first discuss a project that uses educational records from the state of Oregon and school districts in California linked to Internal Revenue Service data to investigate the degree to which schools' National School Lunch Program (NSLP) data aligns with students' Internal Revenue Service (IRS) reported household income. Free or reduced-price lunch program enrollment is based on NSLP income-eligibility thresholds and student and family reported income. Preliminary results, however, show that school reports of students' NSLP enrollment aligns poorly with measures of NSLP eligibility constructed from students' IRS-reported household income.

We then discuss how educational administrative records may benefit survey and decennial census operations, including an ongoing project to examine occupational coding of teachers using these data. In addition to this ongoing project, we showcase the capacity of these data to enhance survey frames, supplement respondent collected data, and aid in refining survey items.

#### **Leveraging Access to and Use of Department of Defense (DoD) Data: A Case Study of Unraveling Military Attrition Through New Approaches to DoD Data Integration**

Josh Goldstein, David Higdon, Sallie Keller, Bianica Pires, and Stephanie Shipp (Social and Decision Analytics Laboratory, Biocomplexity Institute of Virginia)

The Department of Defense and each of the military services collect a vast amount of administrative and survey data about the military population yet do not readily make these data available for research or for developing standard reporting based on cross-sectional and/or longitudinal analyses. Recent

National Academies reports highlight the need to utilize the vast amounts of administrative and survey data that DoD collects for exploratory research programs. They note that the Army could use these data to address questions about how social and organizational factors affect the behavior of individuals and their units. It is proposed that the Army Analytics Group's Person Data Environment (a data enclave that allows researchers to remotely access confidential data) be examined to determine the feasibility of the approach for data-driven analysis.

A primary goal for this research is to assess the ability to access DoD data sources for use outside of current operational purposes by demonstrating the value of leveraging these data to study important recurring issues to the military. This is done within the context of a case study on military attrition, with a focus on the Army. A conceptual model for individual, unit, and organization interaction is developed. This is placed in the context of military attrition and statistical models are developed for assessing the patterns, outcomes, and consequences of attrition on the Army based on DoD administrative and survey data combined with federal will assesses (1) the complexity of accessing the DoD data, (2) the challenges in repurposing the data to support the research questions, (3) the application of statistical methods not traditionally used with these data, and (4) the path forward for developing a data framework that can support researcher access for on-going cross-sectional and longitudinal DoD analyses.

#### **Measuring Innovation in New Ways using Non-Traditional Data Sources**

Stephanie Shipp, Sallie Keller, Gizem Korkmaz, and Bianica Pires (Social and Decision Analytics Laboratory, Biocomplexity Institute of Virginia), and Carol Robbins (National Center for Science and Engineering Statistics, National Science Foundation)

The National Science Foundation (NSF), through the National Center for Science and Engineering Statistics (NCSES), is exploring the application of analytics and research to measure Science and Engineering Indicators using non-survey sources of data. In partnership with Virginia Tech, the research focuses on non-statistically designed data sources and repurposing these data to measure concepts of interest, such as innovation, science, technology, engineering, and math (STEM) pathways, and currently unmeasured goods and services, such as open source software. We use our Data Science Model to assess the quality of these alternative data, including representativeness, timeliness, accuracy, consistency, completeness, reliability, and relevance; to profile, prepare, and link data; and to explore the data and use it in statistical analysis.

Novel techniques to capture these data and to create new measures are tested and refined. Data collection methodologies beyond traditional survey collections, including exploring opportunity data (e.g., social media) with spatial or temporal designs, are being developed such that metric creation from these data can be repeated. This research aims at benefiting NCSES in continuing its development of data collections and acquisitions to provide better statistical products to its data users.

The findings from three projects will be briefly described. These include (1) measuring company innovation through opportunity data – data discovery is instrumental in assessing whether companies tell their innovation story publicly and if so how, such as through company websites, portal, news media, blogs, and databases that aggregate company information; (2) identifying non-traditional STEM pathways – finding and assessing the quality of new data sources that measure post-baccalaureate and non-PhD education and assessing the quality of the data; and (3) examining the feasibility of measuring the outputs of open source software as a first step to measuring its value to the economy.



**Interdisciplinary Insights for Investigating the Intersection of Race/Color and Social Outcomes among Diverse Hispanics Communities: Implications for Statistical Measurements and Analysis**

Howard Hogan PhD (U.S. Census Bureau), Nancy López PhD (University of New Mexico), Ruth Enid Zambrana PhD (University of Maryland, College Park)

What does research in demography, economics, political science, sociology, health and education tell us about the intersection of the race/color and social outcomes within diverse Hispanic communities? What difference does question format make in our ability to detect whether there is a “color line” within these communities? The goal of this presentation is to engage researchers, practitioners, community members, and policy makers in a conversation about what we know from interdisciplinary research and the implications that question format have for providing the data necessary to develop policies that address the needs of different communities.

A guiding premise of the conversation is that the collection of demographic data on race, gender, class, ethnicity and other axes of stratification is primarily for use by the Federal government in enforcing the civil rights laws. These data are used to monitor historic and ongoing inequalities in multiple sectors including voting rights, education, employment, housing, criminal justice, and law enforcement. Our presentation will summarize a cross section of disciplines (e.g., demography, economics, political science, sociology, education, health) on the intersection of race/racialization, social inequalities and the color line within Hispanic communities. We will also provide a content analysis of the trajectory of data policy on classifying Hispanic origin and race as it relates to both the legislative/legal and larger social context.

We conclude with suggestions for research to improve question formats, including alternative question formats that could be used to understand social outcomes.

## Concurrent Session D-4

### Current Research for Improving Race and Ethnicity: Federal Data Maintenance; Collection; and Presentation

**Research Examinations within the Federal Statistical System for Improving Federal Race/Ethnicity Data**

Carolyn Hronis (Energy Information Administration)

This presentation will describe the particular areas where research has been conducted with the goal to improve the quality of race and ethnicity information collected and presented by Federal statistical system agencies. Specifically, four areas have been researched, including:

- The use of two separate questions versus one combined question to collect race and ethnicity, and changes to question phrasing to reduce race and ethnicity nonresponse;
- The classification of a Middle Eastern and North African group and establishment of it as a minimum category that is distinct from the White category;

- Clarification of agency requirements and flexibilities for the collection of detailed race and ethnicity data; and
- The terminology used for race and ethnicity classifications and other language in the current standards

### **Update on Race and Ethnicity Question Proposals for the 2020 Census**

Nicholas Jones (U.S. Census Bureau)

This presentation will provide an update on the U.S. Census Bureau's race and ethnicity question proposals for the 2020 Census. Following the 1997 race and ethnicity standards for the federal government set by the U.S. Office of Management and Budget, the planned race and ethnicity questions for the 2020 Census will use a separate questions format to collect data on race and ethnicity. The presentation will show and discuss the questions that are planned for submission to Congress by March 31, 2018. The presentation will point out the 2020 Census design features which include some improvements in comparison to the 2010 Census race and ethnicity questions, based on research from the 2015 National Content Test, and discuss how respondents can self-identify their race/ethnicity with these new design enhancements.

### **Potential Improvements for 2020 Census Race and Ethnicity Conversion and Bridging**

Benjamin Bolender (U.S. Census Bureau)

Often, race data is collected in one set of categories and used in another. The Population Estimates and Projections area of the U.S. Census Bureau receives vital statistics data that follow the 1977 Office of Management and Budget (OMB) standards, but we develop population estimates in the 1997 standard categories (some states still reported in the old categories until 2016). To transition between the two schemas, the Census Bureau has relied on race bridging and imputation. This project focuses on a description of the various techniques currently used by the Census Bureau and highlights recent proposals for major improvements to the process. Using a combination of record linkage, imputation, new bridging factors, survey data, and the area's innovative Demographic Characteristics File, we look to develop a new kind of bridging that allows for more accurate results and the ability to update factors as we move into the future.

## **Concurrent Session D-5**

### **Advancing Disclosure Limitation Methods in Federal Data Releases**

#### **Towards Developing Synthetic Datasets for the Economic Census**

Katherine J. Thompson (U.S. Census Bureau), Hang Kim (University of Cincinnati)

The U.S. Census Bureau is researching the development of synthetic industry-level microdata on a subset of industries covered by the Economic Census (EC). The EC collects a core set of data items from each establishment in a sector (core items), as well as other industry-specific items. To ensure consistency, key collected items are subjected to predetermined edit rules which are often written in

the form of linear constraints: ratio edits (restrictions) on pairs of items; range edits on single items; and balance edits/additivity constraints. Items that violate these edit rules are replaced with consistent (imputed) values.

To create synthetic economic data under these restrictions, Kim, Reiter, and Karr (2018) developed a two-stage process, utilizing a nonparametric Bayesian prediction model and multiple imputation. This approach preserves correlation structure, allowing for more sophisticated second data analyses than simple tabulations. The prior research was performed on industries in the manufacturing sector and applied the same set of edit rules all establishments in a sector (in reality, edits can be modified or disabled for part-year reporters).

In this paper, we apply their synthetic data generation approach to industries in other sectors. The Economic Census is a multi-purpose program whose macrodata are used to produce benchmark estimates of the economy and whose microdata are used for economic models. In addition to satisfying the predetermined set of edits imposed by the Economic Census on the studied data items, the synthetic data must yield marginal totals approximately equal the published margins in industry levels and satisfy formal privacy restrictions. In this presentation, we discuss the nuanced challenges in modeling viable multipurpose synthetic data, presenting examples and selected results along with proposed revisions to the original modeling procedures.

#### **Formal Privacy and Synthetic Data for the American Community Survey**

Michael H. Freiman (U.S. Census Bureau), Rolando Rodriguez (U.S. Census Bureau), Jerome P. Reiter (Duke University, U.S. Census Bureau), Amy Lauger (U.S. Census Bureau)

The U.S. Census Bureau is expanding the use of formal privacy to improve disclosure avoidance methods and to enable quantification of privacy loss. This presentation discusses the particular challenges of applying formally private algorithms to the American Community Survey (ACS), including the data's high dimensionality coupled with sample size limitations and the use of complex survey weights. We describe research on model-based approaches to creating synthetic data for the ACS, focusing on topics such as health insurance and employment.

#### **An Integrated Approach to Providing Access to Confidential Data**

Jerry Reiter (Duke University and U.S. Census Bureau)

I describe a vision for an integrated system for disseminating microdata collected by government agencies. The system includes (i) capability to generate synthetic data intended for wide access, coupled with (ii) means for approved researchers to access the confidential data via secure remote access solutions, glued together by (iii) a verification server that allows users to assess the quality of their analyses with the synthetic data so as to be more efficient with their use of remote data access. I describe verification measures that satisfy differential privacy. I apply the framework to data from the Office of Personnel Management, examining pay differentials in the federal government.

**Challenges and Experiences Adapting Differentially Private Mechanisms to the 2020 Census**

Simson L. Garfinkel (U.S. Census Bureau)

With public and commercial access to ‘big data’ sources accelerating and already at historically unprecedented levels, with widespread access (e.g. through Cloud services) to cheap, large-scale computing clusters, and with increasing sophistication of statistical/mathematical/machine-learning techniques, traditional procedures for limiting disclosure from published survey products are increasingly inadequate. In this context, the mathematically provable guarantee of privacy protection provided by differential privacy (DP) – in particular, its robustness to arbitrary ‘attacker’ prior knowledge, theoretical savvy or computational capacity – offers a natural, rigorous path forward for provably protecting the privacy of survey respondents. Recognizing this need, a team of experts at the United States Census Bureau has been tasked with adapting state-of-the-art DP methods to protect the 2020 Decennial Census.

Although DP is a natural framework for providing rigorous privacy guarantees in the face of increasing sophistication of possible attackers, it is challenging to adapt well-studied DP mechanisms to a product with the size and complexity of the Decennial Census while maintaining reasonable accuracy. Rendering the Decennial Census’s data products is complicated by a number of factors, including: 1) high-sensitivity queries, 2) nesting of people within households, 3) generating integer microdata from which queries may be answered, 4) enforcing nonnegativity without inducing enormous positive bias in high-order sums, 5) efficiently capturing sparsity patterns, 6) efficiently modeling variables with an unbounded number of levels, 7) exactly maintaining publicly known constraints at multiple levels of a large geographic hierarchy. In this talk, we will describe our efforts to overcome these challenges and to satisfy the stakeholder requirements.

## Concurrent Session E-1

### Cognitive Lab, Usability, and Survey Development Research

#### **Evidence-based Standards and Guidelines for Mobile Survey Instrument Design**

Lin Wang (U.S. Census Bureau), Christopher Antoun (U.S. Census Bureau), Russell Sanders (U.S. Census Bureau), Elizabeth Nichols (U.S. Census Bureau), Erica Olmsted Hawala (U.S. Census Bureau), Brian Falcone (U.S. Census Bureau), Ivonne Figueroa (U.S. Census Bureau), Jonathan Katz (U.S. Census Bureau)

Mobile devices have become an important tool in survey data collection. However, due to the small screens of such devices, we face the challenge of usability in mobile survey instruments. The user interface (UI) design of mobile survey instruments is crucial to survey data quality because it is the UI through which the respondent reads survey questions and provides responses. To guide mobile survey instrument UI design for ensuring and improving the quality of data entry, we need implementable standards for the basic elements of mobile survey instruments (e.g., touch target size) and guidelines for the building blocks of mobile survey instruments (e.g., date of birth format). The U.S. Census Bureau initiated a project in 2015 to develop a suite of standards and guidelines for mobile survey instrument UI design. The standards and guidelines are based on empirical evidence. The evidence was generated through either literature review or behavioral experiments: First, a literature review was conducted for each standard and guideline topic to collect evidence from existing empirical studies in the public domain. For topics that did not have sufficient evidence, we designed and conducted behavioral experiments to generate empirical evidence.

In this paper, we present 9 standards and 28 guidelines for UI design of self-administered mobile survey instruments. The standards cover text presentation, touch target sizing, luminance contrast and color combination between foreground and background in display screen. The guidelines address issues in the following areas: question stem and instruction presentation, response options design, and screen design. Limitations and future work will also be discussed.

#### **Usability Testing Methodology for the 2017 Economic Census Web Instrument**

Rebecca Keegan (U.S. Census Bureau), Temika Holland (U.S. Census Bureau), Krysten Mesner (U.S. Census Bureau), Aryn Hernandez (U.S. Census Bureau)

The Economic Census is the U.S. Census Bureau's benchmark survey of the American economy, collected every 5 years from millions of American business establishments in more than 1,000 industries. For the 2017 Economic Census, the Census Bureau created a new, entirely web-based data collection instrument. This instrument underwent extensive usability testing with participants at their place of business, representing a wide range of industries and company sizes.

Participants completed several tasks pertinent to the actions necessary for responding to the actual survey. In addition, a screen recording software was used to record the actions and words of the participants as they moved through the tasks. This enabled a multi-method approach to assessing the usability of the new instrument.

The metrics observed during testing evaluated the instruments' performance in terms of efficiency, accuracy, and user satisfaction. Efficiency was measured by time on task, accuracy was measured by

successful completion of the task, and user satisfaction was measured using the system usability scale administered to each participant.

This paper will describe the methodologies and metrics used to assess this new online instrument, focusing on the value added by this multi-method approach. It will also discuss selected findings, identifying problematic areas for users, and recommendations for improvements to the design of the instrument.

### **Defining Bullying: A Split-Ballot Experiment Across Three Federal Agencies**

Melissa Cidade (Avar Consulting), Rachel Hansen (National Center for Education Statistics, Department of Education)

Bullying in America's schools is a primary concern for several federal agencies, including the Department of Education, the Department of Justice, and the Department of Health and Human Services, among others. However, examining different sources of bullying victimization data leads to different understandings of the magnitude of the issue. What these estimates obfuscate is that each of these surveys uses a different definition of bullying.

To ameliorate the issue of apples-to-oranges comparisons in the rate of bullying victimization, a working group of federal partners collaborated to determine a uniform definition of bullying, published by the Center for Disease Control (CDC) in 2014. In response to this uniform definition, researchers at the Department of Education changed the way they asked about bullying in school on the 2015 collection of the School Crime Supplement (SCS) to the National Crime Victimization Survey (NCVS). Since the SCS has been conducted biennially since 1999, researchers had a quagmire: how do they continue to collect the trend data on national bullying estimates while bringing the questionnaire into alignment with the newly published CDC definition? The answer was to embed a split-ballot experiment on the 2015 administration of the SCS.

Our project is to outline the methodological and institutional challenges to changing the way that bullying victimization is asked on the SCS. First, we will discuss the concepts embedded in the CDC definition, and ways to operationalize them to survey questions. Then, we will outline the research procedures used to develop and field the split-ballot experiment. We will conclude with results of the experiment, as well as discuss our experiences with best practices for coordinating changes in survey instruments across the three federal agencies involved in the administration of the SCS.

### **Health Insurance in the American Community Survey: Multiple Types of Coverage and Respondent Write-ins**

Edward R. Berchick (U.S. Census Bureau), Monica S. Wiedemann (U.S. Census Bureau)

Estimates of health insurance coverage rates depend on national surveys, such as the American Community Survey (ACS). In addition to offering a series of coverage-type checkboxes, the health insurance question on the ACS allows respondents to write in information about their health insurance coverage. Respondents may choose to use the write-in field for a variety of reasons. They may not know how to classify their health insurance coverage type and/or they may want to provide additional information about the coverage type that they reported via checkbox. Previous research suggests that the use of the write-in field contributes to an overestimation of direct-purchase insurance in the ACS

(Mach & O'Hara 2011; Lynch & Kenney, 2011). Changes to the health insurance landscape that have occurred in recent years may have affected respondents' reporting habits. In this paper, we examine the use of write-ins in the ACS. We focus on the demographic predictors of write-in use and consider the relationship between write-in use and multiple types of health insurance coverage. Results provide additional information on how respondents report their health insurance coverage in surveys.

### **Improving the Anchoring Vignette Methodology with Visual Vignettes**

Mengyao Hu (University of Michigan), Sunghee Lee (University of Michigan)

Self-assessment questions are routinely used in surveys. Ideally, responses from these questions reflect only the responding individuals' true state. This, however, is rarely the case. Respondents' answers to self-assessment questions not only reflect their true state, but also how respondents use the scales, also known as reporting heterogeneity. To address reporting heterogeneity particularly in cross-cultural comparisons, King et al. (2004) proposed the use of anchoring vignettes, which are a set of questions, each describing a hypothetical person's situation related to the construct measured in a few sentences. Anchoring vignettes have become widely used in comparative studies. However, there are several critical practical challenges associated with this method. In particular, vignettes significantly increase survey time and the cognitive burden placed on respondents given the complexity of verbal vignette descriptions. Further, the fulfillment of the measurement assumptions is not often assured.

To address these issues, we introduce a new technique - visual anchoring vignettes, which use images to convey the information about vignette persons, eliminating the long verbal descriptions. The use of visual vignettes offers two main advantages: 1) by omitting long verbal descriptions, we can significantly reduce respondents' cognitive burden, and reduce survey time, and 2) with the ability to standardize and customize information in the visual presentation of the vignette items, the measurement assumptions are more likely to be met. This study aims to 1) design visual anchoring vignettes and 2) compare the performance of visual and standard verbal vignettes. Data was collected through an online survey experiment including a sample of White, African American and Hispanic respondents. We focused on four health domains, sleep, affect, mobility and pain. The findings of this research will contribute to not only survey methodology but also epidemiology and public health, with the general implications for future design of anchoring vignettes.

## **Concurrent Session E-2**

### **Combating Nonresponse an Update from the Field**

#### **Comparing Response Rates Across Surveys**

John Dixon (Bureau of Labor Statistics), Benjamin Cover (Bureau of Labor Statistics), Kirk Hagemeyer (Bureau of Labor Statistics), Nicholas Johnson (Bureau of Labor Statistics),rew Kato (Bureau of Labor Statistics), Randall Powers (Bureau of Labor Statistics), Demetrio Scopelliti (Bureau of Labor Statistics), Jason Tehonica (Bureau of Labor Statistics)

Historically, survey response rates have been used as a measure of how representative a survey is of the sampled population. There are many reasons why response rates differ among surveys including the content of the survey, the amount of time available to collect data, the mode of data collection, and

constraints on who can respond. Research has shown an inconsistent relationship between survey response rates and nonresponse bias, but changes in response patterns can be informative. For example, changes in patterns that are similar between surveys might indicate a common cause, while rate changes unique to a survey might point to a change in methods. This report groups similar surveys to allow for comparisons. For example, response rates from surveys collected from randomly sampled households are shown together while those collected from establishments are together. This presentation will describe the process and issues involved in collecting, computing, and comparing response rates across multiple surveys, as well as different approaches for visually displaying the data.

### **Moving Towards a User-Friendly Expenditure Diary Survey**

Safia Abdirizak (Bureau of Labor Statistics), Brett McBride (Bureau of Labor Statistics)

One of the goals of surveys at the BLS is to collect accurate data without unduly burdening the providers of that data. The Consumer Expenditure Survey (CE) makes an effort to respect this goal by undergoing periodic revisions to its questionnaire and procedures. In January of 2017, we implemented the latest set of revisions that relaxed diary placement procedures and redesigned the layout of the diary. The CE diary survey is composed of two independent diaries that each are used to collect one week of household expenditures.

Once revised, diary procedures no longer required interviewers to place each diary separately, removing a between-weeks visit. In addition, we eliminated restrictive placement procedures that limited interviewers to a 7-day window for diary placement. These procedural changes were motivated by a desire to facilitate interviewers' efforts to contact respondents. In this presentation, we examine the contact rates, level of interviewer effort, and respondent participation associated with those changes.

The redesigned diary streamlined the data entry process by revising how sections were displayed on the form. The previous design had sections spanning multiple pages whereas the current diary has all sections viewable across two pages. The purpose of this revision was to limit the diary-keeping burden placed on respondents. We looked at the change in number of expenditures reported in each diary section before and after the form revision. This presentation can inform how procedural and design changes can affect participation rates and the data that are collected.

### **Stemming the Rising Tide of Nonresponse**

Barbara Rater (National Agricultural Statistics Service), Linda J Young (National Agricultural Statistics Service)

USDA's National Agricultural Statistics Service (NASS) conducts over 500 state and national agricultural surveys a year and conducts a census of the nation's 2.0 million farmers once every five years. At the heart of data collection is a statistical agency's ability to persuade a respondent to voluntarily respond to surveys. Over the last 15 years, response rates have declined on average of 10-to-15 percent on average. In January 2016, a cross-agency team was chartered and charged with finding ways to strengthen survey processes, decrease respondent burden and improve survey participation. To increase buy-in and support of reluctant respondents, NASS established communication and outreach strategies to strengthen relationships. NASS improved interviewer training and delivery to develop more effective and confident interviewers. The team created and documented more consistent processes for more efficient pre-survey activities and have expanded the use of response history to



apply the appropriate data collection strategy. Efforts to create a more accurate sampling pool to reduce the number of non-respondents are underway. To optimize the number of times a potential respondent is contacted and help reduce the number of survey refusals and inaccessibles, the agency is using available data to develop a multidimensional respondent burden index. As a consequence, early analysis of survey response rates indicates that the team's efforts to strengthen survey processes and decrease burden are paying off and response rates are stabilizing. Further research and evaluations are needed. This presentation describes the challenges faced during the initial phase of the project, actions taken for improvements of the data collection process, and results of studies designed to assess the efficacy of proposed solutions.

### **Cleaning Out the Gutter: Identifying and Eliminating Deadwood from a Sampling Frame Using Trees**

Andrew J. Dau (National Agricultural Statistics Service), Gavin R. Corral (National Agricultural Statistics Service), Jodie M. Sprague (National Agricultural Statistics Service), Linda J. Young (National Agricultural Statistics Service)

The National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) publishes more than 400 reports annually that contribute to decision making at both the farm and the farm policy levels. These reports are based on survey data. Many of the survey samples are drawn from the NASS list frame, which is a list of all known farms in the United States. The need for a complete list frame is vital in setting valid and accurate estimates for agriculture. NASS has an on-going process for identifying new farms and adding them to the list frame. However, farms also go out of business and the list experiences aging. This results in the list no longer being current. In this study, we investigate the use of trees to identify aging records (farms), referred to as deadwood, within a sample in advance of data collection. For each record identified as potential deadwood, the survey data are collected in a face-to-face interview. Identifying and removing the deadwood improves future samples and has the potential to lower survey non-response. The efficacy of this approach in identifying deadwood is discussed.

## **Concurrent Session E-3**

### **Issues of Informed Consent for Using Administrative Records**

#### **Pursuing Consent for Record Linkage in an Establishment Survey: Results from a National Survey**

Lauren Harris-Kojetin (National Center for Health Statistics), Manisha Sengupta (National Center for Health Statistics)

Linking survey data to administrative records can help lower survey respondent burden, decrease survey costs, and increase analytical content. Survey-administrative data linkage is increasingly being conducted with population surveys, typically asking survey respondents for consent to provide identifying information about themselves needed to link administrative data to their survey responses. Little has been published on linkage for establishment surveys or obtaining consent from establishment survey respondents to provide identifying information about the people they serve (e.g., schools about students). NSLTCP includes establishment surveys that collect data on providers and the people they serve in two sectors. Because response rates to the 30-minute questionnaire were lower than expected in previous waves, NCHS is considering ways to shorten the questionnaire. A potential option is using

administrative data to replace some of the questionnaire items about people served. We explore establishment survey respondents' willingness to provide protected health information (PHI) about people served that is necessary in order to conduct probabilistic matching to administrative data on these people. The HIPAA Privacy Rule permits HIPAA-covered entities to disclose PHI without patient authorization for public health purposes and for research. In 2016, almost three-fourths of adult day centers and two-thirds of residential care communities were HIPAA-covered entities. One-fifth of providers in both sectors would be willing to provide date of birth for people they served, but less than one-tenth would be willing to provide multiple elements needed to increase the matching rate. Willingness to provide PHI varied by ownership, chain affiliation, and MSA status. For residential care communities, a higher percentage of small places (4-10 beds) would be willing to provide PHI compared to larger places. The presentation will include findings looking at the association between HIPAA status and willingness to provide PHI elements, and discuss possible implications for the viability of linkage.

### **Assessing Consent Bias in Linkage Studies**

Lisa B. Mirel (National Center for Health Statistics), Cordell Golden (National Center for Health Statistics), Cindy Zhang (National Center for Health Statistics)

Linked survey and administrative data can be used to facilitate richer analyses by augmenting the information collected from the surveys with vital or administrative data. However, not all survey participants grant consent for linkage and this can affect the resulting estimates. The National Center for Health Statistics (NCHS) has a data linkage program that is designed to expand the analytic utility of the Center's population-based surveys. In order to conduct the linkages, participants must be linkage eligible which means they provided consent for linkage and have sufficient personally identifiable information (PII) to be included in the linkage algorithm. There has been a growing reluctance of survey participants to provide PII to interviewers. Therefore, in recent years, changes to survey design have been implemented to reduce the amount of PII collected and explicit questions about linkage consent have been added to surveys. This has increased the linkage consent rates but has limited the number of participants who are eligible for linkage based on availability of certain key PII elements. This talk will examine the changes in linkage consent rates and the effect on estimates by assessing bias (e.g., comparing estimates from the full sample to the linkage eligible sample). In addition, we will explore the effects on estimates after adjusting the sample weights for linkage consent. The results will be discussed in terms of implications for analyses and future directions of the NCHS Data Linkage Program.

### **Challenges to Informed Consent from Administrative Data Linkage and Secondary Usage**

Carl Ramirez (U.S. Government Accountability Office)

The growing potential for and scope of survey data linkage to existing administrative data, and secondary usage of survey data for other research purposes present statistical agencies with challenges and risks as well as exciting opportunities. This paper summarizes the regulatory and standards environment (e.g. the "Common Rule" for treatment of human research subjects) and explores possible risks and challenges of two specific practices. It also identifies possible information gaps even in light of high-quality research being done on public attitudes to these practices.

Informed consent and methods for preventing disclosure and re-identification of de-identified data products are key determinants of successful privacy and confidentiality protections. The recently revised draft Common Rule provides new guidance on, but also some exemptions from, requirements for

informed consent in surveys and protections applicable to federal government survey research to prevent microdata from re-identification or unanticipated secondary use.

Effective informed consent may be challenging under this scenario, for example: a prospective research participant begins a survey, under some condition of informed consent. It is voluntary and the respondent can decline to answer any question or end participation at any time. This decision is influenced by what we ask them. But what inferences would respondents make from questions not asked? On the one hand, we reduce burden by not asking for what we can obtain elsewhere, but on the other, how if at all should we inform respondents of this? For example, would knowing that income, if not asked, will become part of their response, affect their decision to answer other questions?

Finally, there are increasing threats to confidentiality through re-identification given the tools and records available to secondary users in today's research environment. Several examples are presented, and possible mitigation policies are discussed.

## **Concurrent Session E-4**

### **Estimation Challenges in Complex Surveys**

#### **Incorporating the Finite Population Correction into the Variance Estimation of a National Business Survey**

Sadeq R Chowdhury (Agency for Healthcare Research and Quality), David Kashihara (Agency for Healthcare Research and Quality), Matthew Thompson (U.S. Census Bureau)

The Medical Expenditure Panel Survey (MEPS) - Insurance Component (IC) is a large national annual survey of private business establishments as well as state and local governments. It is a major source of information on employer-related health insurance in the United States. The sample for MEPS-IC is selected using a single-stage stratified design from two list-based frames. In many sampling strata the sampling rate is very high but the finite population correction (FPC) was never included in the variance estimation until 2016. This presentation will discuss how the FPC factor was incorporated into the Taylor Series variance estimation in various sampling strata with different types of sampling. It will also present the impact of incorporating the FPC on MEPS-IC variance estimates.

#### **Finding an Estimator that Minimizes Revisions in a Monthly Indicator Survey**

Nicole Czaplicki (U.S. Census Bureau), Yarissa Gonzalez (U.S. Census Bureau), Laura Bechtel (U.S. Census Bureau)

Estimates from the Advance Monthly Retail Trade and Food Services Survey (MARTS) are released approximately nine working days after the reference month and provide an early estimate of total monthly sales for certain industries. One month later, the MARTS sales estimate is superseded by the preliminary estimate from the Monthly Retail Trade and Food Services Survey (MRTS); the MRTS estimate may be further revised to incorporate data from late reporters. The MARTS estimates serve as a forecast of the MRTS estimates. Large revisions between the two corresponding estimates is undesirable, especially when the revision reverses the direction of the month-to-month change.

Consequently, the U.S. Census Bureau is investigating methodological enhancements to the current procedures designed to minimize these revisions.

This paper presents an empirical study that investigates whether this can be accomplished by modifying or changing the MARTS estimator, applying four different estimators to six years of historical data. Specifically, we compare the currently-used link relative estimator and a modified link relative estimator (developed by another indicator survey) to two different adjustment weighting estimators whose model assumptions are consistent with the survey data. As part of this research, we consider alternative adjustment cells, incorporating an additional size category into the currently-used industry categories as motivated by our empirical analysis. We also examine estimation effects due to existing analyst imputation procedures of selected units.

For each considered estimator, we assess the predictive effectiveness of the advance (MARTS-based) on the preliminary (MRTS-based) estimates in terms of levels, month-to-month changes, and year-to-year changes. Additionally, we present mean absolute revisions and mean squared revisions to examine the magnitude of revisions. We use these statistics to compare each estimator and make a recommendation accordingly.

#### **Effect of Nearest Neighbor Imputation on Variances Calculated by Fay's Balanced Repeated Replication**

Bradley Rhein (Bureau of Labor Statistics), Leland Righter (Bureau of Labor Statistics), Chester Ponikowski (Bureau of Labor Statistics)

The Occupational Requirements Survey (ORS) is an establishment survey conducted by the Bureau of Labor Statistics (BLS) for the Social Security Administration (SSA). The survey collects 70 data elements that cover information on the vocational preparation and the cognitive and physical requirements of occupations in the U.S. economy, as well as the environmental conditions in which those occupations are performed. Since some sample units are not willing or able to provide data for all data elements, missing data element values are imputed using a nearest neighbor imputation procedure. In cases where there are multiple eligible donors, a random selection process selects a donor. Variance estimates are generated after imputation using the Fay's Balanced Repeated Replication method. Since imputation runs on the full sample before variance estimation, and not for each replicate during variance estimation, the variances are deemed to be underestimated. This paper presents the research results for comparing the current method with a method where imputation occurs at each replicate.

#### **Investigation of the NCHS Data Presentation Standards for Proportions: A Simulation Study**

Frances McCarty, PhD (National Center for Health Statistics), Jennifer Parker, PhD (National Center for Health Statistics)

Producers of official statistics, such as government agencies, often develop and recommend presentation standards to be used for estimates produced from their data products. These standards are intended to ensure that reported estimates are reasonably accurate and precise. NCHS recently released a report that includes statistical standards for the presentation of estimates (specifically, proportions). The goal of this study was to investigate how the new standards for the presentation of proportions would operate. The analyses focus on addressing four main questions: 1. How do 4 different specifications compare with one another in terms of the number of estimates that would be presented?,

2. When estimates are suppressed using the new recommendations, which criterion or combination of criteria most often lead to suppression of an estimate? 3. What types of estimates are being suppressed?, and 4. What sample characteristics are associated with a lower frequency of estimate presentation?

We used a sampling-based study to investigate the impact of selected standards on the presentation of prevalence estimates (proportions). Data obtained from IPUMS Health Surveys were used to create a known population (NHIS years 1997-2015). We generated 1000 samples from the population of records. For each sample, we estimated the prevalence of 7 health related outcomes for a defined subpopulation with estimates based on region and race/ethnicity. We used four presentation standards, 2 based on relative standard error and 2 based on confidence interval width, to determine the presentation status for each estimate. We show the proportion of estimates presented using each standard and use graphical methods to investigate the conditions that lead to presentation/suppression of estimates.

### **An Alternative Way of Estimating a Proportional-Odds Model with Complex Survey Data**

Phillip S. Kott (RTI International), Peter Frechtel (RTI International)

When fitting an ordered categorical variable with  $L > 2$  levels to a set of covariates onto complex survey data, it is common to assume (heroically) that the elements of the population fit a proportional-odds logistic-regression model. This means the probability that the categorical variable is at or below some level is a binary logistic function of the model covariates. Moreover, except for the intercept, the values of the logistic-regression parameters are the same at each level.

The conventional “design-based” method used for fitting the proportional-odds model is based on pseudo-maximum likelihood (PML). We compare estimates computed using PML with those computed by assuming an extended version of the proportional-odds logistic regression model. In this variation from the standard model, the expected values of the model errors are assumed only to be uncorrelated with the covariates, an assumption that can almost always fit the population. We show theoretically and with a numerical example how extended-model estimates and their standard errors can be computed in a design-sensitive manner. An analogous estimation method for the general cumulative logistic regression model easily follows as does a test for the “parallel-lines assumption” underpinning the proportional-odds model (which is a special case of the general cumulative logistic regression model).

When a proportion-odds model fits the population only approximately, an obvious advantage of design-sensitive extended-model estimation over PML estimation is the following. The estimated fraction of the population at or below each level is forced to equal the expected fraction of the population at or below that level.

## Concurrent Session E-5

### Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps

#### **Statistical Methods for Combining Multiple Data Sources**

Trivellore Raghunathan (University of Michigan)

Policy makers and data users are demanding ever-increasing granularity for statistics: wanting more geographic detail, more frequent releases of statistics, and more information about subpopulations. Large-scale probability sample surveys have long been the foundation for producing many national statistics for the United States, but combining survey data with other data sources, or combining multiple administrative data sources, has many potential advantages over the survey paradigm. A number of recent studies have identified information domains that would benefit from drawing on alternative data sources to provide key statistics beyond what is possible or practical through a federal survey. In this presentation we review statistical methods for combining information, identify research needs, and propose steps that can be taken to facilitate a new paradigm for producing federal statistics, which would shift from sole reliance on probability surveys to a system that relies on probability surveys along with administrative and private-sector data, making use of the strengths of each data source. The presentation will provide an overview of statistical methods that have been proposed for combining information, where the choice of method depends on the statistical purpose, the nature of the available data, and privacy and other considerations. When individual records from multiple datasets for each person or entity are available, they can sometimes be linked through statistical models. When aggregate statistics are available or linkage cannot be done, multiple frame methods or modeling can be used. The presentation will also outline research that is needed in the area of statistical methodology and describe a framework for promoting the development of methods for combining data sources.

#### **Quality Frameworks for Statistics Using Multiple Data Sources**

Robert Groves (Georgetown University)

Survey researchers have developed quality frameworks for classifying and examining different potential sources of error in surveys. However, unlike survey data, nonsurvey data sources are not created with the purpose of creating statistics. Thus, combining data from multiple data sources will also require a new or modified quality framework. Some quality dimensions, such as timeliness and granularity have often been undervalued as indicators of quality, but they will become increasingly relevant with statistics based on multiple data sources. Administrative and private-sector data have their own challenges and errors. These errors arise for multiple reasons, such as mistakes in understanding or interpreting metadata; errors in entity linkage; and incomplete or missing information. Unlike handling data from traditional surveys, these errors usually need to be dealt with after the data have been obtained and been through cleaning and processing. The use of administrative and private-sector data not only shifts the focus of reducing errors into the post data gathering stage, it also adds a new error source that is not usually encountered in surveys: linkage errors. The presentation will review frameworks that go beyond the total survey error framework and discuss important dimensions of quality for nonsurvey data and estimates from combining different sources. These frameworks include additional dimensions that better

capture user needs, such as timeliness, relevance, accuracy, accessibility, coherence, integrity, privacy, transparency, and interpretability. The presentation will also discuss tradeoffs between different quality aspects of data and provide examples.

### **Combining Data Sources While Protecting Privacy**

Brian Harris-Kojetin (National Academies of Sciences, Engineering, and Medicine)

Many statistical agencies are currently seeking to evaluate and use multiple data sources to enhance their statistical programs. Moving to an environment in which multiple datasets are combined can change the threats to privacy. Federal statistical agencies are subject to a number of privacy and confidentiality laws that apply to their statistical data, but new legal and policy issues may arise when linking records from different data sources. Because linked datasets can offer greater privacy threats than single datasets, the panel recommends that federal statistical agencies develop and implement strategies to safeguard privacy while increasing accessibility to linked data sets for statistical purposes. This will require collaborative research with academia and industry to develop new techniques to address potential breaches of the confidentiality of their data. The panel also recommends the creation a new entity that will provide a secure environment for analysis of data from multiple sources, coordinate acquisition and use of data, and identify and facilitate research on the challenges that are common across statistical agencies. The entity should follow the principles and practices for federal statistical agencies and permit data access only for statistical purposes. The panel's proposed new entity should assist federal statistical agencies in identifying data sources that can most effectively inform the creation of national statistics, help develop techniques to use those data to compute national statistics while respecting privacy and other protection obligations on the data, and nurture the expertise required for these activities. While adhering to confidentiality, privacy, and data security requirements, statistical agencies and the new entity should strive to provide both federal and external researchers access to data for exclusively statistical purposes in a timely manner that is not administratively burdensome.

## Concurrent Session F-1

### Question Evaluation and Cognitive Interviewing

#### **Cognitive Interviewing Methodology in 2018: Current Trends and Recent Challenges**

Kristen Miller (National Center for Health Statistics)

Since the 1980s, cognitive interviewing methodology has grown to become a widespread and integrated practice for question evaluation within the Federal Statistical System. The method reached the pinnacle of institutionalization when, in 2016, guidelines for conducting cognitive interviewing studies were added as an addendum to the OMB Standards and Guidelines for Statistical Surveys. Though fully established, the method has not been static. In the past 5 years, more articles and books have been published than ever before, with authors developing and refining interviewing and analytic techniques, and redefining the methodology—once understood only as a pre-test method—as a study in validation.

This presentation will discuss the current trends in cognitive interviewing methodology as well as the various ways that findings from those studies are being used across the field of survey research. As a validation method, the studies are being used to evaluate existing survey data, to investigate whether questions perform comparatively across respondent groups, as well as to determine overlap and redundancy in survey questions. The presentation will also discuss current challenges to cognitive interviewing programs, such as representation and recruitment, as well as strategies being used to overcome those challenges.

#### **Results of a Cognitive Interview Evaluation of the Revised Race Question, with Special Emphasis on the Newly Proposed Middle Eastern/North African Response Option**

Stephanie Willson (National Center for Health Statistics), Sheba Dunston (National Center for Health Statistics), Merarys Rios (U.S. Census Bureau)

Efforts are underway to modify the standardized race and ethnicity questions present on many surveys in the Federal Statistical System. Specifically, two changes are being considered. One includes combining what was two separate questions on race and ethnicity into one question. A second change is the adoption of a new response option for respondents of Middle Eastern or North African (MENA) descent. This paper summarizes findings from a cognitive interview evaluation of the newly proposed race question, with special emphasis on the new MENA category. The Collaborating Center for Questionnaire Design and Evaluation Research conducted 89 cognitive interviews with respondents of different racial and ethnic backgrounds.

The main finding was that the detailed categories and the combined race/ethnicity feature of the question caused some respondents to report their race in a way that was slightly different from what they normally do, or in a way that was slightly different from how they normally think of themselves. Respondents could report their race in different ways, and sought to provide an answer that was commensurate with the intent of this particular question as they understood it. Specifically, there were four distinct patterns that formed the foundation for how respondents decided to report their race and ethnicity. These patterns are labeled as cultural, ancestral, administrative, and social. Respondents used these different patterns alone and in combination when thinking about how to report their race.



Moreover, the patterns that they chose to frame their responses were informed not only by their personal life experiences, but also by the structure of the question, specifically, the level of detail in the categories and the introduction of atypical categories (e.g., MENA). As such, response patterns varied among racial and ethnic groups. These patterns will be discussed in detail.

### **Training Cognitive Testing Interviewers in Different Settings and Languages**

Alisú Schoua-Glusberg (Research Support Services and IMPAQ International)

Cognitive testing requires from the interviewer a special set of skills unlike those of survey interviewing. In survey interviewing we value the interviewer's ability to read questions exactly as worded and probe minimally to elicit codable answers, but beyond that, a critical skill is persuading respondents in a sample to agree to the interview. In cognitive testing, respondents are volunteers so the focus is strictly on the interview interaction.

This presentation will discuss tailoring training for cognitive interviews, in particular for studies conducted in different corners of the world, in different languages and cultures where there is no tradition of cognitive testing. Experiences discussed will range from studies involving different languages in the US, to videotraining of interviewers in Africa and Asia.

### **Survey Translation and the Place of Expert Review in the Question Evaluation Tool Kit: Development of Best Practices**

Patricia Goerman (Center for Survey Measurement, U.S. Census Bureau), Mikelyn Meyers (Center for Survey Measurement, U.S. Census Bureau), Yazmin García Trejo (Center for Survey Measurement, U.S. Census Bureau)

It is widely considered a best practice to include respondent pretesting as part of the survey translation process (Cross Cultural Survey Guidelines, 4th ed). The U.S. Census Bureau has a Pretesting Standard that delineates requirements to ensure that any data collection instrument can be administered as intended by interviewers and understood and responded to appropriately by respondents (U.S. Census Bureau, 2015). In 2003, the requirement to include pretesting of survey translations was incorporated into the Census Bureau's Pretesting Standards (U.S. Census Bureau, 2003). The U.S. Office of Management and Budget has recently released Cognitive Interview Standards and Guidelines for all U.S. federal agencies (OMB, 2016), which also discuss testing the accuracy of translations. Most of the pretesting methods addressed in the Census Bureau standards involve respondent testing. The standards include expert review but describe it as a minimal type of testing for use in the event that time and resources do not allow for respondent testing. In large survey organizations, there can be a variety of materials in need of translation, including actual survey questionnaires, respondent letters and brochures, interviewer instructions, instructional videos and training manuals. When an agency wishes to survey people who speak as many languages as possible, there will inevitably be constraints on resources. Materials may need to be prioritized in terms of whether they will actually be seen by respondents and to what extent they need pretesting. Due to resource limitations, expert review is a method often used in the absence of respondent testing. Recent discussions have involved defining and modernizing the Census Bureau's approach to methodological expert reviews in the context of translation and how such reviews can best fit into the overall translation process. In this presentation, we will discuss the design of best practices for expert review in the context of survey translation.

**Recently Resettled Refugees and their Experience with the Annual Survey of Refugees – Findings from Cognitive and In-Depth Interviews**

David Dutwin (SSRS), Hamutal Bernstein (Urban Institute), Susan Sherr (SSRS), Arina Goyle (SSRS), Robert Santos (Urban Institute), Nicole Deterding (Business Strategy Consultants), Erica Meade (U.S. Department of Health and Human Services)

Since the early 1980s, the Administration for Children and Families' Office of Refugee Resettlement has conducted the Annual Survey of Refugees to report measures of economic self-sufficiency and integration among refugees entering the U.S. in the previous five fiscal years. Beginning in 2016, HHS undertook a redesign of the survey to improve the quality and relevance of collected data. Key goals were to better reflect contemporary social science concepts of integration and to improve the interview experience to increase the reliability of data on this complex and diverse population. As part of the redesign, over 100 refugees are participating in cognitive and in-depth qualitative interviews aimed at assessing their comprehension of survey questions, the applicability of standardized survey measures, and their experiences during survey administration. This pretest was designed to examine potential differences across national and ethnic origins; time in the US; family status at resettlement; and the gender of the respondent. We also collect data on respondents' comfort with a government-sponsored survey in the current policy environment, even given the promise of confidentiality. To improve the quality and cultural competence of administration, the pretest interviews specifically explore respondents' feelings about being interviewed; preferred modes of interview; willingness to be interviewed over time; and sensitive and/or difficult topics. Pretest interviews are conducted in Arabic, Nepali, Sgaw Karen, Somali, and Kiswahili, allowing analysis of qualitative data across cultures, languages, and ethnicities that are rarely examined in US-based survey research. Results will provide insights for surveys of vulnerable populations on sensitive subjects, specifically in environments of political vulnerability.

## Concurrent Session F-2

### Conducting Randomized Experiments in Establishment Surveys

**Conducting Experiments in Establishment Surveys: Obstacles and Opportunities**

Jaki S. McCarthy (USDA/National Agricultural Statistics Service)

Diane K. Willimack (U.S. Census Bureau)

While much advancement in survey methodology rests on the results of randomized experiments conducted in surveys, the majority of these experiments have been conducted in surveys of households or individuals. The seeming paucity of experiments in establishment surveys may be the result of obstacles not faced in household surveys. This paper will provide an overview of several key differences between household and establishment surveys that present obstacles for conducting randomized experiments in this setting. For example, establishments may be selected into many survey samples, raising concerns regarding the additional burden experiments may impose. Or large or unique establishments may be treated with special handling which is incompatible with experimental treatments.

While these and other obstacles can pose significant challenges, survey organizations have several potential strategies to address them, which will be discussed in this paper. With careful planning that takes into consideration unique features of establishment surveys and the complexities of their production environment, agencies can create more opportunities for conducting randomized experiments in establishment surveys. This will result in stronger evidence-based improvements to the field of establishment survey methodology. The papers to follow in the session will provide examples that illustrate some of the issues discussed in this overview.

### **Obstacles in Planning Establishment Survey Experiments - Census of Agriculture Content Test and Agricultural Resource Management Survey**

Kathy Ott (USDA National Agricultural Statistics Service), Tyler Wilson (USDA National Agricultural Statistics Service)

Experiments in establishment surveys require consideration of factors unseen in household surveys. For example, consideration of the burden on establishments, tailored data collection, wide variety in size of sampled establishments, availability of unit level control data, and coordination across survey programs. In agricultural surveys, sample units can range from small farms to multimillion dollar operations. The largest operations may be sampled repeatedly over time and across survey programs.

This paper will discuss how these issues impacted the planning, operations, and results for two experiments, one embedded in a production survey (the Agricultural Resource Management Survey (ARMS)) and one stand-alone experiment for the 2017 Census of Agriculture (COA) Content Test.

Alternatives to data collection procedures for ARMS intended to increase response rates were tested in several split sample experiments. Because ARMS collects complex detailed information, NASS did not want to impose this burden in a standalone experiment. However, planning involved concessions such as removing large sample units or those receiving special handling in data collection, and using the large amount of auxiliary data to selectively apply procedures.

Because the COA is conducted only every 5 years, experiments designed to improve response rates or data quality were stand-alone experiments. Prior to the 2017 COA, several split sample experiments were conducted to evaluate alternative questionnaires and data collection procedures. Considerations unique to establishments were also made in planning these experiments, such as removing large, complex or unique units or those sampled in other NASS surveys in the months surrounding the experiment due to concerns about their level of burden. In addition, NASS was also able to use auxiliary data to tailor the sample.

### **Using Email to Solicit Response in an Establishment Survey**

Joshua Langeland (Bureau of Labor Statistics)

In an era of declining survey response rates and increasing costs, evaluating the effectiveness of alternative contact strategies and modes of data collection is an important goal for organizations conducting establishment surveys. Business respondents are likely to differ in how they prefer to be contacted by survey organizations and provide data, and each method is likely to have different costs.

The US Bureau of Labor Statistics Occupational Employment Statistics (OES) survey conducted an experiment in order to evaluate the effectiveness of using email for solicitation of survey response from establishments. Sampled businesses were randomized into two groups: the Test group received an email inviting the respondent to complete the survey online, while the Control group, depending on establishment size, received either a paper mail invite or a survey package, which offered both the online and mail-return options.

The analysis focused on four outcomes: response rates, response mode preference, response time, and cost. The Test units were found to respond more slowly than the Control group but achieved an equivalent overall response rate. The results also showed that the Test units had a higher propensity to respond via web and a lower cost per response than the Control group (approximately 21% lower). This presentation will discuss experimental design details, data collection procedures, analysis results and recommendations for future OES data collection experiments.

### **Experimenting with Alternative Question Designs for “Other, Specify” Product Information in Establishment Surveys**

Diane K. Willimack (U.S. Census Bureau), Kevin Linares (formerly U.S. Census Bureau), Brian Kriz (JPSM/UMd alumnus)

The 2017 Economic Census (EC), a mandatory self-administered survey conducted every 5 years by the U.S. Census Bureau, will implement the new North American Product Classification System (NAPCS), collecting detailed information about all products sold by an establishment, along with their associated revenues. Past economic censuses requested these details only for goods and services produced within an establishment’s industrial sector. In contrast, NAPCS requires full accounting for all revenue-generating products, regardless of associated industrial sector.

While the 2017 EC electronic instrument asks respondents to select from an explicit list of common products in their industry, many respondents will need to report products that are not listed, in order to satisfy NAPCS requirements. Usability testing was unable to provide clear direction for a design that effectively prompted respondents to report these details in “other, specify” write-in spaces.

To inform questionnaire design for collecting NAPCS details, Census Bureau researchers proposed field testing two alternative question designs in a randomized experiment. However, replicating detailed EC product reporting would impose undue burden on respondents, particularly when data would not be used for official statistics. Instead, cases being rotated out of a longitudinal panel for the Services Annual Survey (SAS) were identified for this research, and questionnaire versions were prepared in a manner expected to encourage “other, specify” write-in reporting for out-of-sector products not appearing in an explicit product list. This design strategy mimics essential survey conditions associated with experienced reporters common in establishment surveys, while limiting burden and without jeopardizing responses needed for producing official statistics.

This paper will describe details about the design of this experiment, including the sample selection strategy, the alternative question designs tested, the data analysis and methodology, and the results and their implications for obtaining data using “other, specify” write-in spaces.

## Concurrent Session F-3

### Extending Research through Access and Application of Federal Data

#### **Researching the Psychology Workforce using Federal Statistics**

Luona Lin (American Psychological Association), Karen Stamm (American Psychological Association), and Peggy Christidis (American Psychological Association)

As straightforward as it seems, the question “how many psychologists are there in the U.S.?” is difficult to answer. Using the American Psychological Association’s definition of a psychologist as an individual with a doctoral degree in psychology, we explore how several federal data sources can be used to examine the size and characteristics of the psychology workforce. The psychology workforce includes individuals with doctoral degrees in psychology, but is represented through different types of doctorates, work settings, and work activities and occupations.

Various federal data sources capture different portions of the psychology workforce and have different strengths and weaknesses. Occupation, field and level of degree, or both can be used to define the psychology workforce. For example, in the 2016 Occupational Employment Statistics, the occupations of clinical/counseling/school psychologists, industrial/organizational psychologists, all other psychologists, and psychology postsecondary teachers include an estimated 160,000 individuals. Yet there is no distinction between level of degree; these estimates include master’s level individuals. The American Community Survey includes occupation and level of degree but does not include field of degree above bachelor’s degrees, and yields an estimate of approximately 188,800 individuals in the workforce coded in the psychologist occupation, of which 96,000 held doctorate or professional degrees. The National Survey of College Graduates (NSCG) includes all degree types as well as field of degree that enables us to estimate individual with psychology doctorates. In the 2013 NSCG, about 220,000 individuals held doctorate/professional degrees in psychology, of which 173,000 were in the workforce.

Despite the differences across various data sets, similarities were found in the demographic and employment characteristics of the psychology workforce using different data sets. Caveats in estimation can usually be explained by the specific portion of the workforce the data set captures. Through comparing the demographic and employment characteristics from the 2013 SDR, 2013 and 2015 NSCG, and 2015 ACS, we gain a more comprehensive and in-depth understanding of the psychology workforce.

#### **Intersectionality and Quantitative Methods for Better Serving Vulnerable Communities by Understanding the Simultaneity of Race-Gender-Class Inequalities: College Graduation and Income for College Graduates in New Mexico**

Nancy López PhD (University of New Mexico), Michael O’Donnell JD, MS (University of New Mexico), Carmela Roybal MA (University of New Mexico), Lucas Pedraza MA (University of New Mexico), Jeffrey Mitchell PhD (Bureau of Business and Economic Research, BBER, University of New Mexico)

How do education attainment, employment, work and wealth intersect to shape inequality and opportunity among diverse vulnerable communities in New Mexico? Do U.S.-born Hispanic, White,

Native American, Black and Asian men and women (ages 25-64) with four-year college degrees experience comparable levels of employment, earn parallel salaries and accumulate the same levels of wealth in New Mexico? What are the limitations of race-only, gender-only, ethnicity-only, class-only, and nativity-only approaches to mapping and interrupting inequality? How can we create new knowledge about intersecting inequalities in education, employment and wealth for social inclusion policy in New Mexico and beyond? How can a deliberate focus on the complex ways in which race, gender, class, ethnicity and nativity are intertwined at the micro/individual, meso/institutional and macro/structural levels lead to insights that can advance scholarship, research, teaching, practice for social inclusion policy in New Mexico and beyond? The purpose of this study is to:

(1) examine trends in intercategory inequality/social inclusion in educational attainment, employment and wealth across different groups by race-nativity gender-ethnicity.

(2) examine intracategorical variations in educational attainment, employment and wealth within a given racial/ethnic group.

Using fifteen years of U.S. Census data (2000-2014), specifically the American Community Survey (ACS), we detail both intracategorical and intercategory inequalities (McCall 2001, 2005) between and within groups by race, gender, and nativity. We find surprising gaps in social inequalities that would ordinarily remain unseen in conventional statistical analysis that do not engage intersectional knowledge projects. We argue that intersectionality can provide value added tools for understanding intersecting race-gender-class configurations of inequality in a given sociohistorical and political-economic context.

#### **Data Linking and Large Scale Assessments: Linking PIAAC and BLS Data for Labor Policy**

Katie Seely-Gant (Energetics Technology Center), Lisa M. Frehill (Energetics Technology Center)

Numeracy skills are increasingly crucial for quality, long-term employment in the 21st century workplace, while occupational gender segregation remains a key factor in accounting for the difference in pay between fully employed women and men. (American Association for University Women 2016; Hegewisch & Hartmann 2014; England and Allison 2009) This paper links data from the Program for the International Assessment of Adult Competencies (PIAAC) and the U.S. Bureau of Labor Statistics (BLS) O\*NET database to examine the relationship between numeracy skills, gendered wage gaps, and occupational gender segregation in the United States.

Representing part of a larger project, this paper explores the relationship between numeracy skills and occupational gender segregation by creating and analyzing a new dataset, constituted from the U.S. PIAAC and data from the U.S. Bureau of Labor Statistics (BLS) occupational-level data, including data elements from the O\*NET database. These merged enable us to answer the following research questions:

- To what extent are numeracy skills associated with occupational gender segregation?
- To what extent are there correlations between assessed numeracy skills (PIAAC), numeracy requirements (O\*NET), and occupational gender segregation?

**Examining the Principles of Open Government**

Jennifer Goode, Ph.D. (U.S. Census Bureau), Marie-Claude Jipguep-Akhtar, Ph.D. (Howard University)

The inclination towards an open government is not new. Sixty years ago, Parks (1957) wrote “From the standpoint of the principles of good government...there can be little question but that open government and information availability should be the general rule from which exceptions should be made only where there are substantial rights, interests, and considerations requiring secrecy or confidentiality and these are held by competent authority to overbalance the general public interest in openness and availability.” In 2009, the Open Government Directive identified transparency, participation, and collaboration as “the cornerstones of an open government.” Yet, in practice, the focus is predominantly on transparency, arguably because transparency can facilitate meaningful collaboration and participation. While collaboration and participation can facilitate transparency, the former direction makes more sense in the context of open data and the “window into the inner-workings” of government data transparency is assumed to provide. This paper takes a look at the nature of the relationships between federal agencies and stakeholders, and offers ways in which the principles of open government can be best operationalized.

## Concurrent Session F-4

### Improvements in Data Collection Methodology

**Cost Effective Mail Survey Design**

Doug Williams (Westat), Sherm Edwards (Westat), Pamela Giambo (Westat), Grace Kena (Bureau of Justice Statistics)

Policy needs for estimates at local levels have greatly increased, especially for data that are of uniform quality so that comparisons across areas and over time are valid. Currently, data collection by mail is the methodology that is best suited to achieve these goals, although mixed Internet and mail methods are continuing to become more competitive. This paper describes experiments that were designed to explore cost-effective data collection methods that could be undertaken by local communities to achieve reliable statistics.

The National Crime Victimization Survey – Companion Survey (NCVS-CS) is a prototype of a survey that the Bureau of Justice Statistics is testing that local jurisdictions could field to give reliable information on victimizations and policing issues for comparisons to other areas and over time. The field test of this study included a randomized block experimental design to test the use of incentives (\$0, \$1, and \$2) and delivery methods (FedEx, regular mail) in a mail survey conducted in the 40 largest metropolitan areas of the U.S.

This paper describes the experimental design and the results of the experiment that was conducted using more than 200,000 sampled addresses. The treatments are compared with respect to response rates, sample composition, and key outcome variables such as victimization and responses to community and policing items. In addition, we describe the relative cost for obtaining a completed interview under each of the treatments. It is well known that incentives increase response rates in mail

surveys, but this metric helps answer the question of which approach achieves the largest number of completed interviews for a fixed cost.

### **Integration of Multiple Data Sources to Inform a Responsive Design**

Peter Siegel (RTI International), Jennifer Wine (RTI International)

As more data are available from administrative and commercial records, as well as surveys, it is important to explore how these data can help improve the quality of the resulting data. Administrative data can be available at different times during a data collection and, in some studies, it may be difficult to link external data to sample members. We discuss the integration of data from multiple data sources to define a unit respondent and to inform a survey responsive design for the cross-sectional 2015-16 National Postsecondary Student Aid Study (NPSAS:16) which defined unit respondent based on data, regardless of source, rather than solely on survey response. During the latter stages of data collection, data from several sources were examined to identify survey nonrespondents to target for subsequent data collection intervention.

NPSAS:16 conducted a responsive design in two phases, which was challenging given that sample members were selected on a flow basis over several months and, therefore, started data collection at different times. Cases that did not complete the survey, but had over half of the data necessary to be a unit respondent, were given an abbreviated interview to collect the remaining necessary items. In addition, nonrespondents who had sufficient data to be a unit respondent went through a multiple imputation process to identify cases where the variance due to imputation was largest for key variables available only from the survey, not from any administrative sources. These sample members were then given a different abbreviated survey containing these key variables. The successful integration of administrative data allowed implementation of a responsive design within the constraints of the cross-sectional, protracted sample design.

### **Optimal Sample Size Allocation to Mixed Modes: A Case Study Using the Residential Energy Consumption Survey**

Peter Frechtel (RTI International), Phillip S. Kott (RTI International)

The Residential Energy Consumption Survey (RECS) is a household-level survey administered by the U.S. Energy Information Administration (EIA) used to estimate the energy-consumption behaviors of U.S. residents. Most of the survey items can be collected cheaply by mail or web. However, history shows that one set of items involving the measured square footage of the home needs to be collected in person by trained interviewers.

One of the goals of the RECS is to obtain accurate values for measured household square footage. To accomplish this in such a way that would limit data collection costs, it makes sense to allocate part of the sample to web/mail and part of the sample to face-to-face. The optimal split depends mainly on (1) how expensive a completed face-to-face survey is compared to a completed web/mail case, and (2) how well measured square footage can be modeled given the items that can be collected cheaply via mail or web.

In this study, we present a method for finding the optimal allocation of the sample size to the two modes given the unit costs of each mode and the fixed total cost of the survey, and we test the method



using public-use data from the 2005 RECS. Given our assumptions for the unit costs, preliminary results suggest allocating between 20% and 25% of the total sample to the face-to-face mode. It helps in this application that some of the easy-to-collect items are good predictors of the total measured square footage of a residence, such as the respondent's personal estimate of its square footage, its total number of rooms, and its number of windows. For a single-family home, other good predictors are the residence's number of stories, and whether it has an attic, a basement, and/or a garage.

### **Design Issues for a Longitudinal Employer Health Insurance Survey to Facilitate Analysis of Policy Changes**

Steven Machlin (Agency for Healthcare Research and Quality), David Kashihara (Agency for Healthcare Research and Quality)

The Medical Expenditure Panel Survey (MEPS) Insurance Component (IC) is an annual survey of employers that produces national and state-level estimates of employer-sponsored insurance (including offered plans, costs, and number of enrollees). A longitudinal extension was implemented to this survey for three years (2013-14, 2014-15 and 2015-16) to measure detailed health insurance transition estimates of U.S. businesses, with the first iteration starting just before implementation of the Patient Protection and Affordable Care Act in 2014. This paper describes the varying policy and methodological issues considered in developing the final design for each of the three surveys. Among these considerations were the impacts of the gradual phasing in of particular changes in law as well as survey budget constraints. Transitional estimates made possible by the longitudinal design will be presented, including shifts in health insurance offerings and premiums by selected characteristics of businesses.

## Concurrent Session G-1

### Challenges and Solutions to Collecting Information on Sexual Violence

#### **Improving the Measurement of Sexual Victimization Among Children through a Redesign of the National Survey of Children's Exposure to Violence**

J. Michael Brick (Westat), Jennifer Bronson (Bureau of Justice Statistics), David Finkelhor (University of New Hampshire), Darby Steiger (Westat), Brecht Donoghue (Office of Juvenile Justice & Delinquency Prevention)

The National Survey of Children's Exposure to Violence (NatSCEV) was administered by telephone in 2008 (NatSCEV-I), 2011 (NatSCEV-II), and 2014 (NatSCEV-III). These surveys mark the first comprehensive attempt to measure children's exposure to violence in their daily lives across various settings (i.e., home, school, community) and produced national estimates on sensitive topics (direct victimization and indirect violence) from a hard-to-reach, vulnerable population (i.e., children ages 0 to 17). After three survey administrations, there is a need to assess the current content, design, and data collection methodology to determine how to improve the work and increase response rates as prior NatSCEV iterations encountered challenges in maintaining methodological rigor in light of increasing survey research costs and declining response rates. Through a cooperative agreement with the Bureau of Justice Statistics and the Office of Juvenile Justice and Delinquency Prevention, Westat has been conducting methodological review work along with the University of New Hampshire Crimes Against Children Research Center that will inform the research design and survey content of the next NatSCEV. The focus of this presentation will be on the sexual victimization data collected through the Juvenile Victimization Questionnaire (JVQ), which is the core of the NatSCEV. The presentation will discuss the challenges of collecting sexual victimization data from parent proxies (for children ages 2-9) and from children and youth (ages 10-17), data collections modes, parental consent, survey question wording, the role of incentives, and other important considerations.

#### **Key Issues When Collecting and Publishing General Population Estimates of Rape and Sexual Assault**

David Cantor (Westat), Darby Steiger (Westat), John Hartge (Westat), Shannan Catalano (Bureau of Justice Statistics)

This presentation discusses three issues associated with the collection and dissemination of data on rape and sexual assault (RSA). One is the issue of validation. This is an especially prominent when measuring RSA because of the wide ranging estimates and the multiple ways RSA incidents are conceived. Methodologists concerned with measuring RSA have focused on methods to reduce underreporting. However, a number of individuals have criticized these methods by claiming they lead to overreporting. The second set of issues are ethical concerns ranging from protecting the privacy and confidentiality of survey responses to minimizing harm that might arise from participation in the study. The third issue is the dissemination of results of studies related to RSA, given the politically charged nature of the topic. Publication of results can be met with distortions or misunderstandings that can mislead consumers of the information and affect the credibility of the results.

**Collecting Data on Rape and Sexual Assault in an Institutional Setting**

Jessica Stroop (Bureau of Justice Statistics)

The panel will discuss the challenges faced in collecting good data on Rape and Sexual Assault in the National Survey of Youth in Custody (NSYC-1, NSYC-2, and NSYC-3) and the National Inmate Surveys (NIS-1, NIS-2, NIS-3, and NIS-4). These surveys are part of Bureau of Justice Statistics' (BJS) National Prison Rape Statistics Program (NPRSP), a series designed to collect multiple measures of the incidence and prevalence of sexual assault in all correctional facilities, including prisons, jails, juvenile facilities, military and Indian country facilities, and Immigration and Customs Enforcement (ICE) facilities. Challenges to collecting these data range from the measurement strategies used, including measuring direct experience from inmates and juvenile residents, the use of behaviorally specific language to screen for explicit sexual activities as well as the measurement of coercion separate from screening for activity, and the protocol for implementing these strategies in a facility setting. The NSYC and NIS series mitigate many of these challenges through the use of an Audio Computer Assisted Self-Interview (ACASI) mode of administration, providing the inmates and residents the opportunity to respond while protecting their confidentiality, and through the use of an Alternative Survey in the design for each collection. The ACASI mitigates literacy issues and protects the interviewers from having to enact mandatory reporting protocols in the juvenile setting, while the randomized questions from the Alternative Survey further mask the anonymity of inmate responses. These collections were developed in response to the Prison Rape Elimination Act of 2003 (PREA; Public Law 108-79) that requires BJS to carry out a comprehensive statistical review and analysis of the incidence and effects of prison rape for each calendar year and provide a list of institutions in the sample, separated into each category and ranked according to the incidence of prison rape in each institution.

## Concurrent Session G-2

### Innovations in Measuring and Reducing Respondent Burden

**Assessing Respondents' Perceptions of Burden in the American Community Survey**

Jessica Holzberg (U.S. Census Bureau), Jonathan Katz (U.S. Census Bureau), Gerson Morales (U.S. Census Bureau), Mary Davis (U.S. Census Bureau)

The American Community Survey (ACS) has been a "target of criticism for...excessive burden" and "generates a small but continuous stream of complaints to members of Congress" (National Academies of Sciences, Engineering, and Medicine, 2016). In response to these concerns, the American Community Survey Office has studied several possible methods of reducing respondent burden, including but not limited to reducing the number of interview contact attempts and leveraging alternative data sources to allow for the removal of questions from the survey.

To date, evaluations of the survey and efforts to improve the respondent experience have used objective measures, such as response rates and the time it takes a respondent to complete the survey. It is unclear if or how these changes to the ACS have affected respondents' perceptions of the survey experience from their own, subjective perspective. This work is an important step to addressing this gap.

We conducted ten focus groups with ACS respondents to learn about their perceptions of the ACS and its burden. A total of 46 respondents participated in these focus groups, and the focus groups ranged in size from three to eight participants. Five of the focus groups were conducted with people who had self-responded to the ACS through the mail or on the internet. The other five were conducted with people who had responded in a telephone or in-person interview. Focus groups were held in the Washington, DC; Chicago, IL; and Houston, TX areas.

In this paper, we describe our findings on respondents' experience with the ACS to identify potential sources of burden. We also discuss feedback on survey questions about overall perceptions of burden, as well as individual features of the ACS. We conclude with suggestions for researchers looking to measure perceptions of burden in their surveys.

### **Estimating Reporting Burden for Statistical Surveys**

Anna Hamelin (Energy Information Administration), Brian Hewitt (Energy Information Administration)

Accurately estimating burden per response on establishment surveys is a difficult task. According to the Paperwork Reduction Act regulations in 5 CFR 1320(b)(1), nine components of burden need to be evaluated for estimating reporting burden. Some of the components are easily understood while other components are no longer applicable or confusingly similar to other components. Respondents are able to provide quantifiable responses to some overarching components while other discrete sub-components are not easily measureable and result in respondent confusion. We will cover the approach outlined 5 CFR 1320(b)(1) for evaluating burden and the challenges that arise from attempting to measure burden using this approach. There will be some recommendations on how to improve estimating burden per response in an electronic data collection environment.

### **Exploring Sampling Techniques to Reduce Respondent Burden**

Yijun Wei (National Agricultural Statistics Service), Valbona Bejleri (National Agricultural Statistics Service)

In an effort to control respondent burden USDA's National Agricultural Statistics Service (NASS) is exploring sampling approaches that allow for coordination of multiple samples drawn within a year from the same population. Permanent random numbers (PRN) are often used to limit the overlap within a survey or between different surveys. At NASS both design-based and model-based approaches use PRNs to help reduce the respondent burden. Simulation studies examine the use of a coordination function with design-based approaches with regard to control of overlap among samples. Then simulations based on data from several NASS survey designs are conducted to explore model-based and model-assisted alternative approaches. Sample overlap rates and observed distribution of frequencies of selection of an individual unit measure sample interdependence and respondent burden.

### **Response Likelihood to an Establishment Survey with a Simple Questionnaire Following an Establishment Survey with a Complex Questionnaire**

Joseph Rodhouse (National Institute of Statistical Sciences), Tyler Wilson (National Agricultural Statistics Service), Heather Ridolfo (National Agricultural Statistics Service)

Much research has been conducted on respondent burden. Often, this research has focused on how respondent burden can negatively impact response likelihood and overall response rates. Among the

many factors that contribute to respondent burden, two that have received attention are burden caused by the complexity of the questionnaire and the total number of surveys respondents have received. Research has shown that increased complexity in the questionnaire results in lower likelihood to respond, and that a high frequency of survey contacts decreases response likelihood. However, recent research from Sinibaldi and Karlsson (2016) indicates that response likelihood might be increased for a survey with a simple questionnaire when a survey with a complex questionnaire precedes it – a result that may be explained in part by Cialdini's (2001) theory of reciprocal concessions. Building on Sinibaldi and Karlsson's work, this research extends the investigation of how complex questionnaires impact response likelihood to simpler questionnaires to the establishment survey realm. To do this, six data sets from the U.S. Department of Agriculture, National Agricultural Statistics Service are used: the 2013 and 2014 Agricultural Resource Management Surveys (phase III) and 2014 and 2015 March and June Agricultural Surveys. The ARMS III surveys represent the complex-questionnaire in the analysis, while the March and June Agricultural Surveys are the simple-questionnaires surveys. Overall, the results provide new support that the complexity of the prior-survey's questionnaire matters for response likelihood to subsequent surveys.

## Concurrent Session G-3

### Bayesian Methods for Surveys

#### **Bayesian Estimation Under Informative Sampling with Unattenuated Dependence**

Matthew R Williams (Substance Abuse and Mental Health Services Administration), Terrance D. Savitsky (U.S. Bureau of Labor Statistics)

Survey sample designs often include features such as unequal probabilities of selection, clustering, and stratification. These complex features are a challenge to include in Bayesian methods. Results for asymptotic consistency using survey weighted pseudo-likelihoods are available for pseudo-posteriors, however the class of designs often assume that the sample becomes more independent (joint inclusion probabilities factor) with increasing population and sample size. We can extend this class of results to include designs in which the extent of non-factorizing joint inclusion probabilities grows slowly rather than decays. The most common example is the household cluster setting in which the number of households increases but the size of the households is bounded. We show that the sampling dependence within households need not attenuate. Related sampling designs such as the systematic sample are briefly discussed. Some simulation results are provided along with an example analysis from the National Survey on Drug Use and Health.

#### **Scalable Bayes Clustering for Outlier Detection Under Informative Sampling**

Terrance Savitsky (Bureau of Labor Statistics)

Government surveys of business establishments receive a large volume of submissions where a small subset contain errors. Analysts need a fast-computing algorithm to flag this subset due to a short time window between collection and reporting. We offer a computationally-scalable optimization method based on non-parametric mixtures of hierarchical Dirichlet processes that allows discovery of multiple industry-indexed local partitions linked to a set of global cluster centers. Outliers are nominated as those clusters containing few observations. We extend an existing approach with a new "merge" step

that reduces sensitivity to hyperparameter settings. Survey data are typically acquired under an informative sampling design where the probability of inclusion depends on the surveyed response such that the distribution for the observed sample is different from the population. We extend the derivation of a penalized objective function to use a pseudo-posterior that incorporates sampling weights that "undo" the informative design. We provide a simulation study to demonstrate that our approach produces unbiased estimation for the outlying cluster under informative sampling. The method is applied for outlier nomination for the Current Employment Statistics survey conducted by the Bureau of Labor Statistics.

### **Some Exercises in Covariate Selection for a Bayesian Crop Yield Forecasting Model**

Habtamu Benecha (USDA National Agricultural Statistics Service), Nathan B. Cruze (USDA National Agricultural Statistics Service)

Since 2011, the National Agricultural Statistics Service (NASS) Research and Development Division has produced model-based forecasts of crop yield for various commodity crops. These models incorporate several NASS survey estimates as well as covariates through a linear model component. An ongoing need is to identify and select the most meaningful covariates from among a pool of potential covariates. In this paper we consider several potential techniques for identifying and assessing the contributions of selected covariates used in Bayesian hierarchical models for crop yield, including the evaluation of several candidate models based on the Deviance information criterion (DIC) and distance-based metrics comparing early-season forecasts against NASS end of season official estimates.

### **Are We Under-Estimating Food Insecurity?**

Christian Gregory (Economic Research Service, USDA)

This paper addresses measurement error in food security estimation in the United States. In particular, it looks at the likelihood of over- or under-reporting of the conditions that comprise the food security module (FSM), the data collection administered in many United States surveys to assess and monitor food insecurity. We address this question by using a generalization and extension of the measurement model currently used by the Economic Research Service for official food security statistics in the United States. In the context of the food security module, this model's parameters are only partly identified, but we learn about the likely values of these parameters by using a Bayesian framework. Our results suggest that, under conservative assumptions, there is significant under-reporting of more severe food security items, particularly those in the child module. The paper shows how this model can be used to bound food insecurity prevalence.

## Concurrent Session G-4

### Combining Different Levels of Data Sources

#### **Potential Uses of Individual-level Administrative Records Data in the 2012 Survey of Business Owners**

Adela Luque (U.S. Census Bureau, presenter), Renuka Bhaskar (U.S. Census Bureau), Sharon Ennis (U.S. Census Bureau), James Noon (U.S. Census Bureau)

In an effort to reduce respondent burden and operational costs while preserving quality, the U.S. Census Bureau is researching uses of administrative records (AR) data in surveys as well as the decennial census. While there is evidence of how individual-level AR data can help household surveys, less is known about whether and how they can benefit business surveys. Using the 2012 Survey of Business Owners (SBO), we start researching that question. In particular, we explore how AR data on race and ethnicity as well as veteran and foreign-born status may help inform current imputation methods. In addition, we look into whether coverage of anonymized unique person identifiers called Protected Identification Keys (PIKs) could be increased in the SBO. This work represents an expansion of the efforts already being undertaken by the SBO's use of race and ethnicity AR and previous census data in their 2012 survey.

Our findings indicate that the survey could greatly benefit from the use of Veteran Administration AR data as well as the Social Security Administration Numident's foreign born information. Our study also shows how PIK coverage could be increased in the SBO to further reduce costs while maintaining data quality.

#### **Linking USGS Water Use Data to Detailed Industries for Environmental Input-Output Modeling of the U.S. Food System**

Sarah Rehkamp (USDA Economic Research Service), Patrick Canning (USDA Economic Research Service), Catherine Birney (University of Texas-Austin)

Water is an important input in the U.S. food system and faces stress due to climate change, population growth, and changing dietary patterns. In this research, our goal is to link water use data from the U.S. Geological Survey (USGS) to industries in the Bureau of Economic Analysis (BEA) benchmark input-output tables. This data development is foundational to environmental input-output (EIO) modeling. The EIO framework allows us to measure how and where resources are used throughout the economy, a methodology endorsed by the United Nations and National Research Council. We further narrow our research focus from the national economy to the U.S. food system.

To create the value-added dataset, we allocate freshwater withdrawals published by USGS every five years in broad categories to detailed economic industries at the county-level. The county-level data allow us to account for geographical heterogeneity of water resources and use (i.e. crops grown on irrigated land versus dryland). The disaggregation of the USGS water data is facilitated by county-level data on industrial output by commodity (Census of Agriculture, USDA; County Business Patterns, DOC; Economic Census, DOC) and county-level irrigation by commodity group (Census of Agriculture, USDA; Farm and Ranch Irrigation Survey, USDA). These supplementary government data sources are necessary to allocate the water use data since a large portion of water is self-supplied and purchased water

coefficients do not apply. We then aggregate this county-level data up to the national level for the EIO analysis.

This linked dataset not only allows us to estimate the amount of water used throughout the U.S. food system, but is also part of Food Environment Data System, FEDS, a data product currently being developed at the Economic Research Service that includes other environmental impacts such as energy, greenhouse gas emissions, and agricultural land.

### **Incorporating OSHA Administrative Records in the Survey of Occupational Injuries and Illnesses**

Brooks Pierce (U.S. Bureau of Labor Statistics)

The Survey of Occupational Injuries and Illnesses (SOII) is the primary survey of nonfatal work-related injuries and illnesses in the U.S. The Occupational Safety and Health Administration (OSHA) has instituted new rules governing employer electronic reporting of work-related injuries. This raises the possibility that OSHA administrative data could be combined with SOII survey data to improve estimates of workplace injury risks. However, if some employers fail to report, OSHA records may offer biased information. This work describes practical challenges of various methods of combining the data. Simulations compare the efficacy of different methods under likely reporting scenarios. Linkage of actual OSHA reports to SOII frame records suggest that industry misclassification in OSHA records is a potentially important hurdle to overcome.

### **Data Linkage with an Establishment Survey**

Jennifer Sayers (National Center for Health Statistics), Scott Campbell (NORC), Clinton Thompson (National Center for Health Statistics), Geoff Jackson (National Center for Health Statistics)

The National Hospital Care Survey (NHCS), conducted by the National Center for Health Statistics (NCHS), collects health care statistics from a sample of 581 hospitals across the United States. The NHCS provides information on health and health care utilization from hospital inpatient stays and outpatient and emergency department visits, along with provider and patient characteristics, though it is currently not nationally representative. Since the NHCS collects personally identifiable information from patients seen in the sampled hospitals, it presents a unique opportunity for data linkage of this establishment survey to administrative and vital records. These types of data linkages will allow for robust estimates of indicators previously unattainable from using the survey alone.

NCHS has linked the 2014 NHCS to the National Death Index (NDI), which allows researchers to study mortality post hospital discharge, along with specific causes of death. This presentation will describe the processing of the data for linkage, outline the methodology used, and present initial findings from the linkage, including match-rates and summary statistics. The presentation will conclude with future directions for linkage of NHCS data, as well as the potential applications of the NHCS-NDI linked file.



**Evaluation of Health Care Event Reporting in a National Household Survey Using Medical Provider Data**

Jerrod Anderson (Agency for Healthcare Research and Quality), Emily Mitchell (Agency for Healthcare Research and Quality), Adam Biener (Agency for Healthcare Research and Quality)

The Medical Expenditure Panel Survey (MEPS) is a nationally representative health survey conducted annually by the Agency for Health Research and Quality (AHRQ). Respondents to the Household Component (HC) of MEPS provide detailed information on health care events in addition to socioeconomic data. For a subset of respondents, medical providers that are associated with health events reported by the household are then contacted to obtain more precise information on event details and expenditures. While the motivation for conducting this follow-back survey, called the Medical Provider Component (MPC), is primarily to collect data for improving the quality and completeness of expenditure data for household-reported events, in this analysis we integrate the MPC information to determine the quality of the event reporting data in the HC survey. Using machine learning methods and setting the MPC data as the validation data set, we build a model to identify characteristics of reporting accuracy and to predict reporting accuracy for those respondents for whom provider data was not collected. The prediction models are built to take into account the structure and data flow of the survey in order to make the models operationally relevant.

## Concurrent Session G-5

### Disclosure Limitation and Privacy Protection

**Marrying Demand for Statistical Information with Disclosure Control: The Canadian Experience in Developing an Automated Dissemination Tool in an Open-Data World**

Zixin Nie (Statistics Canada), Claude Girard (Statistics Canada)

With ever growing pressure coming from open-data related initiatives to provide the public with more access to data and statistics, Statistics Canada has been looking for ways to standardize and automate the dissemination of statistical results while staying faithful to its pledge to protect the confidentiality of respondents' data.

One of the main bottlenecks in expanding the release process is disclosure control, which has traditionally been the stronghold of a handful of local experts. By working to integrate their expertise more fully into the release process, Statistics Canada has succeeded in extending its series of generalized systems to include G-TAB; which is now the agency's official standard tool for dissemination. The tool applies consistent, rigorously tested, standardized and approved confidentiality rules for dissemination of data, and also provides a common platform for the creation of tabular outputs for publication.

In this presentation we will discuss the processes by which we create rigorous rules for standardization of confidentiality and dissemination models. We will also discuss some of the recent and ongoing challenges the agency has faced in devising an all-encompassing dissemination tool that is capable of yielding results with ascertained statistical quality and confidentiality protection.

**Disclosure Control and Random Tabular Adjustment**

Mark Stinner (Statistics Canada)

Statistical agencies are interested in publishing useful statistical data but doing so may lead to the disclosure of individuals' private data. This is a problem as it leads to a trade-off between the utility of the published data and the risk of disclosure of confidential data. Disclosure control can be seen as the use of methods to deal with this problem by assessing and controlling the risk of disclosing confidential data while also providing researchers with useful statistical data.

In this paper a disclosure control model based largely on Bayesian decision theory is described as well as a method of disclosure control called random tabular adjustment (RTA). This method controls the risk of disclosure by randomly adjusting the data and is intended to be an alternative to the common practice of suppressing cells at many statistical agencies. It fits naturally into the disclosure control model described. Comparisons with cell suppression and applications to real survey data are described.

**A Note on Multiplicative Noise Perturbation for Privacy Protection**

Xiaoyu Zhai and Tapan Nayak, Department of Statistics, George Washington University, Washington, DC, USA.

Most statistical data agencies aim to collect and publish data while protecting the privacy of individual respondents for legal reasons and maintaining public trust. Randomized response is one common method for privacy protection and multiplicative noise perturbation is one form of randomized response that is useful for positive quantitative variables. In this method, each respondent reports on the product of his/her true value and a random number generated from a given noise distribution. While this mechanism protects respondent's privacy, it also increase the variance of estimators. We assess this variance inflation for the Horvitz-Thompson estimator of a finite population total. We find that several commonly suggested privacy measures and variance inflation mover precisely in opposite directions and they depend only on the first two moments of the noise distribution. One implication of this result is that efforts to select a noise distribution to increase statistical efficiency, as some researchers have made, is fruitless if the privacy level is held constant. We also bring out certain deficiencies of existing privacy measures and argue that for noise one should not use a distribution that has point masses.

## Concurrent Session H-1

### New Uses of Combining Public and Private Data

#### **Linking a Retail Gasoline Price Survey with Commercial Data**

Maura Bardos (Energy Information Administration), Amerine Woodyard (Energy Information Administration), Jeramiah Yeksavich (Energy Information Administration)

As a part of ongoing modernization efforts, the U.S. Energy Information Administration (EIA) is conducting research on utilizing third-party sources to supplement publicly available data. EIA is uniquely situated since a number of its surveys collect information that is also compiled and sold by commercial vendors. These commercial vendors can provide almost real-time frequency of data that when linked with surveys, have the potential to reduce respondent burden and enhance data products. However, much is unknown about vendor's sector visibility, data definitions, processing, and data quality. This study examines an example of integrating of survey and commercial records for a weekly business survey.

The Motor Gasoline Price Survey (EIA-878) is a weekly mandatory survey of about 800 retail gasoline stations across the country. The data collected are used to create point-in-time estimates of gasoline prices at the national, regional, and selected state and city levels by grade and formulation, resulting in 276 published price estimates. Data collection, processing, and dissemination are completed within the same day. In summer 2017, EIA obtained two commercial sources of gasoline price data for research purposes. EIA purchased price data from a commercial vendor for about 110,000 stations. We also created a tool to obtain gas prices via a crowdsourced website.

We use geospatial analysis to match survey data at the station-level to commercial records and present descriptive statistics on linkage rates, availability of prices, and data quality for the commercial sources. Using the matched file, we compute estimates by city, state, and region over time and analyze the congruence between the EIA-878 and commercial sources. Based on this research, we provide an assessment of the extent to which commercial sources could be incorporated into the EIA-878. We conclude with a discussion of implications for future efforts to integrate survey and commercial datasets.

#### **Utility of Open Information Sources to Develop Estimates of Selected Crime and Justice Indicators: Measuring Arrest-Related Deaths**

Duren Banks (RTI International), Michael Planty (RTI International), G. Lance Couzens (RTI International)

Open source information—information publicly available, often through the web—has gained widespread attention as a potential alternative data collection technique to traditional surveys by design. Open sources are often freely available, produce little burden on the public, are often massive in size, and timely. However, these data can also be hard to manage, expensive to harvest, and have suspect data quality. How then can Federal Agencies responsibly use these sources for statistical purposes?

This paper extends this question to a study in measures of police use of force. Existing national-level measures of police-involved deaths have been hampered by systems that rely on voluntary reporting by a decentralized system of more than 18,000 independent law enforcement agencies. Specifically, the Bureau of Justice Statistics' (BJS) Arrest-Related Deaths (ARD) program was designed to be a census of all persons who died during the process of arrest. However, the voluntary nature of the system and varying data collection methodologies led to concerns about coverage and quality (e.g., Banks, Couzens and Planty, 2015). A capture-recapture assessment validated these concerns and demonstrated that the ARD program captured less than half of the estimated number of law enforcement homicides (Banks, Couzens and Planty, 2015).

In response to these data issues, in 2015 BJS redesigned the ARD program and implemented a pilot study to assess whether open source information could be used in concert with traditional police records to improve quality. This presentation first describes the types of estimates that might benefit from open source review, how search terms are selected and optimized, and techniques we used to validate our findings. Further we describe the challenges with sustaining a reliable and valid open source collection over time. We end with a general discussion of the challenges and limitations associated with relying solely on open source data.

#### **Evaluation of Vendor School and Teacher Lists for the 2015-16 National Teacher and Principal Survey**

Maura Spiegelman (National Center for Education Statistics), David Sheppard (U.S. Census Bureau), Quentin Brummet (U.S. Census Bureau)

This research analyzes the quality of commercial vendors as a potential sampling frame replacement or supplement for the National Teacher and Principal Survey (NTPS), a system of related school, principal, and teacher questionnaires sponsored by the National Center for Education Statistics (NCES) and collected by the U.S. Census Bureau. Sampled schools are asked to complete a School Questionnaire and Principal Questionnaire, as well as a Teacher Listing Form (TLF) that rosters eligible teachers and forms the sampling frame for the Teacher Questionnaire. Analysis first explores the coverage and eligibility rates from 3 different commercial vendors of school and teacher lists for the predecessor survey to the NTPS, followed by a detailed examination of the coverage, eligibility, match, and completeness rates of commercial teacher lists for the 2014-15 NTPS Pilot Test.

For the 2015-16 NTPS, teachers from schools that did not complete a TLF were sampled from commercial lists when vendor data were available. While teachers sampled from vendor lists were less likely to complete the teacher questionnaire than teachers sampled from school-reported rosters, the inclusion of teachers from vendor lists improved the overall survey response rate and will be continued in future NTPS administrations.

## Concurrent Session H-2

### Innovative Uses of R in Federal Agencies

#### **FDA's Approach to R Shiny Standardized, Interactive Tools**

Jimmy Wong (U.S. FDA)

Several offices at the U.S. Food and Drug Administration are developing and using R Shiny apps as standardized clinical review tools. Shiny innovates many statistics-related disciplines by providing programmers a framework to transform static code to interactive web applications. We identify recurrent analyses and visualizations in medical product reviews that could be streamlined and implement them as Shiny tools. FDA staff benefit from utilizing these tools, as they can efficiently produce outputs through a user interface instead of updating code to satisfy new data sets. We promote standardization in sections of regulatory review to encourage unambiguous interpretations across FDA staff. As Shiny is a new cutting-edge technology, discussions are needed to understand how it could be useful in different offices at the FDA. We initiated an R Shiny users group to promote collaboration and communication in developing Shiny apps. We created an R Shiny wiki to function as a central hub for the available R Shiny apps and users group information. Our approach helps to strengthen FDA as an efficient, leading-edge regulatory agency promoting and advancing public health.

#### **NHANES Weights: Use of R for Data Visualization, QA/QC, and Calibration**

George Zipf (National Center for Health Statistics)

The National Health and Nutrition Examination Survey (NHANES) is a nationally representative sample that collects data on the health and nutrition status of the United States civilian non-institutionalized population. Around 5,000 exams are conducted each year. This includes a blood draw, and lab analysis of certain analytes on sub-samples. Consequently, accurate weights are critical to national level estimates, particularly given sub-sampling and non-response.

NHANES weights are reviewed using a QA/QC protocol that includes visualization and use of several calibration methods to achieve the final weights. R has visualization capabilities that allow for comparison of multiple characteristics of a distribution (e.g. mean, quantiles, outliers, etc.) across multiple strata, and survey techniques that allow for easy post-stratification and other weights techniques. Once functions are created, programming is minimal, visualization is easily understood by non-statisticians, and survey weight adjustments can be easily compared.

This presentation gives an overview of the approach and provides specific examples.

#### **R: Innovating at the Bureau of Labor Statistics**

Arcenis Rojas, Bureau of Labor Statistics

Over the last few years numerous individuals at the BLS have developed innovations that improve data processing, analysis, or both using the R programming language. This innovation has led to gains in efficiency and better communication of our data products to both internal and external stakeholders. Some examples include a Shiny application that scrapes external resource and automates repetitive tasks saving a program office 80-100 hours of work per sample of data, a web scraping application that

also creates visualizations to simplify data analysis, and R packages that simplify or extend statistical analysis. While the BLS has historically focused its efforts primarily on processing data to ensure accuracy and objectivity, using R has allowed us to shift some of the focus toward improving data products by making them clearer and easier to understand while also improving processes.

## Concurrent Session H-3

### Statistical Methods and Testing

#### **The Effect of the Conservation Reserve Program on Rural Economies: Deriving a Statistical Verdict from a Null Finding**

Timothy Wojan (Economic Research Service), Jason Brown (Kansas City Federal Reserve Bank), Dayton Lambert (University of Tennessee)

The objective of this paper is to provide error probabilities for null findings, allowing applied economists to more confidently conclude when “not significant” can in fact be interpreted as “no substantive effect.” The example used to demonstrate the method is the Economic Research Service’s (ERS) 2004 Report to Congress that was charged with statistically identifying any unintended negative employment consequences of the Conservation Reserve Program (CRP) (Sullivan et al., 2004). The report failed to identify a statistically significant negative effect of the CRP on employment growth, but the authors correctly cautioned that the verdict of “no negative employment effect” was only valid if their econometric test was statistically powerful. We replicate the 2004 analysis and use new methods of statistical inference to resolve the two critical deficiencies that preclude the estimation of statistical power by economists: 1) positing a compelling effect size, and 2) providing an estimate of the variability of an unobserved alternative distribution using simulation methods. We conclude that the econometric test used in the report had very high power for detecting large negative employment effects from CRP and sufficiently high power for detecting a modest effect. Paradoxically, the unrestricted charge to search for “any effect” had very low power.

#### **State-Level Design-Based Estimates for National Surveys**

Stephen Ash (U.S. Census Bureau), Brian Shaffer (U.S. Census Bureau)

National surveys are being asked to do more with their national samples. In addition to providing national estimates, many national surveys are providing different sub-national estimates including state- and CBSA-level estimates. Small-area estimation is sometimes used as a last resort to produce sub-national estimates but many surveys are also using design-based methods in creative ways with their national sample designs to produce sub-national estimates.

This paper will give an overview of how several U.S. Census Bureau surveys including the American Housing Survey, Current Population Survey, National Crime Victimization Survey, and Fish Hunt: National Survey of Fishing, Hunting, and Wildlife-Associated Recreation use different methods within the design-based framework to produce sub-national estimates. We will review and compare the variety of sample designs and weighting methods for the aforementioned surveys. We will also discuss methods for evaluating the resultant estimates to determine whether they are a reasonable set of state-level estimates.

**Multivariate Small Area Estimation Under Informative Sampling and Nonresponse**

Michael Sverchkov (Bureau of Labor Statistics), Danny Pfeffermann (Bureau of Labor Statistics)

We consider multivariate small area estimation under informative sampling and not missing at random (NMAR) nonresponse. We define a response model that accounts for all different patterns of the observed outcomes (which values are observed and which ones are missing), and estimate the response probabilities by application of the Missing Information Principle. By this principle, we first set up the likelihood as if the missing outcomes were actually observed, and then integrate out the unobserved outcomes from this likelihood over the distribution holding for the missing data. The latter distribution is obtained from the distribution fitted to the observed data. Finally, the integrated likelihood is maximized with respect to the unknown parameters underlying the response model. See Sverchkov (2008) for application of this approach in the univariate case. Once the response probabilities are estimated, we impute the missing outcomes and then apply the approach of Pfeffermann and Sverchkov (2007) to the complete data set (observed and imputed values), to obtain the small area predictors. The performance of the proposed approach is illustrated by mean of simulated data and a real application.

**Semiparametric Panel Data Models Using Neural Networks**

Andrew Crane-Droesch (U.S. Department of Agriculture/Economic Research Service)

This paper presents an estimator for semiparametric models that uses a feed-forward neural network to fit the nonparametric component. Unlike many methodologies from the machine learning literature, this approach is suitable for longitudinal/panel data. It provides unbiased estimation of the parametric component of the model, with associated confidence intervals that have near-nominal coverage rates. Simulations demonstrate (1) efficiency, (2) that parametric estimates are unbiased, and (3) coverage properties of estimated intervals. An application section demonstrates the method by predicting county-level corn yield using daily weather data from the period 1981-2015, along with parametric time trends representing technological change. The method is shown to out-perform linear methods such as OLS and ridge/lasso, as well as random forest. The procedures described in this paper are implemented in the R package panelNNET.

**Testing Significance Tests: A Simulation with Cliff's Delta, *t*-tests, and Mann-Whitney *U***

Tyler Barnes (Department of Veterans Affairs/Wright State University), Scott C. Moore (Department of Veterans Affairs), Katerine Osatuke (Department of Veterans Affairs)

It is very common in survey research to see pairwise significance tests that assume continuous measurement on Likert-type data that are categorical. The purpose of this study is to discover if Cliff's Delta, a non-parametric test appropriate for categorical data, performs better than *t*-tests and the Mann-Whitney *U*. The desirable characteristics of Cliff's Delta being non-parametric and a measure of effect size with a significance value allow for testing both statistical and practical significance. Despite the theoretical and practical advantages of Cliff's Delta, this is not normal practice in most situations.

To control the specific characteristics of the data, we used a simulation to determine the performance of Cliff's Delta in a 5x2x6x3x3 fully-crossed design with 100 replications each. The first condition is sample size (20,50,100, 1000,10000). The second condition is a skewed or normal distribution. The third is mean differences (0-.02, .09-.11, .19-.21, .29-.31, .39-.41, .49-.51). The fourth and fifth conditions are the standard deviation of both items (.5,.75,1). An additional simulation will replace condition one with

unequal sample sizes ( $N_1 = 17$  and  $N_2 = 742$  derived from common comparisons the VA encounters in a large scale organizational survey) while maintaining the remaining conditions.

Pairwise comparisons are routine statistical procedures that can supplement decision making. The importance of choosing the most accurate pairwise comparison cannot be understated. The  $t$ -tests had the lowest error rates overall. When sample sizes were 20, Cliff's delta was generally more powerful and robust than the other statistics. In sample sizes of 50 and 100, non-parametric statistics performed better than  $t$ -tests in the skewed conditions, but not the normally distributed conditions. In the sample sizes greater than 100, the  $t$ -tests performed better than the non-parametric statistics. In the unequal sample size conditions,  $t$ -tests tended to be more powerful in some of the conditions, but the non-parametric statistics tended to be able to detect mean differences in more situations. In general, the choice of pairwise comparison should be the product of the data.

## Concurrent Session H-4

### Administrative Records for Sampling Efficiency

#### Sampling with Administrative Records in the National Survey of Children's Health

Scott Albrecht (U.S. Census Bureau), Jason Fields (U.S. Census Bureau), Keith Finlay (U.S. Census Bureau)

In 2015, the Census Bureau became the data collection agent for the National Survey of Children's Health. The survey requires national- and state-representative samples of children. The two-phase design uses a screener to identify households with children, and then uses a subsampling process to select a single reference child for topical questions.

In its first year of full production (2016), the Census Bureau used a sample frame of addresses in the Census Bureau's Master Address File (MAF). To ensure an adequate sample of completed interviews, the Bureau sought to increase sampling efficiency by augmenting the sample frame with a child-present flag built with a variety of administrative records. Tax, program participation, and other administrative records were used to link children to specific addresses in the MAF. The augmented sample frame was tested against the most recent year of American Community Survey (ACS) microdata. The two goals of the ACS audit were to identify the child-present rates in each stratum, and to understand how households with children differ in observable ways across the strata. The difference in the child-present rate helped determine the degree to which the flagged stratum was oversampled. The oversampling rate varied across states, but on average, flagged households were six times more likely to be selected than unflagged households. The flag performed as predicted; flagged households were approximately ten times more likely to report children than unflagged households (74.9% versus 7.4%).

In the second year of collection (2017), the Census Bureau will incorporate a modeling and optimization approach to further integrate administrative records into the sample frame.



**The Use of Administrative Records and the American Community Survey to Study the Characteristics of Undercounted Young Children in the 2010 Census**

Leticia Fernandez (U.S. Census Bureau), Rachel Shattuck (U.S. Census Bureau), James Noon (U.S. Census Bureau)

Children under age five are historically one of the most difficult segments of the population to enumerate in the U.S. decennial census. The persistent undercount of young children is highest among Hispanics and racial minorities. In this study, we link 2010 Census data to administrative records from government and third party data sources, such as Medicaid enrollment data and tenant rental assistance program records from the Department of Housing and Urban Development, to identify differences between children reported and not reported in the 2010 Census. In addition, we link children in administrative records to the American Community Survey to identify various characteristics of households with children under age five who may have been missed in the last census. This research contributes to what is known about the demographic, socioeconomic, and household characteristics of young children undercounted by the census. Our research also informs the potential benefits of using administrative records and surveys to supplement the U.S. Census Bureau child population enumeration efforts in future decennial censuses.

**Logistic Regression with Linked Data**

Abdelnasser Saïdi (Statistics Canada), Kenneth Chu (Statistics Canada), Abel Dasylyva (Statistics Canada), Félix Labrecque-Synnott (Statistics Canada)

With the recent increased demand for record linkage, the effect of linkage errors is typically disregarded. It is well-known that ignoring linkage errors can lead to biased estimates. This paper focuses on a framework to exploit linked data while taking into account these linkage errors. The first part applies the method of Chipperfield et al. (2011) of making inference with binary data in using logistic regression under the assumption that the only source of linkage error is incorrect links or false positives. A simulation study examined the effectiveness of the proposed method of Chipperfield et al. on 16,000 synthetic data sets representing a range of linkage error and clerical review rates. Simulation results showed that the estimators of the method of Chipperfield et al. are unbiased and exhibit smaller variance than those obtained from clerically reviewed data.

The second part discusses the way of analysing linked data in the solely presence of unlinked records or missing links (false negatives). A pseudo-likelihood approach which assigns weights to linked records is applied. "Non-link" weighting adjustment is effective if the weighting is related to the variables that influence the probability of being an unlinked record and to the variable being studied.

The weighting approach was applied to the 2001 Census–tax–mortality linkage. We linked 3.5 million respondents of the 2001 census long form questionnaire and the 2000–2001 T1 Personal Master File (tax data). The objective of the project was to estimate mortality rates by different socio-economic and population groups. The linkage involved 0.2% of incorrect links and a high percentage of unlinked records (11.4%). The adjustment methodology involved only a reweighting approach using proc WTADJUST of Sudaan followed by a calibration raking ratio method. Poisson bootstrap weights were also computed and calibrated. Validation tests showed that the approach adopted is very effective.

**A Method for Assigning Weight to Variable Matching in Record Linkage**

Salam Abdus (Agency for Healthcare Research and Quality), Steven C. Hill (Agency for Healthcare Research and Quality), Marc I. Roemer (Agency for Healthcare Research and Quality)

The goal of this research is to update the weights of variables used to link pharmacy transaction records to a household survey, the Medical Expenditure Panel Survey (MEPS). MEPS households report the names of drugs obtained, and with permission, MEPS contacts pharmacies to collect more detailed information (for example, payments, payers, manufacturer, strength). Successful linkage allows detailed analyses of prescription drug use, expenditures, health status, and other characteristics at the national level.

For sampled persons lacking their own pharmacy data, the detailed drug data are imputed from pharmacy records. Specifically, a similar person-drug pair is matched based on drug name, pharmacy names, and individual characteristics including types of health insurance, geographic region, and medical conditions (“match variables”). The goal of this research is to update the relative weights of match variables developed at the inception of MEPS in 1996.

We use conditional logit regression to determine the relative importance of the match variables in distinguishing true matches from non-matches. A data set containing all the potential matches for the same drug (the potential “donors”) is created, including pairs known to be true matches. The conditional logit models the probability of a true match based on agreement of the match variables. Resulting coefficients are converted into marginal effects, a common scale for the weights of the match variables. A marginal effect is, for example, the change in the probability of a true match when both the recipient and donor have private insurance coverage (the match variable agrees) relative to the recipient having such coverage and the donor lacking it (the match variable disagrees).

The method presented can be adapted by other researchers as an alternative to the Fellegi-Sunter or other frameworks commonly employed in record linkages for federal surveys and other data-matching activities.

**Application of Jaro-Winkler String Comparator in Enhancing Veterans Administrative Records**

Hyo Park (Veterans Affairs), Eddie Thomas (Veterans Affairs), Pheakdey Lim (Veterans Affairs)

This paper presents methodological issues involving an application of the Jaro-Winkler (JW) comparator, a computer matching technique we used to build an integrated database using administrative records from disparate sources within the Department of Veterans Affairs (VA) and the Department of Defense (DoD). The Veteran’s Social Security Number (SSN) was used as the primary unique identifier to link records across sources. However, in many cases, records linked by SSN alone resulted in matching records from different individuals. In this paper, we discuss the utilization of the JW algorithm in conditional matching of multiple data sources to improve the data quality. We present the results before and after applying the matching technique. We also share some insights gained while using the JW algorithm that can extend to other types of probabilistic matching.

## **Concurrent Session H-5**

### **Research Priorities for Enhancing Quality in the Integration of Multiple Data Sources**

Federal statistical agencies are increasingly using non-survey data either alone or in combination with survey data to produce official estimates. Measures that allow users to assess the quality of survey results are well understood and widely adopted. Quality metrics for official statistics produced from non-survey data or from combining survey and non-survey data are less well developed. In this panel session, gaps in quality metrics will be identified. The research needs for filling these gaps will be discussed, and the potential to team with non-federal agencies, such as academia, will be explored. The audience will be engaged in this discussion.

## Concurrent Session I-1

### Re-Engineering and Modernizing the 2017 Economic Census

#### **An Overview of the Improved 2017 Economic Census**

Kimberly Moore (U.S. Census Bureau), William Samples (U.S. Census Bureau)

The U.S. Census Bureau conducts an economic census every 5 years, covering reference years ending in 2 and 7, collecting detailed financial data about employment, operating expenses, and revenues from nearly 4 million business establishments in the United States and its territories. The Economic Census provides the nation with comprehensive, detailed, and authoritative data about the structure and functionality of the U.S. economy. Statistical data are produced for 18 industrial sectors at various levels of geography, providing benchmarks for numerous official statistics including the Gross Domestic Product (GDP), National Income and Product Accounts (NIPAs), and Producer Price Index (PPI); it provides factual support for legislation, policies, and business decisions. In addition, economic census data form the foundation of the Census Bureau's Business Register, which supports numerous annual, quarterly, and monthly surveys that measure the U.S. economy.

The 2017 Economic Census underwent a major re-engineering effort, modernizing procedures for nearly every stage in the survey life cycle. This paper will provide an overview of several new features, including:

- Web-based data collection instrument entered via secured online Portal
- Contact strategies featuring targeted communications and adaptive design techniques
- Response tracking system that provides new data quality metrics to aid data collection monitoring and management
- North American Product Classification System (NAPCS), along with other new inquiries that fill specific data gaps
- Parallel processing techniques that examine editing and data review procedures and integrate automation
- Revamped statistical designs and sampling strategies
- Hot deck imputation methods for missing product data
- Updated dissemination mechanisms
- Embedded research and evaluation using paradata and machine learning techniques
- Project management tools used throughout development and implementation

We will demonstrate how many of these modifications and innovations were fostered by trade-offs among the triple constraints – time, resources, and scope – while maintaining quality, setting the stage for the rest of this session.

**Executing a Multi-Year Multi-Method Electronic Data Collection Re-engineering: Experiences from 2017 Economic Census Development**

Amy E. Anderson Riemer (U.S. Census Bureau)

The 2017 Economic Census will be collected entirely online for the first time. In previous economic censuses, all respondents were offered the option of reporting via paper or electronically. Smaller companies could report on the Web starting in 2012, but larger companies were required to download a software application in order to enter and upload data.

For the 2012 Economic Census, there were over 600 questionnaire versions, tailored by industry. The software application included several features to assist business respondents with the management and collection of potentially thousands of pieces of data from larger companies. A key feature for larger businesses was the availability to upload pre-formatted spreadsheets containing their data.

In order to prepare for the change in platform for the 2017 Economic Census, a multi-year/multi-method research effort took place to identify key requirements, test prototypes, and evaluate early versions prior to implementation in the economic census. In addition to ensuring that key functionalities from the software and legacy Web platform were transferred, researchers took the opportunity to identify and evaluate improvements. Researchers and survey managers also took advantage of annual establishment-level surveys, such as the Annual Survey of Manufactures, for early adoption and evaluation of some of these new features.

This paper will discuss the pretesting research methods used to transition these various response options into a single re-engineered Web system. These methods included requirements gathering, usability testing, respondent debriefings, paradata analysis, and behavior coding. The paper will also address the methodological and practical challenges faced in creating and conducting this research, including lessons learned.

**2017 Economic Census Contact Strategy: Using Data to Make Decisions**

Susanne Johnson (U.S. Census Bureau)

With shrinking budgets and declining response trends, it is paramount for the U.S. Census Bureau to implement cost effective data collection strategies to maximize response for the 2017 Economic Census. Moreover, although respondents' uptake of electronic reporting has generally been quite favorable among the Census Bureau's economic surveys, implementing 100% Web data collection for the 2017 Economic Census contributes an additional layer of risk to ensuring adequate response rates. We developed the 2017 Economic Census contact strategies based on feedback gathered from focus groups and cognitive testing with respondents, results from controlled experiments testing different contact strategies, and lessons learned from the 2012 Economic Census and other economic surveys.

The contact strategies will include an initial mailing, due date reminder for selected firms, and systematic mail follow-up for nonresponse incorporating adaptive design techniques, supplemented by targeted telephone follow-up. Follow-up efforts for nonrespondents begin with the least expensive methods and increase intensity as time elapses past the due date. Some mailings will be targeted to establishments with particular characteristics under evidence-based adaptive design scenarios. In

addition, mail packages will feature design elements found to maintain or potentially improve response rates, while also increasing process efficiency without increasing costs.

The contact strategies research conducted over the past few years has provided invaluable information to improve our methods. This paper will summarize these data analyses and research efforts. We will also touch on some techniques that were studied and found to be inadequate for our purposes. We will demonstrate how results enabled data-driven decisions, and illustrate their impact on components of the mailing and follow-up activities, leading to a comprehensive, integrated, cost-effective contact strategy to maximize response for the 2017 Economic Census.

**From Research to Implementation of Product Estimation in the 2017 Economic Census: Hard, Harder, and Hardest**

William C. Davie Jr. (U.S. Census Bureau), Scot Dahl (U.S. Census Bureau), Katherine Jenny Thompson (U.S. Census Bureau)

The U.S. Census Bureau conducts an Economic Census every five years, producing key measures of American business and the economy. The Economic Census requests information on the revenue obtained from products sold from all large businesses and a sample of smaller businesses. Beginning in 2017, the Economic Census will use the North American Product Classification System (NAPCS) to produce economy-wide estimates of products sold. This marks a major departure from the prior collections, which explicitly linked product codes to industry, and required the development of a single imputation approach for all products that is statistically defensible and operationally feasible.

A research team was assembled to recommend a unified method for treating missing product data. The team's evaluative approach relied on simulation, using empirical data from a purposively selected, small subset of industries as the basis for the study. The research was complicated by the nature of product data, which are characterized by poor item response rates, few available predictors, additivity-within-establishment requirements, many rarely reported products in an industry, and sampling effects. To avoid confounding treatment effects with respondent size effects, the study restricted the analysis variables to a limited set of products within each studied industry. This greatly simplified the evaluations, but left potential implementation challenges uninvestigated. This paper describes the recommended missing data treatment methods and how these methods are being implemented into the 2017 Economic Census production system. Examples are provided to illustrate implementation issues and the modifications and enhancements needed to fully implement the research-based recommendations.

## Concurrent Session I-2

### Record Linkage Techniques to Enhance Criminal Justice and Immigration Statistics

#### **Record Linkage of Bureau of Justice (BJS) Federal Criminal Case Processing Data**

Ryan Kling (Abt Associates), Christopher Cutler (Abt Associates), Mark Motivans (Bureau of Justice Statistics)

The purpose of this presentation is to describe the construction and use of the BJS Federal Justice Statistics Program (FJSP) dyadic linking system that joins an offender's case process across diverse administrative agency data. The links can be challenging, as personal and case identifiers provided by agencies vary in their reliability and completeness. While we will touch upon the successes and challenges encountered linking the data and plans for future improvements of the links, we will also demonstrate their usefulness for describing experiences of non-citizens throughout the criminal justice system. Our goals are to promote the use of the BJS FJSP data and its linking system for external researchers and to share our experiences with others that have similar issues linking records across diverse data.

#### **Record Linkage Application in DHS USCIS Person Centric System**

Damian Kostiuk (U.S. Citizenship and Immigration Services)

The USCIS' person centric project goals for their Verification program are to create an identity management system to specifically address data quality with integration of multiple sources and an optimized search functionality. The system was built around "golden records" (i.e. augmented records to fill null attributes) for all persons associated with immigration based benefits petitions and applications. Our team applied a "gatekeeper" model originally comprised of logistic regression coupled with layers of other techniques such as random forest machine learning method to assess record linkage between numerous immigration systems, and result in a single golden record where we create complete biographic and biometric identities. We explore other methods, and are preparing for a new methodology capable of handling live ingestion.

In terms of search functionality, we developed data retrieval with an equally robust and optimized Elastic Search (search engine) configuration with a custom confidence scoring algorithm. We demonstrated (at time of abstract submission) that manual validation processes in some scenarios could be nearly eliminated using our search engine and database.

#### **Evaluation of Data Matching in Immigration Enforcement Outcome Tool**

Hongwei Zhang, Katherine Shanahan (Department of Homeland Security, Office of Immigration Statistics)

Since its inception, Department of Homeland Security (DHS) has stored immigration-related records across multiple siloed data environments that made tracking an individual's immigration history across DHS Component a challenge. The Office of Immigration Statistics (OIS) in its effort to build an

Immigration Data Integration (IDI) system to support better evidence-based decision making on immigration related issues, developed an enforcement lifecycle tool that tracks how aliens move through the DHS enforcement system. We will discuss challenges that arise from record matching in developing this tool as well as lessons learned in a follow-up study that examines the accuracy of the matching algorithm. OIS strives to keep improving upon the identity resolution as progresses are made on the IDI environment and encourages fellow researchers to provide valuable feedback.

**Data Matching Practice and Application in DHS CBP Arrival and Departure Information System (ADIS)**  
Michael Gorman (U.S. Customs and Border Protection)

The Department of Homeland Security (DHS) U.S. Customs and Border Protection (CBP) Arrival and Departure Information System (ADIS) contains biographic information, biometric indicators, and encounter data consolidated from various systems from DHS and the Department of State (DOS). ADIS facilitates the identification and investigation of individuals who may have violated their admission status by remaining in the United States beyond their authorized terms of entry. ADIS consolidates entry, exit and immigration status update related information into a consolidated person centric record to calculate overstay status. The system matches data utilizing a series of custom and open source matching algorithms designed to facilitate exact and predictive matching capabilities. Damerau–Levenshtein distance is the open source algorithm currently integrated within ADIS for predictive “fuzzy” matching, which replaced Jaro–Winkler distance to improve upon predictive matching accuracy. The system matches information together through a series of tier based matching scenarios. However, before the system matches information, business rules and data normalization standards are applied to augment the data into the system’s logical data model and to determine if the information meets requirements for data ingestion. Once the information is cleared for processing, a series of tier based matching steps are applied to identify possible match candidates or to create a new identity in the system. The majority of the system’s matches are made through exact matches. The data fields used for matching are Last Name, First Name, Birth Date, Gender, Document Numbers, Document Country of Issuance, Biographic System Identifiers and Biometric System Identifiers. To measure accuracy for the algorithms, ADIS data holdings were analyzed by the Department of Energy Lawrence Livermore National Labs for a multi-year period. ADIS has been used to generate the FY15 and FY16 DHS Entry/Exit Overstay Report for Congress and CBP has implemented overstay traveler notifications.

## Concurrent Session I-3

### New Advances in Applications of Geospatial Technology

**Mapping Geographic and Temporal Variations in Select Mortality and Natality Outcomes with R-INLA in Small Areas**

Diba Khan (National Center for Health Statistics), Lauren M. Rossen (National Center for Health Statistics), Brady Hamilton (National Center for Health Statistics), Yulei He (National Center for Health Statistics), Holly Hedegaard (National Center for Health Statistics), Rong Wei (National Center for Health Statistics), Margaret Warner (National Center for Health Statistics)

Hierarchical Bayes models have been used in disease mapping to examine small scale geographic variation. State level geographic variation for less common causes of mortality and natality outcomes



have been reported however county level variation is rarely examined. Due to concerns about statistical reliability and confidentiality, county-level mortality and natality rates based on fewer than 20 counts are suppressed based on Division of Vital Statistics, National Center for Health Statistics (NCHS) statistical reliability criteria, precluding an examination of spatio-temporal variation in less common causes of mortality and natality outcomes at the county level using direct estimates. Existing Bayesian spatio-temporal modeling strategies can be applied via Integrated Nested Laplace Approximation (INLA) in R to a large number of rare causes of mortality and natality outcomes to enable examination of spatio-temporal variations on smaller geographic scales such as counties. This method allows examination of spatiotemporal variation across the entire U.S., even where the data are sparse. We investigated suicide mortality rates from 2005-2015 and teen birth rates from 2003-2015 as two particular applications of the Bayesian spatio-temporal modeling techniques in R-INLA to predict year and county-specific smoothed estimates. Model-based estimates were mapped to explore county level spatio-temporal variation. Estimates from one model were included in a 508-compliant data visualization posted on the NCHS Data Visualization Gallery.

### **Crossing Boundaries: A Case Study on Building Accessible Tools to Combine Public-Use Data**

Matthew Graham (U.S. Census Bureau)

The last ten years has seen an explosion of online data availability from both the U.S. Census Bureau and other Federal agencies. While some datasets can be mixed together on the basis of common coding schemes, merging datasets without such common codes may require the use of spatial operations that are unfamiliar to data users or unavailable within their computational toolboxes. In some particular cases, it may make sense for agencies providing data to offer recommendations on spatial methodologies or to offer the computation services themselves as part of a comprehensive suite of online data services.

This presentation will outline a specific service - OnTheMap for Emergency Management - that the Census Bureau has been offering to the public for several years, in which Census Bureau data products are tabulated for emergency event boundaries from other Federal agencies. In addition, the presentation will break down the key pieces of this service and describe a generalized infrastructure for delivering public-use statistical data for areas that do not conform to existing legal/statistical boundaries.

### **Evaluation of Two Different Interviewing Protocols to Test a Mobile Mapping Instrument for the June Area Survey**

Denise Abreu (National Agricultural Statistics Service), Michael Hyman (National Agricultural Statistics Service), Linda A. Lawson (National Agricultural Statistics Service), Sonia Hickman (National Agricultural Statistics Service)

The National Agricultural Statistics Service is exploring the use of an electronic mobile mapping instrument to improve data collection for its June Area Survey (JAS). The JAS is based on an area sampling frame comprised of segments of land that make up the sampling units. JAS enumerators use a paper aerial photograph to locate and interview all operators within the segment boundary. Fields are drawn off by hand on the aerial photograph and a paper questionnaire is completed. Research conducted in 2014, using a mobile mapping prototype indicated that drawing fields during the interview took longer than is operationally feasible. Testing in 2016 focused on providing enumerators with pre-

delineated fields in the mobile mapping instrument in order to reduce interview time. Two types of interviewing protocols (mock and live) were used to test the use of a mobile mapping instrument in North Carolina and Indiana. This paper provides the results of both sets of interviews and outlines the advantages of each method.

### **The Time Use Data-based Measures of the Wellbeing Effect of Community Development: An Evaluative Approach.**

Wlodzimierz Okrasa (Central Statistical Office, Poland). Dominik Rozkrut (Central Statistical Office, Poland)

The problem of the interaction between the local community development and individual wellbeing - examined from an evaluative perspective - presents a challenging task due to the necessity of elaborating both the measurement and analytical instruments. This research was motivated as much by the knowledge gaps existing in the literature concerning the functional form and methods of parameterization of this relationship, as also by the policy practitioners' demand (addressed to statisticians) for tooling devices to better allocate and use the resources being accrued to local communities.

The usefulness of the proposed methodological framework - assessing the impact of the local development on individual wellbeing through multilevel modeling accounting for spatial effects - is empirically checked using a dual measurement system. To this aim, an analytical database (ADB) was created with data from two independent sources: (i) the Local Data Bank (LDB) to build a multidimensional index of local deprivation (MILD) and to capture variations in geographically-embedded administrative units (2,478 gminas, the country's finest division); and (ii) the Time Use Survey (country-wide large representative sample, in 2013) to construct the U-index ('unpleasant' - using data collected with DRM), treated next as a measure of individual wellbeing. These two kinds of units (gminas-LDB and TUS-respondents, combined within the ADB), allow one to employ multilevel modeling and to check the 'causal' relationship between their respective measures/indexes.

Since one of the implications of the main hypothesis on the interaction between community development and individual wellbeing was the importance of 'place' and 'space' (effect of neighborhood or proximity), a special interest was put on spatial effects - geographic clusters and spatial associations (autocorrelation, dependence). We found evidence that place and space matter for this relationship, proving usefulness of the employed approach as it brings in valuable insights into the analyzed problem and facilitates dealing with it for the policy purposes.

### **Sidestepping the Box: Designing a Supplemental Poverty Indicator for School Neighborhoods**

Doug Geverdt (National Center for Education Statistics), Laura Nixon (U.S. Census Bureau)

School and neighborhood poverty indicators are a familiar feature in educational research, but the scope and specificity of available indicators is limited. As a result, researchers frequently rely on data from proxy neighborhood geographies out of operational necessity rather than analytic choice. This paper examines common constraints of neighborhood data used for educational research and proposes the development and use of school-centered neighborhood poverty estimates based on data from the U.S. Census Bureau's American Community Survey (ACS) and estimation techniques borrowed from spatial statistics. The design experiment developed neighborhood poverty estimates based on 1,793 Ohio

elementary schools. Initial results suggest that the proposed indicator may provide a useful supplement to existing school-level poverty indicators and offer additional clarity about neighborhood economic conditions where schools are located. This analysis situates schools as core institutions within a surrounding neighborhood and potentially offers a new option to help assess neighborhood effects.

## Concurrent Session I-4

### Analyzing Nonresponse – New Research

#### **Correlates of Nonresponse in the 2012 and 2014 Medical Expenditure Panel Survey**

Frances Chevarley (Agency for Healthcare Research and Quality), William Mosher (Johns Hopkins Bloomberg School of Public Health)

This paper analyzes estimates from the Medical Expenditure Panel Survey (MEPS) matched with the National Health Interview Survey (NHIS) to inform MEPS nonresponse estimates. MEPS is a nationally representative panel survey studying health care use, access, expenditures, source of payment, insurance coverage, and quality of care sponsored by the Agency for Healthcare Research and Quality. Each year a new panel begins and each panel has 5 rounds of data collection over 2 ½ years that cover a two-year period.

The MEPS sample is a subsample of the previous year's NHIS sample, conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention (CDC). It is, therefore, possible to measure response and non-response to the MEPS by characteristics of the household measured in the NHIS the year before.

The goal of this paper is to explain MEPS nonresponse rates for 2012-2014. We use demographic and other personal characteristics from NHIS to determine NHIS correlates of MEPS nonresponse rates. Data used are from linked 2011NHIS/2012MEPS and 2013NHIS/2014MEPS files along with additional paradata. We explore the hypothesis that the decline in MEPS response rates may reflect an increased emphasis in survey management on the quality of the collected data, and perhaps shorter field periods.

In past research, characteristics of being non-Hispanic Black, not covered by health insurance, being in the lowest poverty level, and living in the West were associated with the largest decreases in nonresponse rates between 2010 and 2012. We will look particularly at these characteristics to see if those decreases in nonresponse rates persisted during this period of declining response rates.

#### **Subnational Geography and the Overseas U.S. Citizens Population**

Carl Turner (Fors Marsh Group), Colin MacFarlane (Fors Marsh Group), Krysha Gregorowicz (Fors Marsh Group), Jonathan Mendelson (Fors Marsh Group)

In 2017, the Federal Voting Assistance Program (FVAP) conducted a survey of the overseas citizen population who requested an absentee ballot in the 2016 general election. These citizens live across the globe and understanding the variations in this population is important for FVAP, in their efforts to serve this critical segment of the voting population. The Overseas Citizen Population Survey (OCPs) was sent to a sample of 40,000 citizens who were registered to vote. The goal of the survey is to understand the

overseas citizen population, including how these individuals participate in elections from abroad and how they interact with their fellow overseas citizens in the electoral process.

The analysis begins by examining specific issues related to variation in subnational geography within countries and how these factors differ between developed and developing countries. Specifically, how does the subnational geography in which overseas citizens live vary within countries, including GDP per capita, infrastructure quality (e.g., roads and postal serve), distance from a major city, and distance from a major airport hub? Once these variations are understood, we ask how the variations in subnational geography affect the behavior of overseas citizens. Here we focus on several key outcome variables, such as the decision to respond to the survey and the responses to key questions, especially the respondent having reported voting. We also consider whether the geographic location where overseas citizens live affect nonresponse patterns, missing value patterns, and mode choice.

The analysis concludes with a discussion of how geographical differences affect the ability of survey researchers to contact and survey individuals living abroad.

#### **Investigation into Responses for the Occupational Requirements Survey**

Alice Yu (Bureau of Labor Statistics), Erin McNulty (Bureau of Labor Statistics)

The Occupational Requirements Survey (ORS) is an establishment survey conducted by the Bureau of Labor Statistics (BLS) for the Social Security Administration (SSA). The survey collects information on the vocational preparation and the cognitive and physical requirements of occupations in the U.S. economy, as well as the environmental conditions in which those occupations are performed. In fiscal year (FY) 2016 and 2017, the BLS completed data collection for the beginning of ORS production. This paper focuses on the process for computing and analyzing the response rates from the ORS production sample, utilizing Office of Management and Budget (OMB) approved methods and formulas to produce detail statistics –weighted and unweighted at the establishment, occupation, and item levels. The results from this process will be used to help identify important auxiliary variables for use in estimation processes to reduce potential bias due to non-response in future samples.

#### **Understanding Their Apprehension: A Look into the Federal Employee Viewpoint Survey**

Mark Gorsak (Office of Personnel Management), Taylor Lewis (Office of Personnel Management), Judah Frank (Office of Personnel Management)

Non-response exists in practically any survey and could be a key threat to survey quality. Knowing the reason for the non-response could help the survey practitioner determine if groups of non-respondents would answer differently with respect to the survey questions. Information about survey members unable to participate either by refusal or another reason from administrative data and participation data in previous enumerations may provide insights about the non-respondents.

The Federal Employee Viewpoint Survey (FEVS) is a web based survey capturing employee perspectives on workplace practices and organizational policies in Federal agencies. This paper uses data from the 2016 and 2017 Federal Employee Viewpoint Surveys to assess the potential differences between non-respondents in the 2017 FEVS. Survey administrators of the FEVS conducted an experiment in 2017 where a sample about 112,000 full- and part-time Federal employees was given the option of opting out from taking the survey. Participants who decided to opt out were asked about their level of conviction towards taking the survey and primary reason for not participating.

A comparison between the participants in the opt-out experiment focuses on their level of participation in the 2016 FEVS. A study to compare responses from 2016 among the groups of non-respondents in 2017 may shed some light on the effect of results. The hypothesis is that participants who have a steadfast conviction not to respond would have been less likely to participate in the previous survey while participants who waver would have been more likely to respond before. In conclusion, responses from 2016 among both groups of non-respondents are explored to identify any differences from the overall Federal employee population as well as demographics between the non-respondents and respondents.

## Concurrent Session I-5

### Advances in Evaluating Labor Force Statistics

#### **Creating State Specific Occupational Replacement Rates**

Alex Roubinchtein (Washington State Employment Security Department)

The Bureau of Labor Statistics (BLS) concluded that the current replacement rate method undercounts the number of openings in occupational projections. As a result, BLS created a new separations methodology. BLS used replacement and separations results for the 2012 to 2022 and 2014 to 2024 projections. For 2016 to 2026 projections, only separations rates are expected to be used.

The replacement and separations methods measure workers who leave occupations and need to be replaced by new entrants into the vacated jobs.

In Washington state we annually produce forecasts for 2, 5 and 10 years. This year we created and tested a new alternative turnover method. Our alternative method is based on Washington state wage records, making results state specific.

Our alternative rates not only measure when workers leave one occupation for another, or leave the workforce, but also measure openings created by turnover within occupations (i.e. workers stay within an occupation but transfer to different companies).

The data for the alternative rates comes from Washington state wage files. We estimate the numbers of annual transfers between industries, inside industries and in and out of wage files. Then we use occupation-to-industry staffing patterns (shares of occupations for each industry) to convert industry transfers to occupational transfers. Alternative replacement rates are calculated as the shares of total transfers, minus growth or decline, divided by estimated occupational employment for a base period.

For our state the alternative openings are more than two and a half times larger than separations openings. The alternative method measures turnover within occupations, while the separations method does not. The estimated total numbers of job openings with alternative rates are consistent with other data sources on job openings. They also reflect specifics in our state's economic structure.

We used open source R-software for all major calculations.

**Employment Transitions in Washington State – An Examination of Job Flows Among Persons****Employed** in Growing and Declining Industries, 2014-2015

Jonathan Adam Lind (Washington State Employment Security Department)

Each year many American workers experience an employment transition. Examples of employment transitions include the acquisition of a new job, or a change in labor force participation status. In this paper I use data drawn from Unemployment Insurance (UI) administrative records for the State of Washington to explore how Washingtonians transition across employers (or other types of primary activities) over a two-year period. The UI records that serve as my primary data are a particularly rich source of information insofar as they detail the employment and wage histories of approximately 98% of all workers in Washington State. In particular, I examine inflows into growing industries and outflows from declining industries. I define growing and declining industries via a systematic examination of industry-level employment totals over a period of approximately 25 years. My focus on growing and declining industries allows me to examine from which sectors rapidly growing employers have drawn new workers, as well as the destinations of workers previously employed in shrinking industries. In addition, I also explore patterns of migration into, and out of, Washington State, as well as flows into, and out of, public 2- and 4-year institutions of higher education. Finally, I examine the wage trajectories of workers who transitioned out of employment in a declining industry. I find evidence that workers who transition out of employment in declining industries, on average, earn less than their immobile counterparts. Further, I find that transitions away from employment in declining industries do not always result in increased earnings among transitioners.

**Using Administrative Records for Survey Item Replacement: An Analysis of the LEHD-NSCG Annual Earnings Ratio**

Michaela Dillon (U.S. Census Bureau)

Using administrative records in survey operations can potentially improve data accuracy and reduce respondent burden. In this study, we link administrative data on earnings and employer characteristics from the 2009 Longitudinal Employer-Household Dynamics (LEHD) dataset to the 2010 National Survey of College Graduates (NSCG) to understand the alignment of this administrative records information with respondent collected data.

The LEHD program populates its database with state unemployment insurance (UI) and Quarterly Census of Employment and Wages (QCEW) data linked to other government, administrative and Census Bureau records. This information allows the tracking of both aggregate and individual level employment, earnings, and job flows over time. Data is consistently available for most states and U.S. territories from 2000 forward. Established in the 1970s, the NSCG is a biennial survey that collects data on the college-educated population of the United States, highlighting the connection between educational attainment and subsequent labor market outcomes. Additionally, the NSCG places emphasis on the collection of information relevant to measurements of the science and engineering workforce.

Upon linking the LEHD to the NSCG by Protected Identification Key (PIK), we find that while the median percent difference in reported annual earnings is small (0.015%) only one-third (33.9%) of linked earnings values are within five percent of each other. Our analysis then focuses on estimating the relationship of various demographic, employment, and academic characteristics with the LEHD-NSCG earnings ratio. Initial results indicate the earnings differential is smaller for men and increases with age. The differential is largest for those with a bachelor's degree and smallest for those with a Ph.D. Recently

switching employers or prior retirement decreases the difference in reported earnings while full time employment raises it. Lastly, the differential is largest for those employed in the for profit private sector.

## Concurrent Session J-1

### Progress in Measuring Income Statistics

#### **Measuring Trends in the Distribution of Annual Earnings**

David Pattison (Social Security Administration)

The Social Security annual earnings data includes many observations of very low earnings, either from persons who work very sporadically or from workers who begin or end a multi-year earnings spell part way through the year. If the prevalence of these low earnings varies over time, the variations can introduce trends in measures of average earnings or the median and other percentiles of the distribution, trends that might not be representative of actual changes at middle or high earnings.

A further complication is the changing age and sex composition of the workforce. The aging of the baby boom cohort through the prime working years, the large increase in female labor force participation, and recent declines in male participation at younger ages will all have affected the earnings distribution.

Several techniques are available for adjusting the measurement of trends in the earnings to protect against these variations, including the exclusion of earnings below some threshold like the federal minimum wage or a fixed percentage of the national average wage, exclusion of earnings at the beginning or end of an earnings spell, and the age-sex adjustment of earnings. It is possible that these techniques could introduce unintended trends of their own. The minimum wage has varied over time. The national average wage will have been affected by the changing age-sex composition. The sensitivity to choice of technique needs to be checked.

This paper will explore these issues using the Social Security administrative data, which have uncensored annual earnings records back to 1978 and taxable earnings back to 1951. The various techniques will be applied, alone and in combination, and the effects on trends analyzed. The window of earnings in the data is wide enough to consider as well the effects of adjustment on trends in some measures of cohort career average earnings.

#### **Creating a Comprehensive Income Dataset**

Bruce D. Meyer (University of Chicago, AEI, NBER)

Unfortunately, to measure income we almost exclusively rely on government surveys that households are increasingly unwilling to answer or at least answer accurately. For example, in the Current Population Survey (CPS), our official U.S. data source for income and poverty, about half of those receiving pensions, cash welfare, or food stamps do not report that they are recipients. Missing so much income makes the income distribution appear worse at the bottom, understates the effects of government programs in redistributing income and suggests that more people are missed by the safety net of government programs.

Fortunately, the government already collects much of the needed information to measure income well. Income data are available from many sources including surveys, tax records, and administrative data from government programs. Each of these sources has important strengths as well as major limitations:



certain income components tend to be poorly reported in surveys; tax datasets lack important individual background information and information for many key programs such as in-kind benefits and non-taxable cash transfers, which have grown considerably in recent years; and administrative data from programs are typically only available for participants and focused on a specific program. Working at the Census Bureau, I am linking CPS, tax and program data to provide an accurate and comprehensive measure of income for the U.S. population along with associated information on program participation and tax payments and credits. This Comprehensive Income Dataset (CID) will become a critical resource for policymakers and researchers as the best source to analyze poverty, mobility, and the effects of government policies. Because the database would also include details of program and tax payments, the CID would be an ideal platform to simulate possible changes in welfare and social insurance programs and income taxes, as many government agencies currently do.

### **What Does Consumer Heterogeneity Mean for Measuring Changes in the Cost of Living?**

Robert S. Martin (Bureau of Labor Statistics)

The Bureau of Labor Statistics (BLS) aims to measure changes in the cost of living with its Consumer Price Index (CPI). This paper concerns the theoretical assumption of homothetic preferences, which underlies cost of living indexes like the Chained CPI (C-CPI-U). Homotheticity posits that the rate at which a consumer unit is willing to trade one good for another is independent of its income level, implying homogenous expenditure patterns (i.e. consumer units spend the same share of their total expenditures on food and housing regardless of their differences in income). Homotheticity is appropriately viewed as a simplification of reality, but less known about its consequences for measuring aggregate cost of living, though there is a substantial literature on related issues of price index measurement. Notably, a paper in progress by Hottman and Monarch finds evidence of non-homotheticity in demand for imported goods. In this project, I attempt to replicate the Hottman and Monarch method with a panel of household purchasing data from Nielsen HomeScan. This includes estimating a demand system that allows for nonhomotheticity, which in theory could allow researchers to assess heterogeneity in cost of living increases across different income groups.

Reference:

Hottman, Colin J and Ryan Monarch (2018). "Estimating Unequal Gains across U.S. Consumers with Supplier Trade Data". Unpublished.

## **Concurrent Session J-2**

### **Forecasting Methods**

#### **An Assessment of Crime Forecasting Models**

Bruno Gasperini (IMPAQ International), Aaron Heuser (IMPAQ International), Minh Huynh (IMPAQ International), Hautahi Kingi (IMPAQ International), James Moore (IMPAQ International), Chriz Zhang (IMPAQ International)

The efficient allocation of scarce policing resources is a key operational concern for police departments. The desire for a more rigorous approach to policing coupled with more recent computational advances have led to the rapid development of many crime forecasting techniques. The relativeness effectiveness

of these various techniques, however, tend to be compared on an ad-hoc basis. In this article, we systematically compare the forecasting success of four classes of popular models in the literature - 1) simple long-term averaging, 2) classical spatial econometrics, 3) spatially independent counting processes, and 4) spatio-temporal self-exciting point processes - using reported crime data from Portland Oregon provided by the National Institute of Justice. The data contains crime incident reports for four types of crime over a five year period from 2012-2017, which allows us to compare techniques over multiple time horizons and crime types. Using cross-validation methods, we find that the spatio-temporal self-exciting process methods marginally outperform the others over most time horizons and crime types with some notable exceptions. This paper is the first to provide systematic evidence comparing methods in the burgeoning statistical field of crime forecasting.

### **Jointly Predicting U.S. Recessions and Restaurant Downturns: Integrated Models using the ERS Food Expenditure Series**

Timothy Park (Economic Research Service), Howard Elitzak (Economic Research Service, USDA), Abigail Okrent (Economic Research Service, USDA)

The accurate prediction of economic recessions is a continuing area of interest for a wide range of agents including entrepreneurs, consumers, investors, and policy makers. Food industry analysts have also highlighted restaurant recessions as a leading indicator of recessions in overall economy, since food away from home sales are closely linked to changes in consumer spending. Food industry trade journals often note that restaurants have historically led the market lower during the 3-to-6-month periods prior to the start of the last three U.S. recessions.

Our research advances the understanding of restaurant recessions by proposing a valid definition of a restaurant recession aligned with industry views and by testing for the ability to predict these downturns. Current macroeconomic models use the Treasury term spread as a key predictor of overall recession, where the term spread is defined as the difference between the ten-year and the three-month Treasury yields. We build on these specifications using the term spread and examine the predictive power of the Food Expenditure Series data published by the U.S. Department of Agriculture's Economic Research Service (ERS). The series provides estimates of sales of food eaten away from home and food eaten at home from 1929 through 2014, and the estimates are currently being revised and extended.

The proposed econometric model is based on bivariate probit time series models which can be used to predict the joint probability of economy-wide recessions and restaurant recessions. Separate univariate models would imply that the two recession events are not related. We propose methods to evaluate relative model performance using the receiver operating characteristic (ROC) curve for out of sample forecasts. The predictive ability of the models will be evaluated at various decision horizons (three- and six-month intervals) with particular attention to the power of the food expenditure series in enhancing model performance.

### **Clarifying the Confidence Levels of PPI and CPI Forecasts in the USDA's Food Price Outlook**

David Levin (USDA – Economic Research Service)

The USDA's Economic Research Service publishes the Food Price outlook, which consists of forecasts of the annual percent change in the food series of the Producer Price Index and the Consumer Price Index.

Currently, these forecasts are expressed as a fixed one percent interval (for example, the forecasted annual percent change for fresh fruits in 2017 is 2.0% to 3.0%). However, the use of this fixed one percent interval for all PPI and CPI food series may confuse data users, as the interval does not indicate a confidence level, which may lead to users to believe that the forecasts for two different PPI/CPI food series are given with equal confidence, when this belief may not necessarily be accurate. One solution that could remedy this confusion would be including confidence intervals with the published PPI/CPI forecasts. This research explores that possibility by creating backwards-looking confidence intervals for the forecasts of each individual PPI/CPI food series using the historical volatility of each individual PPI/CPI food series, and then evaluating the accuracy of the confidence intervals relative to the realized PPI/CPI values to determine if they may be useful in clarifying the confidence levels of current Food Price Outlook forecasts.

## Concurrent Session J-3

### Ensuring Data Quality and High Response Rates on Federally-Funded Establishment Data Collections

#### **Ensuring Optimal Response Rates on Agency Data Collections Over Time**

Susan Brumbaugh (RTI International), Elizabeth Smith (RTI International), Chris Ellis (RTI International), Mary Cowhig (Bureau of Justice Statistics), Zhen Zeng (Bureau of Justice Statistics)

According to a consensus study report by the National Research Council, and—as well known by federal, academic, and private researchers alike—response rates have been steadily declining for at least the past two decades. This trend presents several challenges to federally-funded programs (e.g., cost containment), not the least of which is the validity of inferences drawn from these surveys' estimates. Further, low unit- or item-level response rates can threaten the power and comparability of estimates—especially on time-series collections. However, there are noteworthy examples of collections that have weathered the response rate storms.

The mission of the Bureau of Justice Statistics (BJS) is to “collect, analyze, publish, and disseminate information on crime, criminal offenders, victims of crime, and the operation of justice systems at all levels of government...(providing) data...critical to federal, state, and local policymakers...” Two collections supporting this mission are the Deaths in Custody Reporting Program (DCRP) and Annual Survey of Jails (ASJ). Since 2009 and 2014, respectively, RTI International has collaborated with BJS to ensure optimal unit- and item-level response rates for both collections.

The DCRP is the only national statistical collection that obtains comprehensive information about deaths in adult correctional facilities, with data being collected annually from approximately 3,000 jails nationwide. The ASJ tracks changes in the size and demographic characteristics of the jail population, capacity and crowding, flow of inmates, and more. ASJ data are collected annually from a nationally representative sample of 930 local jail jurisdictions.

Our presentation will provide an overview of the trends in response rates and strategies used to achieve and sustain high quality results for DCRP (since 2009) and ASJ (since 2014).

**Reasons for Late Response and Nonresponse in Surveys of Government Agencies**

Tim Smith (RTI International), Christian Genesky (RTI International), Danielle Kaeble (Bureau of Justice Statistics), Anthony Whyde (Bureau of Justice Statistics)

Achieving and maintaining high response rates on establishment surveys pose many challenges. Establishment staff offer many reasons for nonresponse (or late response) such as lack of data, time, or interest in the survey topic. Learning more about the reasons voiced by staff might help refine communications and follow-up protocols, incentives, and other means of fostering participation. To the extent that establishments with similar characteristics (e.g., size) share reasons for nonresponse, the data collection protocols might be anticipated and tailored by type.

Data will be presented from the Annual Probation Survey and the Law Enforcement Management and Administrative Statistics Survey – two national collections sponsored by the Bureau of Justice Statistics. When contacted during nonresponse follow-up (and during Help Desk contacts), agency staff were asked why they had not responded by the survey due date or date of contact. Survey staff recorded responses verbatim and coded them into a set of pre-defined categories. Data were examined overall and by agency characteristics to determine patterns by agency type and to identify ways to address the stated reasons through written correspondence and other forms of follow-up.

The data will help validate or reject reasons for nonresponse believed to influence participation in surveys of government agencies (e.g., law enforcement agencies, state and local probation agencies) and identify ways to adapt data collection strategies for future surveys based on known agency characteristics. In addition to changes to communications strategies, other methods of fostering participation might be suggested (e.g., incentives, data collection staff training).

**Shifting Data Quality Follow-Up Methods for a Time-Series Collection of Local and State Agencies**

Chris Ellis (RTI International), Scott Ginder (RTI International), Megan Waggy (RTI International), Mary Cowhig (Bureau of Justice Statistics)

Minimizing item nonresponse and obtaining data of high quality are key goals in the field of survey research. Since 2009, RTI has successfully conducted data quality follow-up (DQFU) on the Bureau of Justice Statistics' (BJS's) Deaths in Custody Reporting Program (DCRP). Since 2016, in conjunction with the DCRP and sponsored by BJS, RTI has also implemented the Annual Survey of Jails (ASJ). Prior to 2016, our standard practice had been to wait until data from approximately all of the 3,000 eligible agencies were received before commencing with DQFU. With the assimilation of ASJ, and based on ASJ's standard practice, the concept of more real-time (or rolling) DQFU was envisioned. We implemented a small trial of this in 2016 and moved to a full rolling DQFU approach in 2017, anticipating that contacts timed to follow the agency's submission date may result in enhanced data quality and greater efficiency.

The DCRP is the only national statistical collection that obtains comprehensive information about deaths in adult correctional facilities. Data is collected annually from approximately 3,000 jails nationwide. The ASJ tracks changes in the size and demographic characteristics of the jail population, jail capacity and crowding, the flow of inmates moving into and out of jails, and use of jail space by other correctional institutions. The data are collected annually from around 875 local jail jurisdictions, representing approximately 3,000 jails nationwide.

We will present lessons learned from transitioning this time-series collection to different methods. We also will share comparative results across multiple cycles of the DCRP-ASJ collection, examining resolution rates, item response rates, DQFU costs, and agency burden.

## Concurrent Session J-4 Innovations in Survey Design

### **Collecting Electronic Health Record Data for the National Ambulatory Medical Care Survey and the National Hospital Care Survey**

Carol DeFrances, Ph.D. (National Center for Health Statistics), Denys T. Lau, Ph.D. (National Center for Health Statistics)

This presentation examines the transition of the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Care Survey (NHCS) into the electronic health record (EHR) realm. With declining response rates, the National Center for Health Statistics has searched for ways to reduce burden on providers and hospitals sampled to participate in NAMCS and NHCS. The need to move into EHR data collection and the opportunities afforded by EHR data will be discussed. Moving to EHR data collection required interagency coordination between NCHS, the Office of the National Coordinator for Health Information Technology, and the Centers for Medicare and Medicaid. It also necessitated the development of data standards and incentives for providers and hospitals to participate. The lessons learned from the 2016 NAMCS and 2016 NHCS EHR data collections will be detailed. The presentation concludes with what NAMCS and NHCS will look like in the future.

### **Data Integration Innovations to Enhance Analytical Capacity**

Steven B. Cohen (RTI International)

The quality and content of national population-based surveys are enhanced through integrated designs that include the conduct of inter-connected surveys to businesses, establishments and health providers. Their analytical capacity is dramatically enhanced through data integration efforts that link additional medical, behavioral, environmental, socio-economic and financial content from multiple sectors. This would include connectivity to existing secondary data sources at higher levels of aggregation and via direct matches to additional health and socioeconomic measures at the individual level acquired from other sources of survey, health system, economic or administrative data. Advances in data science are also serving to facilitate the effective and efficient utilization of statistical methods in concert with big data applications to develop these enhanced analytical platforms and infrastructure.

These design features and analytic enhancements will be discussed with examples drawn from national surveys with integrated designs. The presentation will also provide insights drawn from recent data integration innovations that have served to enhance the analytic capacity and utility of several existing sentinel data resources designed to inform policy and practice. Attention is given to innovative approaches to partnership development and community engagement that facilitate data sharing and use; the analytical advancements achieved in the Medical Expenditure Panel Survey (MEPS) through data integration efforts; enhancements to the analytic capacity and utility of cancer clinical trial data hosted by Project Data Sphere, LLC ("PDS") through data linkage from multiple sources; and modeling

innovations. and modeling innovations. Attention is given to analytic enhancements achieved through data integration to clinical data hosted by PDS, a data platform that provides the research community with broad access to both de-identified patient-level data from oncology clinical trials and analytical tools to assist them in analyzing those data.

### **National Health Interview Survey 2019 Content Redesign – Analytic Implications**

Sarah Lessem (National Center for Health Statistics), Renee Gindi (National Center for Health Statistics), Stephen Blumberg (National Center for Health Statistics), Aaron Maitland (National Center for Health Statistics), Ben Zablotsky (National Center for Health Statistics), Lindsey Black (National Center for Health Statistics), Emily Zammiti (National Center for Health Statistics)

The content and structure of the National Health Interview Survey (NHIS) will be updated in a 2019 questionnaire redesign. The new NHIS will better meet the needs of data users. Major features of the redesign include a dramatically shorter questionnaire, focusing data collection only on one sample adult and one sample child rather than on all persons in a family, and rotating sets of core questions so they are included with fixed periodicity rather than annually. Additional changes that may impact analysts are the method used to collect information about the family (e.g., structure, income, employment) and the possible inclusion of new "dyad" weights to explore associations between the health of parents and children. This presentation will give an overview of the redesign process and timeline, summarize changes to the questionnaire content, and explore the impacts of the questionnaire redesign on analytic activities.

### **Measuring Program Knowledge Longitudinally: The Uses and Methodological Challenges for Federal Agencies**

Matt Messel (Social Security Administration), Mark Sarney (Social Security Administration), David Rogofsky, (Social Security Administration), Laith Alattar (Social Security Administration)

In order to deliver services effectively to the public, federal agencies must ensure that target populations have adequate knowledge about their programs. Yet, little research exists that explores how the public's knowledge of federal program develops over time. In our study, we will use a nationally representative sample from the Understanding America Study (UAS) to examine individual and population-level change in Social Security program knowledge over a two-year period, using a knowledge index developed by federal researchers and an outside researcher. The UAS is mainly funded under an NIA grant but non-federal entities (Discover Card, Boston FRB, etc.) have also sponsored surveys in the UAS. We will examine how change varies by demographic factors (including age, gender, and economic status), as well as by financial literacy. The federal government will the UAS to test program knowledge interventions, and we will discuss the implications of these interventions for measuring population-level knowledge longitudinally. Our presentation will outline methodological features of the 6000-member, address-based UAS panel, which make it a robust tool for gauging knowledge over time, as well as the results of our study. We will conclude by providing examples of how measuring knowledge could help the Social Security Administration to deliver effective services to the public—and how other federal agencies might benefit from similar efforts.

**Balancing Cross-sectional and Longitudinal Design Objectives for the Survey of Doctorate Recipients**

Wan-Ying Chang (National Center for Science Engineering Statistics), Patricia Ruggles (National Center for Science Engineering Statistics)

A key issue in designing a longitudinal panel is identification of the specific topics and potential outcome measures that the panel needs to address. When a long-running survey carries both cross-sectional and longitudinal goals, there are likely to be trade-offs between breadth of coverage and the granularity of the estimates that can be made with respect to each type of goal. The Survey of Doctorate Recipients, a biennial survey conducted since 1973, is reviewing its longitudinal goals after a recent extensive expansion of its cross-sectional estimation capability. The goal of this paper is to examine potential longitudinal outcome measures that would be consistent with an effective, efficient, and sustainable longitudinal design that also meets the survey's expanded cross-sectional objectives.

This paper will propose specific outcome measures that could be examined, based on inputs both from data users and from an advisory expert panel, and will then attempt to quantify some of these trade-offs. In particular, we will produce a range of measured outcomes for specific population subgroups and examine the impact on the estimation reliability for different aggregates of the outcomes. We will consider estimates under the current panel weighting scheme (which includes some oversampling of specific groups) and the degree of any additional oversampling that might be needed to produce useful estimates for selected subgroups. The range of outcome measures that we will examine will include both transition measures (measures of a change in state, for example, from unemployed to employed) and duration-related measures (for example, the length of time needed to achieve specific employment outcomes.)















The Federal Committee on Statistical Methodology and the Council of Professional Associations on Federal Statistics recognize and are abundantly grateful to the following organizations for sponsoring this year's FCSM Research and Policy Conference. Their commitment and support help to ensure the exchange of cutting edge statistical methods and findings among statisticians and other social scientists from the Federal government, academia, and the business community.

