Federal Committee on Statistical Methodology Research Conference Final Program and Abstract Booklet

January 10–12, 2012

Walter E. Washington Convention Center
801 Mount Vernon Place, NW
Washington, DC 20001

Sponsored by:
Agency for Healthcare Research and Quality
Bureau of Economic Analysis
Bureau of Justice Statistics
Bureau of Labor Statistics
Economic Research Service
Energy Information Administration
National Agricultural Statistics Service
National Center for Education Statistics
National Center for Health Statistics
Office of Research, Evaluation and Statistics, Social Security Administration
Statistics of Income Division, Internal Revenue Service
U.S. Census Bureau
U.S. Environmental Protection Agency

Hosted by:
Council of Professional Associations on Federal Statistics

FCSM Federal Committee on Statistical Methodology

FCSM Research Conference Planning Committee

Kevin Cecco, Co-Chair, Statistics of Income Division, IRS

Pamela McGovern, Co-Chair, U.S. Census Bureau

Wendy Barboza, National Agricultural Statistics Service

Benjamin Bridgman, Bureau of Economic Analysis

Chris Chapman, National Center for Education Statistics

Margaret Conomos, U.S. Environmental Protection Agency

Jennifer Edgar, Bureau of Labor Statistics

Thesia Garner, Bureau of Labor Statistics

Stacia Henderson Martin, U.S. Census Bureau

Kim Henry, Statistics of Income Division, IRS

Anna Holaus, U.S. Census Bureau

Tracy Hunt-White, U.S. Department of Education

David Kashihara, Agency for Healthcare and Research Quality

Michael Planty, Bureau of Justice Statistics

Mark Prell, Economic Research Service, USDA

Richard Reeves, Energy Information Administration

Wanda Rivers, U.S. Census Bureau

Elaine Russell, U.S. Census Bureau

Joy Sharp, Department of Transportation

Nancy Sonnenfeld, Centers for Medicare and Medicaid Services

Edward Spar, Council of Professional Associations on Federal Statistics

Hilary Waldron, Social Security Administration

The Federal Committee on Statistical Methodology Members

(July 2011)

Brian Harris-Kojetin, Chair, Office of Management and Budget

Nancy Bates, Secretary, U.S. Census Bureau

Jonaki Bose, Substance Abuse and Mental Health Services Administration

Lynda Carlson, National Science Foundation

Steven B. Cohen, Agency for Healthcare Research and Quality

Steven H. Cohen, National Science Foundation

John Eltinge, Bureau of Labor Statistics

Patrick Flanagan, U.S. Census Bureau

Dennis Fixler, Bureau of Economic Analysis

Barry Graubard, National Cancer Institute

Deborah Griffin, U.S. Census Bureau

William Iwig, National Agricultural Statistics Service

Arthur Kennickell, Federal Reserve Board

Jennifer Madans, National Center for Health Statistics

Rochelle Martinez, Office of Management and Budget

Renee Miller, Energy Information Administration

William Sabol, Bureau of Justice Statistics

Susan Schechter, U.S. Census Bureau

Nathaniel Schenker, National Center for Health Statistics

Rolf Schmitt, Federal Highway Administration

Marilyn McMillen Seastrom, National Center for Education Statistics

Katherine K. Wallman, Champion, Office of Management and Budget

G. David Williamson, Agency for Toxic Substances and Disease Registry

Additional Conference Support

Mae Pattison, Council of Professional Associations on Federal Statistics

Note: Papers and discussant comments will be available in April 2012 on <www.fcsm.gov>.

Federal Committee on Statistical Methodology

2012 FCSM Research Conference

The 2012 Federal Committee on Statistical Methodology (FCSM) Research Conference was initiated by the FCSM. The FCSM is an interagency committee dedicated to improving the quality of federal statistics. The committee's major goals are to:

- Communicate and disseminate information on statistical practices 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 2012 FCSM Research Conference provides a forum for experts from around the world to discuss and exchange current research and methodological topics relevant to federal government statistical programs. The conference will offer papers on a wide range of topics including small area estimation, confidentiality and disclosure, paradata and dashboards, variance estimation, imputation, weighting, sampling, multi-mode and Web data collection, economic statistics, nonresponse and measurement error, record linkage, questionnaire and survey design, cognitive research, administrative records, survey redesign, Bayesian statistical methods, and data quality.

Technical demonstrations will run concurrently on the second and third days of the conference during the first morning session. These sessions will include demonstrations on "Utilizing Technology to Improve Processes, Products, and Communication" and "Data Collection and Monitoring Technology."

Sessions feature papers and demonstrations authored by government, private sector, and academic researchers from six countries. All sessions will include an open discussion and some sessions will include a formal discussion. Papers will be made available on the FCSM Web site <www.fcsm.gov> in April 2012.

In the opening plenary session, Don Dillman from Washington State University will discuss "Who Is This Survey Designed for, the 'Agency' or the Respondent?"

Final Program

Federal Committee on Statistical Methodology Research Conference

Washington, DC – January 10-12, 2012

Tuesday (1/10)

7:30 a.m.
Registration
(Foyer near Room 151A-B)

Coffee
(Foyer near Room 151A-B)

9–10 a.m.
Welcoming Remarks and PLENARY SESSION I
(Room 151A-B)

10–10:30 a.m.
Break
(Foyer near Room 151A-B)

10:30 a.m.–12 noon
CONCURRENT SESSION
I-A: I-B: I-C: I-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

12 noon–1:15 p.m.
Open

1:15–2:45p.m.
CONCURRENT SESSION
II-A: II-B: II-C: II-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

2:45–3 p.m.
Break
(Foyer near Room 151A-B)

3–4:30 p.m.
CONCURRENT SESSION
III-A: III-B: III-C: III-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

Wednesday (1/11)

7:30 a.m.
Registration
(Foyer near Room 151A-B)

Coffee
(Foyer near Room 151A-B)

9–10:30 a.m.
CONCURRENT SESSION
IV-A: IV-B: IV-C: IV-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

10:30–10:45 a.m.
Break
(Foyer near Room 151A-B)

10:45 a.m.–12:15 p.m.
CONCURRENT SESSION
V-A: V-B: V-C: V-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

12:15–1:30 p.m.
Open

1:30–3 p.m.
CONCURRENT SESSION
VI-A: VI-B: VI-C: VI-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

3–3:15 p.m.
Break
(Foyer near Room 151A-B)

3:15–4:45 p.m.
CONCURRENT SESSION
VII-A: VII-B: VII-C:
(Room 151A) (Room 151B) (Room 152A)

Thursday (1/12)

7:30 a.m.
Registration
(Foyer near Room 151A-B)

Coffee
(Foyer near Room 151A-B)

9–10:30 a.m.
CONCURRENT SESSION
VIII-A: VIII-B: VIII-C: VIII-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

10:30–10:45 a.m.
Break
(Foyer near Room 151A-B)

10:45 a.m.–12:15 p.m.
CONCURRENT SESSION
IX-A: IX-B: IX-C: IX-D:
(Room 151A)(Room 151B)(Room 152A)(154A-B)

12:15–1:30 p.m.
Open

1:30–3 p.m.
CONCURRENT SESSION
X-A: X-B: X-C:
(Room 151A)(Room 151B)(Room 152A)

Meeting Rooms:
Concurrent Session A: 151A
Concurrent Session B: 151B
Concurrent Session C: 152A
Concurrent Session D: 154A-B

Medical Assistance available in Room 153

All rooms located at street level in the middle building. Entrance at 7th and M Street, NW, near the Mt. Vernon Square/Convention Center Metro stop is closest to the meeting rooms.

Final Program[1]

Tuesday, January 10

7:30 a.m.–4:30 p.m. **Foyer near Room 151A-B**
Registration

7:30–9 a.m. **Foyer near Room 151A-B**
Coffee

9–9:10 a.m. **Room 151A-B**
Introduction and Welcoming Remarks

9:10–10 a.m.

PLENARY SESSION

"Who is This Survey Designed for, the 'Agency' or the Respondent?"
Don Dillman (Washington State University, USA)

10–10:30 a.m. **Foyer near Room 151A-B**
Break

10:30 a.m.–12 noon **Room 151A**

CONCURRENT SESSION I-A:

INCENTIVES: A LOOK ACROSS STUDIES AND TIME

Chair: Stephen Cohen (National Science Foundation, USA)

More Money? The Impact of Larger Incentives on Response Rates in Two Phase Mail Survey
Cameron McPhee[1] (American Institutes for Research, USA)
Sarah Hastedt (National Center for Education Statistics, USA)

What Has It Gotten Us? Examining Incentives Over Time in a Cross-Sectional Study
Tracy Hunt-White (National Center for Education Statistics, USA)
Jennifer Wine (RTI International, USA)
Peter Siegel (RTI International, USA)

An Investment in Goodwill or Encouraging Delays? Examining the Effects of Incentives in a Longitudinal Study
Karen Grigorian (NORC at the University of Chicago, USA)
Lynn Milan (National Science Foundation, USA)

Discussant: Shelly Martinez (Office of Management and Budget, USA)

Session Organizer: Tracy Hunt-White (U.S. Department of Education)

[1] In the case of coauthors, the presenter is underlined.

10:30 a.m.–12 noon **Room 151B**

CONCURRENT SESSION I-B:

THE NATIONAL CRIME VICTIMIZATION SURVEY (NCVS) SMALL AREA ESTIMATION PROGRAM

Chair: Michael Planty (Bureau of Justice Statistics, USA)

Generic Area Estimations and Local Crime Patterns
Michael Planty (Bureau of Justice Statistics, USA)

Rethinking the NCVS: Subnational Goals Through Direct Estimation
Robert Fay (Westat, Inc., USA)
Jianzhu Li (Westat, Inc. USA)

Rethinking the NCVS: Small Area Estimation Approaches to Estimating Crime
Jianzhu Li (Westat, Inc., USA)
Mamadou Diallo (Westat, Inc., USA)
Robert Fay (Westat, Inc. USA)

Designing a Low(er)-Cost Complement to the National Crime Victimization Survey
J. Michael Brick (Westat, Inc., USA)
W. Sherman Edwards (Westat, Inc., USA)
Sharon Lohr (Arizona State University, USA)

Discussant: James Lynch (Bureau of Justice Statistics, USA)

Session Organizer: Michael Planty (Bureau of Justice Statistics, USA)

10:30 a.m.–12 noon **Room 152A**

CONCURRENT SESSION I-C:

NEW APPROACHES TO OLD PROBLEMS IN IMPUTATION

Chair: Daniel Kasprzyk (NORC at the University of Chicago, USA)

Census Count Imputation: Mean Squared Error and Collapsing Strategies
Andrew Keller (U.S. Census Bureau)

Implementation of Improvements to the Allocation Routine for Health Insurance Coverage in the CPS ASEC
Joanna Turner (University of Minnesota, USA)
Michel Boudreaux (University of Minnesota, USA)

Evaluations of Imputation Methods to Improve the American Community Survey Estimates of the Group Quarters Population for Small Geographies
Mark Asiala (U.S. Census Bureau)
Michael Beaghen (U.S. Census Bureau)
Alfredo Navarro (U.S. Census Bureau)
Lynn Weidman (U.S. Census Bureau)

Wage Imputation in the OES Survey: A Model-Based Approach
Jane Osburn (Bureau of Labor Statistics, USA)

Session Organizer: Charlie Hallahan (U.S. Department of Agriculture)

10:30 a.m.–12 noon **Room 154A-B**

CONCURRENT SESSION I-D:
ISSUES IN INTERNATIONAL MIGRATION AND U.S. IMMIGRATION

Chair: Sheila Thompson (National Center for Education Statistics, USA)

Estimates of International Migration for United States Natives
Christopher Dick (U.S. Census Bureau)
Eric Jensen (U.S. Census Bureau)
David Armstrong (U.S. Census Bureau)

Forecasting Immigration Trends With Bayesian Structural Time Series Models
Scott Borger (Office of Immigration Statistics, USA)
Anthony Kassekert (Office of Immigration Statistics, USA)

Imputing Legal Status of Foreign-Born Persons on Surveys: Two Approaches
Dean Judson (Decision Analytics, USA)
Sharon Long (University of Minnesota, USA)

Estimating Personal Transfers From the United States
Rachel Soloveichik (Bureau of Economic Analysis, USA)
Anne Flatness (Bureau of Economic Analysis, USA)

---

12 noon–1:15 p.m.
Open Lunch

1:15–2:45 p.m. **Room 151A**

CONCURRENT SESSION II-A:
SIMULATION MODELS OF FEDERAL SURVEYS

Chair: Lawrence Cox (National Institute of Statistical Sciences, USA)

The Case for Simulation Models of Federal Surveys
Lawrence Cox (National Institute of Statistical Sciences, USA)

Simulating NHIS Field Operations
Bor-Chung Chen (Federal Railroad Administration, USA)

WSSM: World's Simplest Survey Microsimulator
Alan Karr (National Institute of Statistical Sciences, USA)

Discussant: John Eltinge (Bureau of Labor Statistics, USA)

Session Organizer: Lawrence Cox (National Institute of Statistical Sciences, USA)

1:15–2:45 p.m. **Room 151B**

CONCURRENT SESSION II-B:
ENHANCING SURVEYS THROUGH DATA LINKAGE

Chair: Paul Beatty (National Center for Health Statistics, USA)

Retrospective Linking of ECLS-K and ECLS-B Reading Scores
Carolyn Fidelman (National Center for Education Statistics, USA)

The Role of Community Health Centers in Providing Safety-Net Access to Health Care
Sandra Decker (National Center for Health Statistics, USA)
Frederic Selck (Johns Hopkins University, USA)

Use of Synthetic Data in Testing Administrative Records Systems
K. Bradley Paxton (ADI, LLC, USA)
Thomas Hager (ADI, LLC, USA)

Discussant: Robert Parker (Consultant on Federal Statistics, USA)

Session Organizer: Nancy Sonnenfeld (Centers for Medicare and Medicaid Services, USA)

1:15–2:45 p.m. **Room 152A**

CONCURRENT SESSION II-C:
USING MULTI-MODE TECHNIQUES

Chair: Jennifer Edgar (Bureau of Labor Statistics, USA)

Return to Sender: Improving Response Rates in Two-stage Mail Surveys
Saida Mamedova (American Institutes for Research, USA)
Cameron McPhee (American Institutes for Research, USA)

Multi-mode Survey Response Process for a Complex Economic Survey
Thomas Falconer (U.S. Census Bureau)
Richard Hough (U.S. Census Bureau)

Redesigning the National Surveys of Long-Term Care Providers
Lauren Harris-Kojetin (National Center for Health Statistics, USA)

Discussant: Bill Mockovak (Bureau of Labor Statistics, USA)

Session Organizer: Jennifer Edgar (Bureau of Labor Statistics, USA)

1:15–2:45 p.m. **Room 154A-B**

CONCURRENT SESSION II-D:
ECONOMIC STATISTICS

Chair: Charlie Hallahan (U.S. Department of Agriculture)

Measuring the Economic Impact of Intellectual Property Reliant Industries
Brittany Bond (U.S. Department of Commerce)
Jocelyn Burston (U.S. Department of Commerce)
David Langdon (U.S. Department of Commerce)

Modernization of Benchmarking Economic Time Series at the U.S. Census Bureau
Irene Brown (U.S. Census Bureau)

Using a Composite Index of Financial Conditions Indicators to Predict Turning Points in the U.S. Business Cycle
Gad Levanon (The Conference Board, USA)
Jean-Claude Manini (The Conference Board, USA)
Ataman Ozyildirim (The Conference Board, USA)
Jennelyn Tanchua (The Conference Board, USA)

Early Estimates for the Retail Trade with a Mixed Model
Pieter Vlag (Statistics Netherlands)

Session Organizer: Mark Prell (Economic Research Service, USA)

2:45–3 p.m. **Foyer near Room 151A-B**
Break

3–4:30 p.m. **Room 151A**

CONCURRENT SESSION III-A:
USE OF ADMINISTRATIVE RECORDS DATA IN THE 2020 CENSUS

Chair: Shelly Martinez (Office of Management and Budget, USA)

Possible 2020 Census Designs and the Use of Administrative Records: What is the Impact on Cost and Quality?
Maryann Chapin (U.S. Census Bureau)

2010 Census Simulation: Comparing Administrative Records and Decennial Data
Amy O'Hara (U.S. Census Bureau)

Public Attitudes Toward the Use of Administrative Records in the U.S. Census: Does Question Frame Matter?
Nancy Bates (U.S. Census Bureau)
Monica Wroblewski (U.S. Census Bureau)
Joanne Pascale (U.S. Census Bureau)

Discussant: John Thompson (NORC at the University of Chicago, USA)

Session Organizer: Amy O'Hara (U.S. Census Bureau)

3–4:30 p.m. **Room 151B**

CONCURRENT SESSION III-B:
ADVANCED TECHNIQUES FOR EVALUATING AND ADJUSTING FOR UNIT NONRESPONSE ACROSS DIFFERENT MODES OF DATA COLLECTION

Chair: Jonaki Bose (Substance Abuse and Mental Health Services Administration, USA)

Comparing Weighting Methods of Adjusting for Logistic Unit Nonresponse
Phillip Kott (RTI International, USA)
Dan Liao (RTI International, USA)

A Response Propensity Based Evaluation of the Treatment of Unit Nonresponse for Selected Business Surveys
Katherine Jenny Thompson (U.S. Census Bureau)
Katrina Washington (U.S. Census Bureau)

Nonresponse and Panel Attrition in a Mobile Phone Panel Survey
Marek Fuchs (Darmstadt University of Technology, Germany)

Rich List Data and Adjusting for Nonresponse Bias
Paco Martorell (RAND, USA)
David Loughran (RAND, USA)
Jacob Klerman (Abt Associates, USA)

Session Organizer: Chris Chapman (National Center for Education Statistics, USA)

3–4:30 p.m. **Room 152A**

CONCURRENT SESSION III-C:
USE OF PARADATA AND DASHBOARDS

Chair: Lynn Langton (Bureau of Justice Statistics, USA)

Comparing Estimates and Item Nonresponse Rates of Interviewers Using Statistical Process Control Techniques
Robyn Sirkis (U.S. Census Bureau)

Paradata in the Consumer Expenditure Survey
Laura Paszkiewicz (Bureau of Labor Statistics, USA)

The Implementation of Dashboards in Governments Division Surveys
Terri Craig (U.S. Census Bureau)
Carma Hogue (U.S. Census Bureau)

Development of a SAS Dashboard to Support Administrative Data Collection Processes
Marianne Reifschneider (Energy Information Administration, USA)
Susan Harris (Energy Information Administration, USA)

Session Organizer: Michael Planty (Bureau of Justice Statistics, USA)

3–4:30 p.m. **Room 154A-B**

CONCURRENT SESSION III-D: INCOME

Chair: Thesia Garner (Bureau of Labor Statistics, USA)

Reconcile This! Exploring a Balance Edit Approach in the Consumer Expenditure Survey
Brandon Kopp (Bureau of Labor Statistics, USA)
Scott Fricker (Bureau of Labor Statistics, USA)
Nhien To (Bureau of Labor Statistics, USA)

Explaining Long-term Differences Between Census and BEA Measures of the Income of Persons
Arnold Katz (Bureau of Economic Analysis, USA)

Influential Values and Robust Estimation in the Annual Survey of Public Employment and Payroll
Joseph Barth (U.S. Census Bureau)
Yang Cheng (U.S. Census Bureau)
Mary Mulry (U.S. Census Bureau)
John Tillinghast (U.S. Census Bureau)

Getting More From Survey Income Measures: Empirically-Based Recommendations for Improving Accuracy and Efficiency
John Czajka (Mathematica Policy Research, USA)
Gabrielle Denmead (Consultant, USA)

Session Organizer: Thesia Garner (Bureau of Labor Statistics, USA)

Wednesday, January 11

7:30 a.m.-4:45 p.m. **Foyer near Room 151A-B**
Registration

7:30-9 a.m. **Foyer near Room 151A-B**
Coffee

9–10:30 a.m. **Room 151A**

CONCURRENT SESSION IV-A:
SMALL AREA ESTIMATION

Chair: Charles Day (Substance Abuse and Mental Health Services Agency, USA)

Establishing a Feasibility of Small Area Estimation From the National Ambulatory Medical Care Survey (NAMCS) Data
Vladislav Beresovsky (National Center for Health Statistics, USA)
Donald Malec (National Center for Health Statistics, USA)

Small Area Estimation for Governments Surveys
Bac Tran (U.S. Census Bureau)
Yang Cheng (U.S. Census Bureau)

Small Area Prediction for a Unit Level Lognormal Model
Emily Berg (National Agricultural Statistics Service, USA)
Hukum Chandra (Indian Agricultural Statistics Research Institute, India)
Ray Chambers (University of Wollongong, Australia)

Discussant: Jerry Maples (U.S. Census Bureau)

Session Organizer: Kim Henry (Internal Revenue Service, USA)

9–10:30 a.m. **Room 151B**

CONCURRENT SESSION IV-B:
MULTI-MODE AND WEB DATA COLLECTION

Chair: Elise Christopher (National Center for Education Statistics, USA)

Evaluating Data Collection Mode Options in the 2010 National Survey of College Graduates
John Finamore (National Science Foundation, USA)

Developing a Multi-Mode, Longitudinal Study to Understand College Student Outcomes Using Becker's Human Capital Framework
Matthew Soldner (National Center for Education Statistics, USA)
Tracy Hunt-White (National Center for Education Statistics, USA)
Jennifer Wine (RTI International, USA)
Natasha Janson (RTI International, USA)

Considering a Digital 2020 Census
Michael Thieme (U.S. Census Bureau)

Discussant: Jennifer Hunter Childs (U.S. Census Bureau)

Session Organizer: Tracy Hunt-White (U.S. Department of Education)

9–10:30 a.m. **Room 152A**

CONCURRENT SESSION IV-C:
REDUCING MEASUREMENT ERROR

Chair: Chris Anguelov (Social Security Administration, USA)

Exploring Underreporting and Respondent Records Usage in the Consumer Expenditure Survey
Neil Tseng (Bureau of Labor Statistics, USA)
Brandon Kopp (Bureau of Labor Statistics, USA)
Janel Brattland (Bureau of Labor Statistics, USA)
Jeanette Davis (Bureau of Labor Statistics, USA)
Emily Geisen (RTI International, USA)
M. Christopher Stringer (U.S. Census Bureau)

Statistical Issues in the Financial Well-Being of the Aged
Chris Anguelov (Social Security Administration, USA)
Howard Iams (Social Security Administration, USA)
Patrick Purcell (Social Security Administration, USA)

Evaluation of Gross Vacancy Rates From the Decennial Census Versus Current Surveys: Early Findings From Comparisons With the 2010 Census and the 2010 ACS 1-Year Estimates
Arthur Cresce, Jr. (U.S. Census Bureau)

Evaluating an Alternative Data Source for Editing MEPS Drug Prices
Marc Zodet (Agency for Healthcare Research and Quality, USA)
Steven Hill (Agency for Healthcare Research and Quality, USA)
Samuel Zuvekas (Agency for Healthcare Research and Quality, USA)

Session Organizer: Hilary Waldron (Social Security Administration, USA)

9–10:30 a.m. **Room 154A-B**

CONCURRENT SESSION IV-D: TECHNICAL DEMONSTRATIONS

UTILIZING TECHNOLOGY TO IMPROVE PROCESSES, PRODUCTS, AND COMMUNICATION

Chair: Jaki McCarthy (National Agricultural Statistics Service, USA)

Cropscape: The New Visualization, Querying, and Dissemination Web Portal for the NASS Cropland Data Layer
Audra Zakzeski (National Agricultural Statistics Service, USA)

Visualizing Agricultural Data via the Web
Irwin Anolik (National Agricultural Statistics Service, USA)

Testing Production Data Capture Quality
K. Bradley Paxton (ADI, LLC, USA)
Steven Spiwak (ADI, LLC, USA)
Douglass Huang (ADI, LLC, USA)
James McGarity (ADI, LLC, USA)

Statipedia: A Wiki Workspace for Federal Statistical Staff
Peter Meyer (Bureau of Labor Statistics, USA)

Session Organizer: Wendy Barboza (National Agricultural Statistics Service, USA)

10:30-10:45 a.m. **Foyer near Room 151A-B**
Break

10:45 a.m. -12:15 p.m. **Room 151A**

CONCURRENT SESSION V-A:

WEIGHTING COMPLEX SURVEY DATA

Chair: Tamara Rib (Internal Revenue Service, USA)

Comparing the Selection of One Person per Household to the Selection of All Household Members: Can Less be More?
Vincent Iannacchione (RTI International, USA)
Bonnie Shook-Sa (RTI International, USA)

Methods for Incorporating an Undersampled Cell Phone Frame When Weighting a Dual-Frame Telephone Survey
A. Elizabeth Ormson (NORC at the University of Chicago, USA)
Kennon Copeland (NORC at the University of Chicago, USA)
Kirk Wolter (NORC at the University of Chicago, USA)
Kathleen Santos (NORC at the University of Chicago, USA)
Stephen Blumberg (National Center for Health Statistics, USA)

Longitudinal Survey Weight Calibration Applied to the NSF Survey of Doctorate Recipients
Michael Larsen (The George Washington University, USA)
Siyu Qing (The George Washington University, USA)
Beilei Zhou (The George Washington University, USA)
Mary Foulkes (The George Washington University, USA)

Methods for Adjusting Survey Weights When Estimating a Total
Richard Valliant (University of Michigan, USA)
Kimberly Henry (Internal Revenue Service, USA)

Session Organizer: Kim Henry (Internal Revenue Service, USA)

10:45 a.m.–12:15 p.m. **Room 151B**

CONCURRENT SESSION V-B:

EVALUATING THE SIPP RE-ENGINEERING, COMPARING SIPP AND SIPP-EHC

Chair: Bob Belli (University of Nebraska-Lincoln, USA)

A Preliminary Evaluation of the Residence History Data in the 2010 and 2011 Field Tests of the Re-Engineered Survey of Income and Program Participation
Matthew Marlay (U.S. Census Bureau)
Peter Mateyka (U.S. Census Bureau)

Measuring School Enrollment in the 2011 Re-SIPP Field Test
Stephanie Ewert (U.S. Census Bureau)
Sarah Crissey (U.S. Census Bureau)

"I Don't Remember": Effects of Recall Period on Reported Job and Program Participation Duration
Rebecca Chenevert (U.S. Census Bureau)
Renee Ellis (U.S. Census Bureau)

An Evaluation of Field Test Data From Re-Engineered SIPP Using Administrative Records and 2008 SIPP
Graton Gathright (U.S. Census Bureau)
Martha Stinson (U.S. Census Bureau)
Lori Reeder (U.S. Census Bureau)

Session Organizer: Jason Fields (U.S. Census Bureau)

10:45 a.m–12:15 p.m. **Room 152A**

CONCURRENT SESSION V-C:
MONITORING AND USING PARADATA AND METADATA

Chair: Chris Stringer (U.S. Census Bureau)

Designing a Paradata Application in a CAPI Environment
Barbara O'Hare (U.S. Census Bureau)
Matthew Jans (U.S. Census Bureau)

Using Paradata to Monitor Survey Production, Cost, and Quality Within an Adaptive Total Design Framework
David Cunningham Hunter (RTI International, USA)
Susan Mitchell (RTI International, USA)
Lisa Carley-Baxter (RTI International, USA)

Producing Control Charts to Monitor Response Rates for Business Surveys in the Economic Directorate of the U.S. Census Bureau
Yarissa Gonzalez (U.S. Census Bureau)
Broderick Oliver (U.S. Census Bureau)

A Web-Based Approach for Combining Metadata, Search, and Data Profiling
Jeff Butler (Internal Revenue Service, USA)

Session Organizer: Jennifer Edgar (Bureau of Labor Statistics, USA)

10:45 a.m.–12:15 p.m. **Room 154A-B**

CONCURRENT SESSION V-D:
STATISTICAL METHODS

Chair: Ryan Greenaway-McGrevy (Bureau of Economic Analysis, USA)

An Exploratory Components Procedure to Uncover the Q Matrix in Cognitive Diagnostic Testing
Catherine Close (University of Minnesota, USA)
Mark Davison (University of Minnesota, USA)
Ernest Davenport, Jr. (University of Minnesota, USA)

The Multivariate Differential Effects Value-Added Model
Robert Meyer (University of Wisconsin-Madison, USA)
Emin Dokumaci (University of Wisconsin-Madison, USA)

Genuine Exact Two-Stage Methodologies for Producing Assigned Accuracy Estimators for a Gamma Mean
Kevin Tolliver (U.S. Census Bureau)

Quantification of Profile Matches
Ernest Davenport, Jr. (University of Minnesota, USA)
Mark Davison (University of Minnesota, USA)

Session Organizer: Benjamin Bridgman (Bureau of Economic Analysis, USA)

12:15-1:30 p.m.
Open Lunch

1:30–3 p.m. **Room 151A**

CONCURRENT SESSION VI-A:
VARIANCE ESTIMATION

Chair: Michael Planty (Bureau of Justice Statistics, USA)

Can Post-Imputation Calibration Mitigate Possible Bias in the Multiple Imputation Variance Estimator?
Benjamin Reist (U.S. Census Bureau)
Michael Larsen (The George Washington University, USA)

Condition Indexes and Variance Decompositions for Diagnosing Multicollinearity in Linear Model and Generalized Linear Model Analysis of Survey Data
Dan Liao (RTI International, USA)

Within PSU Sort Research to Reduce Variances for the Survey of Income and Program Participation (SIPP)
Sarah Tekansik (U.S. Census Bureau)
Stephen Mack (U.S. Census Bureau)

Weighting and Variance Estimation Under Responsive Designs and Related Forms of Sample-Driven Resource Allocation for Complex Surveys
John Eltinge (Bureau of Labor Statistics, USA)

Session Organizer: Michael Planty (Bureau of Justice Statistics, USA)

1:30–3 p.m. **Room 151B**

CONCURRENT SESSION VI-B:
RESPONSE AND NONRESPONSE IN ESTABLISHMENT SURVEYS

Chair: Manisha Sengupta (National Center for Health Statistics, USA)

Survey Data Collection Methods in Establishment Surveys
Brad Edwards (Westat, Inc., USA)
David DesRoches (Mathematica Policy Research, USA)

A Comparison of Establishment Collection Mailing Methods
Chris Ellis (RTI International, USA)
Margaret Noonan (Bureau of Justice Statistics, USA)
Tim Flanigan (Bureau of Justice Statistics, USA)
Scott Ginder (Bureau of Justice Statistics, USA)
Hope Smiley McDonald (RTI International, USA)

Does it Pay to Try Again? Using Production Metrics From the Recruiting Process on an Establishment Survey to Design Recruiting Protocols
Sara Zuckerbraun (RTI International, USA)
John Loft (RTI International, USA)
Patricia LeBaron (RTI International, USA)
Manisha Sengupta (National Center for Health Statistics, USA)

Developing Response Metrics for the Economic Census
Joanna Fane Lineback (U.S. Census Bureau)
Broderick Oliver (U.S. Census Bureau)
Diane Willimack (U.S. Census Bureau)

Session Organizer: Nancy Sonnenfeld (Centers for Medicare and Medicaid Services, USA)

1:30–3 p.m. **Room 152A**

CONCURRENT SESSION VI-C:
MEASURING, MONITORING, AND IMPROVING THE QUALITY OF SURVEY DATA

Chair: John Bushery (Retired, U.S. Census Bureau, USA)

Survey Quality Indicator Measures
Donsig Jang (Mathematica Policy Research, USA)
Flora Lan (National Science Foundation, USA)

A Quality Control Program for NASS' National Operations Center
Jeffrey Boone (National Agricultural Statistics Service, USA)
Joseph Parsons (National Agricultural Statistics Service, USA)

Quality Assurance for EPA's National Coastal Survey
Marla Smith (U.S. Environmental Protection Agency)
Sarah Lehmann (U.S. Environmental Protection Agency)
Treda Grayson (U.S. Environmental Protection Agency)

Achieving Information Quality via Continuous Quality Improvement
Shawna Waugh (Energy Information Administration, USA)

Session Organizer: Benjamin Bridgman (Bureau of Economic Analysis, USA)

1:30–3 p.m. **Room 154A-B**

CONCURRENT SESSION VI-D:
PROTECTING CONFIDENTIAL DATA AND PREVENTING DISCLOSURE

Chair: Demetra Collia (Bureau of Transportation Statistics, USA)

Applicability of Basic Separability Principles to Enhance the Operational Efficiency of Synthetic Tabular Data Generation Procedures in Multi Dimensional Table Structures
Ramesh Dandekar (Energy Information Administration, USA)

A Stochastic/Deterministic Hybrid Search Method for Solving the 3-Dimensional Cell Suppression Problem for Deeply Hierarchical-Structured Tables
Matt Fetter (National Agricultural Statistics Service, USA)

k-Anonymization May Be NP-complete, but Can It Be Practical?
David Wilson (RTI, International, USA)

Establishing a Secure Data Center With Remote Access
Jeffrey Gonder (National Renewable Energy Laboratory, USA)
Evan Burton (National Renewable Energy Laboratory, USA)
Elaine Murakami (Federal Highway Administration, USA)

Session Organizer: Joy Sharp (Department of Transportation, USA)

3–3:15 p.m. **Foyer near Room 151A-B**
Break

3:15–4:45 p.m. **Room 151A**

CONCURRENT SESSION VII-A:
LINKING DATA TO ASSESS NONRESPONSE AND MEASUREMENT

Chair: William Sabol (Bureau of Justice Statistics, USA)

Nonresponse Bias Analysis of BMI Data in the ATUS Eating & Health Module
Karen Hamrick (Economic Research Service, USA)

Estimating Mental Illness in an Ongoing National Survey
Joe Gfroerer (Substance Abuse and Mental Health Services Administration, USA)
Sarra Hedden (Substance Abuse and Mental Health Services Administration, USA)
Peggy Barker (Substance Abuse and Mental Health Services Administration, USA)

Jonaki Bose (Substance Abuse and Mental Health Services Administration, USA)
Jeremy Aldworth (RTI International, USA)

Telling Truth From Y's: Accuracy of Self-Reported Condom Use Assessed by a Semen Y-Chromosome Biomarker for Unprotected Sex
Janet Rosenbaum (University of Maryland, USA)
Johan Melendez (Johns Hopkins University, USA)
Eve Rose (Emory University, USA)
Gina Wingood (Emory University, USA)
Ralph DiClemente (Emory University, USA)
Jonathan Zenilman (Johns Hopkins University, USA)

Linking Children From the National Health and Nutrition Examination Survey to Medicaid Enrollment and Claims Data
Lisa Mirel (National Center for Health Statistics, USA)
Gloria Wheatcroft (Center for Medicare and Medicaid Services, USA)
Cordell Golden (National Center for Health Statistics, USA)
Kenneth Schoendorf (National Center for Health Statistics, USA)

Session Organizer: Michael Planty (Bureau of Justice Statistics, USA)

3:15–4:45 p.m. **Room 151B**

CONCURRENT SESSION VII-B:
QUESTIONNAIRE AND SURVEY DESIGN

Chair: David Kashihara (Agency for Healthcare Research and Quality, USA)

Measurement Strategies for Identifying Holders of Certificates and Certifications
Sharon Boivin (National Center for Education Statistics, USA)
Isaiah O'Rear (National Center for Education Statistics, USA)

Evaluation of the 2008 American Community Survey Employment Status Question Change
Alfred Gottschalck (U.S. Census Bureau)
Braedyn Kromer (U.S. Census Bureau)
David Howard (U.S. Census Bureau)
David Hedengren (U.S. Census Bureau)

National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR) Cell Phone and Debit Card Test
Elke McLaren (U.S. Census Bureau)
Aniekan Okon (U.S. Census Bureau)
Denise Pepe (U.S. Census Bureau)

The Missing Link: From Concepts to Questions in Economic Surveys
Diane Willimack (U.S. Census Bureau)
Ger Snijkers (Statistics Netherlands)

Session Organizer: David Kashihara (Agency for Healthcare Research and Quality, USA)

3:15–4:45 p.m. **Room 152A**

CONCURRENT SESSION VII-C:
CHALLENGING ISSUES FOR TODAY AND TOMORROW: INNOVATIVE APPROACHES

Chair: Barry Nussbaum (U.S. Environmental Protection Agency)

Consequences of Social Security Number Randomization
Bert Kestenbaum (Social Security Administration, USA)

Taxable Property Values: Exploring the Feasibility of a Survey
Peter Schilling (U.S. Census Bureau)
Brian Zamperini (U.S. Census Bureau)

Measurement Error in the Relationship Status of Same-Sex Couples in the 2009 American Community Survey
Matthew Virgile (U.S. Census Bureau)

Statistics on Temporal and Spatial Activity Based on Mobile Phone Network Data
Edwin de Jonge (Statistics Netherlands)
Merijn van Pelt (Statistics Netherlands)
Marko Roos (Statistics Netherlands)

Session Organizer: Margaret Conomos (U.S. Environmental Protection Agency)

7:30 a.m.-12:15 p.m. **Foyer near Room 151A-B**
Registration

7:30-9 a.m. **Foyer near Room 151A-B**
Coffee

9–10:30 a.m. **Room 151A**

CONCURRENT SESSION VIII-A:
MULTIPLE IMPUTATION

Chair: Patrick Flanagan (U.S. Census Bureau)

MI Double Feature: Multiple Imputation to Address Nonresponse and Rounding Errors in Income Questions Simultaneously
Jörg Drechsler (Institute for Employment Research, Germany)
Hans Kiesl (University of Applied Sciences, Germany)

Multiple Imputation: Limitations Born From Missing Data Among Covariates
Kenneth Steve (Bureau of Transportation Statistics, USA)
Li Leung (Bureau of Transportation Statistics, USA)

Effect of Different Imputation Methods on Factor Analyses of CAHPS Nursing Home Survey
Robert Baskin (Agency for Healthcare Research and Quality, USA)
Judy Sangl (Agency for Healthcare Research and Quality, USA)
Marc Zodet (Agency for Healthcare Research and Quality, USA)

Discussant: Katherine Jenny Thompson (U.S. Census Bureau)

Session Organizer: Mark Prell (Economic Research Service, USA)

9–10:30 a.m. **Room 151B**

CONCURRENT SESSION VIII-B:
ISSUES IN DESIGNING AND REDESIGNING ESTABLISHMENT SURVEYS

Chair: Clarice Brown (National Center for Health Statistics, USA)

Closing the Feedback Loop in Survey Design – How Public Sector Research Needs Inform Evolution of the Census of Agriculture
Luanne Lohr (Economic Research Service, USA)
Donald Buysse (National Agricultural Statistics Service, USA)

Survey Redesign Process for a Business Survey: Lessons Learned From the Business R&D and Innovation Survey
Richard Hough (U.S. Census Bureau)

Comparison of the Data Reliability of Nationally Representative Estimates From the Full (2007) and Half (2008) Samples of the National Hospital Discharge Survey–Implications for Trend Analyses
Margaret Jean Hall (National Center for Health Statistics, USA)
Maria Owings (National Center for Health Statistics, USA)

Integration of the National Hospital Discharge Survey and the National Hospital Ambulatory Medical Care Survey into the National Hospital Care Survey
Carol DeFrances (National Center for Health Statistics, USA)
David Woodwell (National Center for Health Statistics, USA)

Session Organizer: Nancy Sonnenfeld (Centers for Medicare and Medicaid Services, USA)

9–10:30 a.m. **Room 152A**

CONCURRENT SESSION VIII-C:
DISCLOSURE

Chair: Jacob Bournazian (Energy Information Administration, USA)

Generalized Domain Size Threshold for Analysis Restrictions for Remote Analysis Servers
Avinash Singh (NORC at the University of Chicago, USA)
Joshua Borton (NORC at the University of Chicago, USA)
Allan Crego (NORC at the University of Chicago, USA)

Logistic Regression With Variables Subject to Post Randomization Method
Yong Ming Jeffrey Woo (The Pennsylvania State University, USA)
Aleksandra Slavkovi (The Pennsylvania State University, USA)

Disclosure Avoidance Through Noise Infusion and Synthetic Data: The Case of the Quarterly Workforce Indicators
John Abowd (Cornell University, USA, and U.S. Census Bureau)
Lars Vilhuber (Cornell University, USA, and U.S. Census Bureau)

Kaj Gittings (Louisiana State University, USA)
Kevin McKinney (U.S. Census Bureau)
Bryce Stephens (Bates White Economic Consulting, USA)
Simon Woodcock (Simon Fraser University, USA)

Discussant: William Winkler (U.S. Census Bureau)

Session Organizer: Richard Reeves (Energy Information Administration, USA)

9–10:30 a.m. Room 154A-B

CONCURRENT SESSION VIII-D: TECHNICAL DEMONSTRATIONS

DATA COLLECTION AND MONITORING TECHNOLOGY

Chair: Joseph Parsons (National Agricultural Statistics Service, USA)

Technology Advancing Data Collection: Implementing Thin Client Computer Assisted Personal Interviewing into National Agricultural Statistics Services' 2011 Field Data Collection Program
Michael Gerling (National Agricultural Statistics Service, USA)
Pam Hird (National Agricultural Statistics Service, USA)
James Harris (National Agricultural Statistics Service, USA)

Data Quality Record—A Desktop Application
Judy Lieberman (U.S. Environmental Protection Agency)
Joseph Greenblott (U.S. Environmental Protection Agency)
Patricia Mundy (U.S. Environmental Protection Agency)
Michael Crow (Crow Environmental, USA)

CARI Interactive Data Access System
Carl Fisher (RTI International, USA)

RTI Quality Evaluation System (QUEST)
Susan Kinsey (RTI International, USA)
Derek Stone (RTI International, USA)
Curry Spain (RTI International, USA)

Session Organizer: Wendy Barboza (National Agricultural Statistics Service, USA)

10:30–10:45 a.m. Foyer near Room 151A-B
Break

10:45 a.m.–12:15 p.m. Room 151A

CONCURRENT SESSION IX-A:

BESTING THE ODDS: OPTIMAL REPORTING OF LOGISTIC REGRESSION

Chair: Judith Sangl (Agency for Healthcare Research and Quality, USA)

Understanding Regression Risk Analysis in Multinomial and Ordered Logistic Regressions
Morgen Miller (University of Michigan, USA)
Edward Norton (University of Michigan, USA)
Lawrence Kleinman (Mount Sinai School of Medicine, USA),
Jason Wang (Mount Sinai School of Medicine, USA)
Kasey Coyne (Mount Sinai School of Medicine, USA)

Estimating Variance for Regression Risk Analysis
Jason Wang (Mount Sinai School of Medicine, USA)
Edward Norton (University of Michigan, USA)
Kasey Coyne (Mount Sinai School of Medicine, USA)
Morgen Miller (University of Michigan, USA)
Lawrence Kleinman (Mount Sinai School of Medicine, USA)

Practitioner's Guide to Regression Risk Analysis Using SAS and Stata
Edward Norton (University of Michigan, USA)
Lawrence Kleinman (Mount Sinai School of Medicine, USA)
Morgen Miller (University of Michigan, USA)
Kasey Coyne (Mount Sinai School of Medicine, USA)
Jason Wang (Mount Sinai School of Medicine, USA)

Rank Reversal in Comparative Effectiveness Research
Lawrence Kleinman (Mount Sinai School of Medicine, USA)
Edward Norton (University of Michigan, USA)
Kasey Coyne (Mount Sinai School of Medicine, USA)
Morgen Miller (University of Michigan, USA)
Jason Wang (Mount Sinai School of Medicine, USA)

Session Organizer: David Kashihara (Agency for Healthcare Research and Quality, USA)

10:45 a.m.–12:15 p.m. Room 151B

CONCURRENT SESSION IX-B:

RESPONDENT BURDEN: MEASUREMENT AND REDUCTION

Chair: Marianne Reifschneider (Energy Information Administration, USA)

Reducing Respondent Burden: Evaluating the Progress of the 2011 SIPP-EHC
Rachael Walsh (U.S. Census Bureau)
Jason Fields (U.S. Census Bureau)

Differing Person Nonresponse Rate by Interview Length and Respondent Characteristic: Results From the 2010 and 2011 SIPP-EHC Field Tests
Daniel Perez-Lopez (U.S. Census Bureau)
Rachael Walsh (U.S. Census Bureau)

Does Length Really Matter? Exploring the Effects of a Shorter Interview on Data Quality, Nonresponse, and Respondent Burden
Scott Fricker (Bureau of Labor Statistics, USA)
Brett Creech (Bureau of Labor Statistics, USA)
Jeanette Davis (Bureau of Labor Statistics, USA)
Jeffrey Gonzalez (Bureau of Labor Statistics, USA)
Lucilla Tan (Bureau of Labor Statistics, USA)
Nhien To (Bureau of Labor Statistics, USA)

Discussant: Eileen O'Brien (Energy Information Administration, USA)

Session Organizer: Richard Reeves (Energy Information Administration, USA)

10:45 a.m.–12:15 p.m. Room 152A

CONCURRENT SESSION IX-C:
USING ADMINISTRATIVE DATA FOR SURVEY DATA QUALITY EVALUATIONS

Chair: Patrick Purcell (Social Security Administration, USA)

Linking Medicaid Administrative Records Over Time and Space: Methods and Findings, 2005 to 2007
Shinu Verghese (Mathematica Policy Research, USA)
John Czajka (Mathematica Policy Research, USA)

Misreporting of Food Stamp Participation in the NHANES: Results From a Pilot Study Linking Administrative Records to Survey Data
John Kirlin (Economic Research Service, USA)
Michael Wiseman (George Washington University, USA)
Daniel Schroeder (The University of Texas at Austin, USA)
Lisa Mirel (National Center for Health Statistics, USA)

How Good are ASEC Earnings Data? A Comparison to SSA Detailed Earning Records
Joan Turek (Department of Health and Human Services, USA)
Kendall Swenson (Department of Health and Human Services, USA)
Bula Ghose (Department of Health and Human Services, USA)
Fritz Scheuren (NORC at the University of Chicago, USA)
Charles Nelson (U.S. Census Bureau)
Edward Welniak (U.S. Census Bureau)

Evaluating Job Data in the Redesigned SIPP Using Administrative Records
Graton Gathright (U.S. Census Bureau)
Jeremy Skog (U.S. Census Bureau)
Martha Stinson (U.S. Census Bureau)

Session Organizer: Hilary Waldron (Social Security Administration, USA)

10:45 a.m.–12:15 p.m. Room 154A-B

CONCURRENT SESSION IX-D:
SAMPLING FOR EVERY OCCASION

Chair: Joy Sharp (Bureau of Transportation Statistics, USA)

Expanding the Community Address Updating System Universe
Bryan Schar (U.S. Census Bureau)
James Lawrence (U.S. Census Bureau)
Star Ying (U.S. Census Bureau)
Jim Hartman (U.S. Census Bureau)

Estimation and Sampling of Longitudinal Person-Firm Data
Kevin McKinney (U.S. Census Bureau)

Estimates of External Bias in Impact Evaluations That Select Sites Purposively
Stephen Bell (Abt Associates, USA)
Robert Olsen (Abt Associates, USA)
Larry Orr (Johns Hopkins University, USA)
Elizabeth Stuart (Johns Hopkins University, USA)

Discussant: Michael P. Cohen (American Institutes for Research, USA)

Session Organizer: Joy Sharp (Bureau of Transportation Statistics, USA)

12:15–1:30 p.m.
Open Lunch

1:30–3 p.m. Room 151A

CONCURRENT SESSION X-A:
BAYESIAN STATISTICAL METHODS

Chair: Richard Reeves (Energy Information Administration, USA)

Calibrated Bayes: Spanning the Divide Between Frequentist and Bayesian Inference
Roderick Little (U.S. Census Bureau and University of Michigan–Ann Arbor, USA)

A Semi-Parametric Approach to Account for Complex Designs in Multiple Imputation
Hanzhi Zhou (University of Michigan, USA)
Trivellore Raghunathan (University of Michigan, USA)
Michael Elliott (University of Michigan, USA)

Nonresponse Adjustment Based on Auxiliary Variables Subject to Error
Brady West (University of Michigan-Ann Arbor)
Roderick Little (U.S. Census Bureau and University of Michigan–Ann Arbor, USA)

Discussant: William Anderson (University of Richmond, USA)

Session Organizer: Richard Reeves (Energy Information Administration, USA)

1:30–3 p.m. **Room 151B**

CONCURRENT SESSION X-B:
LINKING AND QUALITY OF ADMINISTRATIVE DATA

Chair: Kathleen Short (U.S Census Bureau)

Pulling Together: Linking Unemployment Insurance and Supplemental Nutrition Assistance Program Administrative Data to Study Effects of the Great Recession
Theresa Anderson (George Washington University, USA, and Economic Research Service, USA)
John Kirlin (Economic Research Service, USA)
Michael Wiseman (George Washington University, USA)

Study of Factors Affecting Record Linkage in Federal Statistical Databases
Yuan Zhao (George Washington University, USA)
Michael Larsen (George Washington University, USA)

Qualitative Assessment of Administrative Records: The Case of State Prisons' Records
Anna Chan (U.S. Census Bureau)

Methods of Assigning Race and Hispanic Origin to Births From Vital Statistics Data
Christine Guarneri (U.S. Census Bureau)
Christopher Dick (U.S. Census Bureau)

Session Organizer: Thesia Garner (Bureau of Labor Statistics, USA)

1:30–3 p.m. **Room 152A**

CONCURRENT SESSION X-C:
DETECTING AND CORRECTING NONRESPONSE ERROR AND BIAS IN LARGE SAMPLE GOVERNMENT DATA COLLECTIONS IN THE UNITED STATES AND CANADA

Chair: Chris Chapman (National Center for Education Statistics, USA)

Total Survey Error in the American Time Use Survey
John Dixon (Bureau of Labor Statistics, USA)
Brian Meekins (Bureau of Labor Statistics, USA)

Experimental Design for Nonresponse Follow-Up of Electronic Questionnaire Survey
Joanne Leung (Statistics Canada)
Jeannine Claveau (Statistics Canada)
Claude Turmelle (Statistics Canada)

Comparison of the American Community Survey Voluntary Versus Mandatory Estimates
Karen King (U.S. Census Bureau)
Michael Starsinic (U.S. Census Bureau)
Alfredo Navarro (U.S. Census Bureau)

Discussant: Jill Montaquila (Westat, Inc., USA, and University of Maryland, USA)

Session Organizer: Chris Chapman (National Center for Education Statistics, USA)

Abstract Booklet

This section represents abstracts received as of October 2011.

The following abstracts have not been edited for content.

CONCURRENT SESSION I-A:
INCENTIVES: A LOOK ACROSS STUDIES AND TIME

More Money? The Impact of Larger Incentives on Response Rates in a Two Phase Mail Survey
Cameron McPhee (American Institutes for Research, USA) and Sarah Hastedt (National Center for Education Statistics, USA)

A growing literature has consistently demonstrated the effectiveness of prepaid cash incentives in boosting survey response rates across a variety of modes. As a result, survey researchers increasingly rely on incentives to improve response rates in surveys. Understanding how incentives can be used to achieve desired response rates under cost constraints is a critical challenge for survey researchers. In this paper we will look at the impact of incentives on the National Household Education Survey (NHES). The NHES utilizes a two phase mail approach in which households are screened to determine if there is an eligible child in the household. If the household contains an eligible child, an in-depth topical survey is sent to the household. Since the final response rate is a function of the response rate at both phases, it is critical to achieve a high response to both surveys. The incentives examined here built on findings from a smaller-scale feasibility test conducted in 2009 which tested topical incentives of $0, $5, and $15. Results from this test indicated that the $15 incentive performed significantly better than the $0 and $5 treatments; however, it was unclear if an incentive between these levels would have provided a similar increase in response at a lower cost or whether further increases in incentive levels would have continued to boost response rates. To further test this, we conducted experiments during a large scale field test that manipulated incentive level and mailing type to determine the optimal incentive level. At the screener level we tested $2 and $5 cash incentives and at the second phase, we tested $0, $5, $10, $15, and $20 cash incentives, as well as USPS Priority Mail delivery (with no incentive). This paper will report the results of these experiments and provide a cost effectiveness analysis.

What Has It Gotten Us? Examining Incentives Over Time in a Cross-Sectional Study
Tracy Hunt-White (National Center for Education Statistics, USA) and Jennifer Wine and Peter Siegel (RTI International, USA)

Surveys researchers have continued to experience declining response rates over the past decades. While survey procedures have become more efficient and the methods for collecting data have become more technologically sophisticated, people have nonetheless become more reluctant to participate. This reduced participation may impact the usefulness of the data collected to describe the population from which the individuals, or sample members, were selected. In order to make the survey worth the expense and effort, researchers are always seeking ways to gain the cooperation of sample members and/or to adjust the survey results to account for those who did not respond. One key strategy that survey administrators have used to increase response rates is incentives. However, even the effectiveness of incentives has declined over time.

This presentation will examine the history of monetary incentives over several cycles of a nationally representative, cross-sectional study of college-age students, the National Postsecondary Student Aid Study (NPSAS). Several experiments were conducted to determine the incentive levels necessary, at different cycles of this study, and the strategies for implementing them. This presentation offers a comprehensive examination of the use of incentives in this study over time. It will also incorporate the results of the latest field test experiment where data from a prior cycle of the study will be used to determine the predictors of response propensity for the current sample and prescribe the incentive amounts offered to these different propensity groups. In previous cycles of NPSAS, all sample members have been offered the same incentive amounts, regardless of their likelihood to respond. This latest experiment will provide an opportunity to examine whether predictors of response can be identified prior to sampling and data collection and test the impact of varying the incentive amounts based on the propensity to respond on response rates.

An Investment in Goodwill or Encouraging Delays? Examining the Effects of Incentives in a Longitudinal Study
Karen Grigorian (NORC at the University of Chicago, USA) and Lynn Milan (National Science Foundation, USA)

The Survey of Doctorate Recipients (SDR) is a biennial multi-mode panel survey of doctorate recipients in the science, engineering, and health fields from U.S. educational institutions sponsored by the National Science Foundation and the National Institutes of Health. In an effort to maintain high response rates, the SDR has experimented with offering monetary incentives in past rounds. In the 2003 SDR, a small late-stage experiment showed a prepaid incentive significantly improved response and yielded significantly higher quality data. The 2006 SDR included a larger controlled experiment to determine the most effective time to offer a late-stage incentive. The 2008 SDR included a controlled experiment to determine the effect of offering an early incentive to new cohort panel members and those that refused to participate in 2006, as well as a late-stage incentive offered to a randomly selected group of nonresponse panel members.

One concern about offering incentives to a longitudinal panel is the potential for conditioning sample members to expect an incentive each round, which could be detrimental to future response rates. To determine whether to include incentives in the 2010 SDR, we examined the effects of past incentives offered in the 2003 and 2006 cycles on the response rate and costs of the 2008 SDR. These data did not show a negative conditioning effect but, in fact, suggested the opposite. Hence, the decision was made to offer a late-stage incentive in the 2010 cycle.

CONCURRENT SESSION I-B:

THE NATIONAL CRIME VICTIMIZATION SURVEY (NCVS) SMALL AREA ESTIMATION PROGRAM

Generic Area Estimations and Local Crime Patterns

Michael Planty (Bureau of Justice Statistics, USA)

The National Crime Victimization Survey (NCVS) has produced national-level estimates of crime and victimization since the early 1970s. However, local policymakers and stakeholders would find the survey data more useful if crime statistics could be produced at local levels. Currently, local areas typically rely on data from police records or results from locally administered crime surveys. However, comparability issues prevent local surveys and crime data to serve as a vehicle for comparisons with other similar areas or to the national levels. To address this need the Bureau of Justice Statistics (BJS) is developing a Small Area Estimation (SAE) program to examine options for producing small area victimization estimates. These options use both direct and indirect estimation procedures, including: 1) generic area typologies, 2) direct sample boost and reallocation strategies, 3) small are estimation modeling techniques, and 4) dual frame data collection and modeling techniques. The decision to implement a particular option will rest on trade-offs between reliability and precision, generalizability, number of and types of local areas reported out, subdomain disaggregations, periodicity, timeliness, and costs.

This paper describes the generic area approach to small area estimation that leverages existing NCVS data to create a typology of generic areas or "like places" using, for example, combinations of MSA, place size, and region. These estimates can be generated for 1-, 3-, or 5-year periods to increase statistical reliability and to allow for disaggregation by key variables. While the generic area offers the least expensive option, the limitations for local stakeholders include concerns about precision, periodicity, disaggregation, and within type homogeneity.

Rethinking the NCVS: Subnational Goals through Direct Estimation

Robert Fay and Jianzhu Li (Westat, Inc., USA)

Now almost 40 years old, the National Crime Victimization Survey (NCVS) provides annual estimates of the number of victimizations by different types of crime and estimates of the characteristics of the victims. The Bureau of Justice Statistics (BJS) is currently investigating several strategies to expand the usefulness of the survey. A major goal is to shift the focus of the NCVS as almost exclusively a national survey to a revised program capable of providing relevant subnational detail useful to state and local governments.

Until now, the NCVS has been conducted exclusively by the Census Bureau. Currently, BJS is investigating a strategy that includes retaining a core NCVS preserving many of the basic features of the current design, including address sampling and personal visit as the primary mode for an initial interview. The core NCVS would continue to be conducted by the Census Bureau. At the same time, BJS is supporting research to determine the possibility of integrating the core NCVS with auxiliary or supplemental data collected by quite different survey strategies, modes, and collection agents.

This paper will detail an investigation of one aspect of BJS's overall strategy: To what extent can strategic sample boosting and allocation of the core NCVS meet a key set of subnational estimation goals through direct estimation? Direct survey estimates of annual crime rates for each of the states would require an unrealistic expansion of the survey, but the analysis reported here examines to what extent defining a restricted set of areas, such as states of over 8 million population, and using 3- or 5-year period averages in a manner similar to the American Community Survey could achieve a set of useful results to address the some of the needs of many NCVS users.

Rethinking the NCVS: Small Area Estimation Approaches to Estimating Crime

Jianzhu Li, Mamadou Diallo, and Robert Fay (Westat, Inc., USA)

The Bureau of Justice Statistics (BJS) is currently pursuing a set of strategies to expand the usefulness of the National Crime Victimization Survey (NCVS). BJS is simultaneously investigating designs for a core NCVS, which will continue to be conducted by the Census Bureau, and designs for a supplemental survey or surveys to be conducted by outside organizations. The overall goal will be to combine these sources into an integrated statistical program.

One facet of this overall research effort is to investigate how much can be accomplished from the core NCVS alone. In a companion paper, we report on the extent that selective boosting and reallocation of the sample, in combination with use of 3- and 5-year period averages, can support estimates for selected states or other subnational areas through direct estimation. In this paper we report on a parallel effort, namely, to investigate the extent to which small area estimation techniques can be applied to the core NCVS to produce useful subnational estimates.

The success of small area estimation projects often depends to a large degree on the availability of predictive auxiliary data. We previously analyzed NCVS data as a guide for the redesign of the core NCVS, but the results are also a starting point for small area estimation. We will report on our efforts to examine other auxiliary variables. The application includes a number of interesting technical challenges. The incidence of many crime variables, including the violent crime rate, is quite low, requiring a more sophisticated technical approach than models based on the normal distribution. We will investigate how to make optimal use of the survey data across time. We will discuss the problem of communicating the interpretation of these estimates to the law enforcement agencies and other, primarily nonstatistical, audiences for the NCVS data.

Designing a Low(er)-Cost Complement to the National Crime Victimization Survey

J. Michael Brick and W. Sherman Edwards (Westat, Inc., USA) and Sharon Lohr (Arizona State University, USA)

The National Crime Victimization Survey (NCVS), a panel survey sponsored by the Bureau of Justice Statistics, is the primary source of victim-reported information about personal and property crime in the United States. Methods for reducing the cost of the face-to-face NCVS data collection while producing high quality crime statistics are a high priority. In addition, estimates for states or smaller geographic areas are of great interest, but the NCVS sample cannot support estimates below the national and regional level. Efforts to produce small area estimates (SAEs) of victimization from the current or core survey have been hampered the varying quality of local sources of administrative data on victimization and the relative rarity of personal crime events.

The goal of this study is to develop a low-cost complementary survey to support a strategy for producing SAEs of crime victimization. One set of possible estimation strategies involve blended estimates drawing on both the core NCVS and the complement survey. An alternative approach is using the complement survey as a source of consistent and relevant correlates of victimization to support model-based SAEs rather than blending.

This paper describes the design of a pilot study to assess the feasibility of achieving these goals. The complement survey uses an address-based sample with mail and telephone data collection. The pilot design includes experiments to assess different data collection approaches. One would use a mail screener for all sampled addresses to (1) allow oversampling of households likely to report victimization; (2) provide correlates to support model-based SAEs; and (3) obtain telephone numbers for conducting the core NCVS interview by telephone. A second approach would use a shorter mail screener only for those addresses for which a telephone number could not be obtained from directory services, and would address only objective (3).

CONCURRENT SESSION I-C:

NEW APPROACHES TO OLD PROBLEMS IN IMPUTATION

Census Count Imputation: Mean Squared Error and Collapsing Strategies

Andrew Keller (U.S. Census Bureau)

For the 2010 census, the count imputation (CI) procedure filled in missing household status and size for the small proportion of addresses (less than one half percent) where this information was unknown. The CI model partitioned records with complete information defined by household characteristics and geography into several cells. We incorporated household characteristics (structure type, enumerator type, nearest neighbor type) and neighborhood geography (tract) as part of our cell definitions to ensure the unknown addresses were imputed from a distribution of complete addresses with similar characteristics. To account for the varying degrees of information from the unresolved addresses, we constructed three imputation types.

The similarity achieved by defining imputation cells by household characteristics and geography sometimes involved a tradeoff of creating cells with sparse counts of complete addresses. The essential goal of this research was to use research data from the 2000 Census to aid in developing collapsing procedures for cells with sparse counts. To measure the effectiveness of possible collapsing procedures, we utilized a mean squared error (MSE) approach. The MSE approach incorporated a composite estimator that examined the three types of imputation and estimated bias and variance. To assess different collapsing procedures, we performed simulations for assigning household status and size to missing data cases using two approaches for creating pseudo-missing data. This work outlines potential methodologies for collapsing sparse cells if more complete data addresses are necessary. In addition, it considers possible sizes for the maximum number of complete data addresses that constitute a sparse cell.

Implementation of Improvements to the Allocation Routine for Health Insurance Coverage in the CPS ASEC

Joanna Turner and Michel Boudreaux (University of Minnesota, USA)

The U.S. Census Bureau's Annual Social and Economic Supplement of the Current Population Survey (CPS ASEC) is a widely used source of data on health insurance coverage. The estimates of health insurance coverage date back to 1987 and are used to monitor state and national trends in uninsurance. About 10 percent of respondents do not answer any of the ASEC supplement questions and the entire supplement for these respondents is imputed. Davern et al. (2007) identified problems with the imputation of health insurance variables in the CPS ASEC. They found the full supplement imputations cases were less likely to have private coverage and more likely to be uninsured.

We collaborated with the Census Bureau and evaluated improvements to the imputation routine to fix the identified problems. The study team recommended switching the order of the hot-deck allocation matrices with public coverage first and removing the nuclear family restriction to the private coverage allocation routine. Using the 2008 estimates, making these modifications reduces the uninsured rate from 17.3 percent to 16.8 percent.

The Census Bureau will implement the recommendations beginning with the 2010 estimates of health insurance coverage released in the fall of 2011. In addition to making these changes going forward the Census Bureau will release a revised time series of estimates. The CPS ASEC has had a number of methodological changes applied over time. In this paper we discuss the decisions that were made to implement the recommendations for previous years, how for back the revised time series goes, and evaluate the revised time series.

Evaluations of Imputation Methods to Improve the American Community Survey Estimates of the Group Quarters Population for Small Geographies

Mark Asiala, Michael Beaghen, Alfredo Navarro, and Lynn Weidman (U.S. Census Bureau)

The Census Bureau has recently developed a new methodology to improve the American Community Survey (ACS) estimates of the group quarters (GQ) population for small areas. What motivated this work is that while the ACS GQ sample was designed to produce estimates at the state level, the estimates of the GQ population contribute to ACS estimates of the total resident population for substate areas such as counties and tracts. Consequently, there are small geographies which either do not have GQ sample or have GQ sample that is not representative of the area, which can lead to distorted estimates of characteristics and/or total population for these geographies. The approach taken is to impute whole person records (and weight them appropriately) to GQ facilities which appear on the sampling frame but were not selected into sample. This paper examines the results of this method using real ACS data, including comparing the results of the new methodology with those of the current ACS methodology.

Wage Imputation in the OES Survey: A Model-Based Approach

Jane Osburn (Bureau of Labor Statistics, USA)

The Bureau of Labor Statistics Occupational Employment Statistics Survey collects data over a three year cycle on the wages and industrial /occupational category of each employee in a total of approximately 1.2 million establishments, spanning the non-farm private and public sectors in the U.S.

Wage imputation in the OES Survey requires a process that matches the wage levels of the respondent units in a given industry / establishment size cell to the wage levels of the non-respondent units in the cell. Currently, OES procedures use the geographic location of establishments as a proxy measure of the wage level of each establishment in an area. Non-respondent wages are imputed with the mean occupational wage of donor cells comprised of respondents in the same MSA/ Industry/ Establishment size cell. The remaining missing data is imputed using a complex hierarchical process that collapses the cells to the State and National levels. For States containing areas with widely varying wage levels, some outcomes could be improved.

The proposed revision uses a statistical estimate of the wage level of each area as a proxy measure of the wage level of each establishment in the area. A linear mixed model is used to estimate the wage level of each area, and these estimates are used to form donor cells from areas with wage levels similar to those of the location of the non-respondent.

A simulation study is used to examine the relative performance of the current and proposed estimators, along with alternative estimators that incorporate an OES-data proxy for the establishment average wage that is available from the Quarterly Census of Employment and Wage Program and is currently being investigated by BLS Statisticians.

Preliminary results suggest that an estimator that combines the proposed estimator and the proxy QCEW wage performs best.

CONCURRENT SESSION I-D:

ISSUES IN INTERNATIONAL MIGRATION AND U.S. IMMIGRATION

Estimates of International Migration for United States Natives

Christopher Dick, Eric Jensen, and David Armstrong (U.S. Census Bureau)

Net international migration (NIM) is a component of change in the Census Bureau's population estimates. This paper focuses on one aspect of NIM, native migration. It is difficult to estimate native migration because of a lack of data on both immigration and emigration of U.S. natives. In the Population Estimates Program, we use a residual method and data on U.S. natives collected by censuses in other countries. In this paper we present an alternative approach that involves calculating a residual using data from the American Community Survey (ACS), Census 2000, and vital statistics from the National Center for Health Statistics (NCHS).

There are both benefits and limitations to using the alternative approach instead of the method currently used in the Population Estimates Program. For example, estimates of native migration are small relative to the size of the native population residing in the United States. Since the alternative approach calculates the residual using data on U.S. natives, even small errors in the estimate of natives may have a large impact on the residual-based estimate of native migration. However, the alternative approach provides two major benefits. First, it uses data from the United States – instead of a large number of countries – allowing for greater understanding of the quality and limitations of the data. Second, we can more frequently and easily update our estimates when using U.S.-based data.

In sum, in this paper we will describe the data and methods that can be used to estimate the international migration of U.S. natives since 2000, provide results from our ongoing research, and discuss the benefits and limitations of each approach

Forecasting Immigration Trends with Bayesian Structural Time Series Models

Scott Borger and Anthony Kassekert (Office of Immigration Statistics, USA)

Historically, statistical forecasts of immigrant and non-immigrant application receipts and approvals have been accomplished through Box-Jenkins/ARIMA models based solely on the previous observations of the forecasted series. These time series provide valuable forecasts but are incapable of informing policy decisions. Questions regarding the impact will an increase in the application fee or the hiring of additional processing staff are neglected by solely examining previous observations of applications or approvals.

We employ a structural time series model which includes several policy dimensions and administrative controls to predict the receipt of applications for naturalization and legal permanent residence in the U.S. In particular, we test to see if application price, administrative work hours, application pool size, and policy shifts can improve out-of-sample predictions. We employ Bayesian estimation because of its statistical advantages in dealing with unit roots and in order to assess the reasonableness of using semi-informative priors to inform the prediction error component.

Imputing Legal Status of Foreign-Born Persons on Surveys: Two Approaches

Dean Judson (Decision Analytics, USA) and Sharon Long (University of Minnesota, USA)

Legislation and Policymaking often run ahead of our ability to produce data usable for implementation. In the case of the Affordable Care Act, this has occurred with respect to state-level estimates of the legal status of the immigrant population. Several provisions that affect unauthorized immigrants and are directly relevant to state planning are included in the bill. For example, unauthorized immigrants are exempted from the

'individualmandate' to maintain coverage and are prohibited from purchasing health insurance coverage in federal or state health insurance exchanges. Because of provisions such as these, unauthorized immigrants are expected to comprise a substantial portion of the remaining uninsured population after the ACA is in full effect. Accordingly, accurate state-level estimates of the size of the unauthorized immigrant population are needed as states determine the appropriate safety net capacity for the remaining uninsured population after 2014. However, no Federal agency makes a state-by-state estimate of persons by legal status. Similarly, no specifically-state-representative Federal survey asks "legal status questions" of the foreign-born. This talk will focus on two experimental methods for imputing legal status on foreign-born persons. The first of these involves developing an imputation model from the Survey of Income and Program Participation and using that model on other target surveys; the other involves developing a latent class model to classify foreign-born persons. Some caveats and other preliminary considerations will be discussed.

Estimating Personal Transfers from the United States

Rachel Soloveichik and Anne Flatness (Bureau of Economic Analysis, USA)

As part of the U.S. International Transactions Accounts, the Bureau of Economic Analysis (BEA) prepares estimates of funds sent by U.S. residents to friends and family abroad. These personal transfers, a major component of remittances, are frequently sent through informal channels making it difficult for BEA to directly measure the monetary value of transfers abroad. As a result, BEA developed a model that estimates personal transfers based on immigrant demographics and income. Every year, BEA applies that model to known population data and calculates total personal transfers sent abroad.

BEA's current model for estimating personal transfers is based on a variety of government and academic sources which provide information on immigrants' propensity to remit and the percent of income remitted. However, recent events such as the financial crisis and the downturn in the housing sector may have changed immigrants' behavior. In August of 2008, the Bureau of Labor Statistics and the Census Bureau jointly conducted a large survey of personal transfers and other immigration topics as part of the Current Population Survey (CPS).

In this paper, we consider updating the BEA model of personal transfer behavior based on newly available CPS data. We evaluate these potential modifications by applying them to known population data to estimate personal transfers sent by U.S. residents to selected foreign countries. We then compare these estimates of transfers sent from the United States with the reported transfers received by selected foreign countries, as estimated in the national accounts of those countries and the industry literature. In theory, total transfers sent should equal total transfers received. However, the two often differ in practice because of differences in the definition of personal transfers, estimation methodology, or source data.

CONCURRENT SESSION II-A:

SIMULATION MODELS OF FEDERAL SURVEYS

The Case for Simulation Models of Federal Surveys

Lawrence Cox (National Institute of Statistical Sciences)

This paper will describe the advantages to survey methodology, operations and cost control presented by simulation models of survey processes. Such models might be limited to survey costs, field operations, or dealing with nonresponse, or, conceivably, might simulate an entire survey including design, estimation and incorporation of ancillary/exogenous information. A brief summary of the April 2010 NISS Survey Simulation Workshop will be provided.

The key discussion questions I hope to stimulate are:
Survey simulators—worthwhile or not?
Where (which survey[s]) and what (operations/components) to simulate?
How to organize and finance such an effort?

An ultimate vision is for a general purpose Survey Simulation Laboratory for Federal surveys. Discussion of the feasibility and desirability of this concept is welcomed.

Simulating NHIS Field Operations

Bor-Chung Chen (Federal Railroad Administration, USA)

Discrete-event simulation modeling has become the most commonly used tool for performance evaluation of stochastic dynamic systems in science and engineering. The field operations of surveys can be classified as one of these stochastic dynamic systems. Simulation modeling provides, if feasible, flexibility to build either

aggregate or detailed models. This presentation gives an overview of simulation modeling methodology and describes the simulation and modeling of simplified field operations for NHIS (National Health Interview Survey). We will describe simulation study steps needed to develop a simulation model that is a valid representation of an existing/proposed real system. We use the 2004 NHIS CHI (Contact History Instrument) data for the input modeling of the simulation. From this study, we have shown that simulation modeling can be used for optimizing the field operations by setting the controllable parameters before a decision is made and implemented. The cost savings might be enormous and would not be at the expense of the response rate.

WSSM: World's Simplest Survey Microsimulator

Alan Karr (National Institute of Statistical Sciences, USA)

The World's Simplest Survey Microsimulator (WSSM) is the first version of an extensible simulation laboratory for surveys. By design, it is deliberately and perhaps even grossly oversimplified.

WSSM's short-term purpose is to support sensitivity analyses that not only demonstrate how even oversimplified models can reflect policy and operational decisions but also to inform the course of more detailed modeling efforts in the future.

The WSSM has three essential characteristics. First, the entire underlying population and the behavior on which the survey is focused are both simulated, and serve as "ground truth" for calculating measures of data quality. Second, the survey responses themselves are simulated. And third, the measures of data quality are based on the fidelity of inferences drawn from the survey responses compared to the same inferences based on the entire population.

The long-term purposes are to provoke suggestions for enrichments and to serve as a foundation for implementing them.

CONCURRENT SESSION II-B:

ENHANCING SURVEYS THROUGH DATA LINKAGE

Retrospective Linking of ECLS-K and ECLS-B Reading Scores

Carolyn Fidelman (National Center for Education Statistics, USA)

Plans for making scores between different cohorts of examinees verifiably construct-equivalent are ideally built into a test's design from the beginning. But it is sometimes in hindsight that we realize we might be able to leverage the results from two very similar studies in order to carry out new comparisons. In cases where test forms for different cohorts are very similar yet have not been thoroughly tested for equivalence before going to the field, can we reformulate the scores in order to achieve a degree of parity adequate for score linking? The psychometric properties of the items of two purportedly similar multistage kindergarten reading tests administered to cohorts five years apart, in 1998 and 2006, were examined and a subset of the test items was used in order to meet the assumptions of item response theory parameter estimation and score linking. Specifically, a unidimensional set of common items was identified from the Stage 1 tests and the unique items measuring that same construct within the three Stage 2 test levels and across the two cohorts were able to be put on the same scale. IRT chain linking within and across cohorts was found to be the best approach for the Early Childhood Longitudinal Study (ECLS) 1998 and 2006 Reading assessments. By putting the scores of all six test forms on one scale, we can offer new score tables for use in studies whose aim is to compare or contrast the two cohorts of examinees. The results from this linking study also remind us of some common ways in which construct validity can be verified and maintained over multiple cohorts or how modifications in structure can be made while maintaining the baseline measurement properties that enable their continued use in such analyses.

The Role of Community Health Centers in Providing Safety-Net Access to Health Care

Sandra Decker (National Center for Health Statistics, USA) and Frederic Selck (Johns Hopkins University, USA)

The number of Community Health Centers (CHCs) has grown over the past decade and is expected to grow due to funding from the Patient Protection and Affordable Care Act. This paper examines the relationship between CHC availability and use of ambulatory health care among nonelderly people, especially those who are uninsured or insured through the Medicaid or Children's Health Insurance Program (CHIP). Data come the 2005-2008 National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS). We use the Dartmouth Institute's Primary Care Service Areas (PCSAs) as our measure of market areas, and examine the increase in CHCs by PCSA over the 2005-2008 time period using patient zip code of residence from the NAMCS/ NHAMCS and geo-coded information on the location of all CHC delivery sites in the U.S. Using a difference-in-differences approach with area and time effects, we investigate how the increase in the number of available CHC delivery sites within PCSAs affect the site of care (physician office, hospital out-

patient department (OPD), hospital emergency department (ED), or CHC)) for all ambulatory care visits, and for visits due to "ambulatory care sensitive" (ACS) conditions. We find that the uninsured and those on Medicaid/CHIP were more likely to have ambulatory care visits at a CHC or an ED compared to others. Even in the short 2005-2008 time period, the proportion of visits that took place at CHCs increased from 2% and 3% of uninsured and Medicaid/CHIP visits to nearly 6% and 10%, respectively. The probability that a Medicaid/CHIP or uninsured ambulatory care visit took place in an ED decreased by 0.1 percentage point for each additional CHC delivery site in a PCSA. The effect for ACS visits is larger for Medicaid/CHIP providing significant evidence of substitution of Medicaid/CHIP ambulatory healthcare visits between CHCs and EDs.

Use of Synthetic Data in Testing Administrative Records Systems

K. Bradley Paxton and Thomas Hager (ADI, LLC)

In the development of the U. S. Decennial Response Integration System (DRIS), which processed the 2010 Decennial Census, we at ADI, LLC supplied some unique test materials called Digital Test Decks◊. These decks appeared to be real Census forms with data printed on them by a variety of respondents, but they were not: they were "synthetic" forms containing known truthed data populated on the forms useful for testing and evaluating data capture performance. The data on these forms was statistically representative of the population, was internally consistent, and realistic.

We have found that such realistic synthetic data is very valuable for a variety of other test purposes applied to complex data classification systems generally, such as record linkage, health care, and intelligence. Over the last few years, we have developed a novel new approach generating of this data called the Dynamic Data GeneratorÔ. This generator can create very large and complex data sets designed for testing which we call a Great Automated Model Universe for Test (GAMUT), from which suitably formatted data streams may be extracted to suit particular test objectives.

Applied to the effective use of Administrative Records (AR) in future Census surveys, we describe the use of a particular GAMUT to test matching accuracy between a Census-like data stream, and a data stream from another Agency. We indicate the value of this approach to those developing AR systems to validate and improve Census data. Example test files will be illustrated, and some key test data properties described, such as longitudinal consistency, realistic name and address morphing, insertion of engineered errors, etc.

Finally, we will show how using these synthetic yet realistic data sets, one can perform timely, cost-effective and precise scoring of system matching results, including analysis of true and false positives to select optimal classifiers.

CONCURRENT SESSION II-C:

USING MULTI-MODE TECHNIQUES

Return to Sender: Improving Response Rates in Two-stage Mail Surveys

Saida Mamedova and Cameron McPhee (American Institutes for Research, USA)

Since 1991, the National Center for Education Statistics (NCES) has used the National Household Education Surveys Program (NHES) to collect data on such topics as early childhood care and education, children's readiness for school, and parent involvement in education. From 1991 through 2007, the NHES used a list-assisted RDD CATI survey. However, like most RDD surveys, NHES response rates have been declining over time and the increase in households converting from landlines to cell phone-only service has raised concerns about population coverage.

In January 2011 NCES began a large-scale field test of a redesigned multi-mode survey on a nationally representative sample of approximately 41,000 addresses in the United States. The primary data collection mode used in the redesigned NHES is a two-phase mail survey. Included in the field test were multiple experiments intended to determine how to best improve response rates and population coverage, such as the use of pre-notice letters, differing incentive levels and different versions of the first-stage screener instrument. Preliminary analysis indicates that the use of pre-notice letters, higher incentives, and shorter instruments improve response rates. However, it is not yet clear how these treatments may interact with each other, potentially confounding initial results, nor do we know if certain treatments garner earlier response, therefore requiring fewer follow-up up mailings. The analyses for this paper will use bivariate statistical tests and logistic regression to determine if certain combinations of treatments led to higher rates of response at multiple mailing waves. Additionally, our analyses will examine whether the most successful first-stage treatment(s) also lead to the highest response at the second stage.

Finally, using demographic data from the sampling frame and the screener questionnaire such as region, urbanicity, parental education, household size, and home tenure, we hope to determine whether the redesigned mail survey improves population coverage compared to the RDD design.

Multi-mode Survey Response Process for a Complex Economic Survey

Thomas Falconer and Richard Hough (U.S. Census Bureau)

Completion of an economic survey can be very challenging. This process becomes more complex when the survey response requires communication with people from multiple business units or divisions within a diverse organizational structure.

The Business R&D and Innovation Survey (BRDIS) is a new survey that replaces the former Survey of Industrial Research and Development (SIRD) collected since the 1950s. The BRDIS was fielded for the first time in 2008. Like the SIRD, the BRDIS is conducted under a joint partnership agreement between the National Science Foundation and the U.S. Census Bureau. The BRDIS expanded data topics include data on R&D expense, R&D costs that are funded by others, R&D employment, innovative activities by companies, patents and intellectual property protections by private sector companies in the U.S. with a sample of approximately 40,000 companies.

The wide range of topical areas contained in the survey does require the survey coordinator to gather data from multiple people within different working units in the companies. The survey offers respondents two primary modes to submit the data, a paper questionnaire and an internet based instrument. To facilitate the collection of data within companies, excel spreadsheets were developed and made available to survey coordinators and the capability to upload the spreadsheets directly into the internet instrument was made available as well. This paper will examine the differences in response rates and data quality across the various modes for survey responses. We will also examine the impact on the timing of response across the various modes and size of the company as defined by the total R&D costs.

Redesigning the National Surveys of Long-Term Care Providers

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

Producing national data on long-term care providers, their staffs, their services, and the people they serve is especially challenging because of the multiplicity of settings and populations, and because of resource limitations. The National Center for Health Statistics (NCHS) previously conducted the National Nursing Home Survey (NNHS) and the National Home and Hospice Care Survey (NHHCS) periodically, with five to seven years between their most recent iterations.

In 2011, NCHS launched its new integrated strategy for obtaining nationally representative statistical information about the supply and use of the spectrum of paid, regulated long-term care (LTC) providers in the United States—the National Surveys of Long-Term Care Providers (NSLTCP). NSLTCP will replace NNHS and NHHCS.

The NSLTCP is being designed to (1) broaden the coverage of LTC providers beyond nursing homes, home health care agencies, and hospices to also include residential care facilities, adult day service centers, and over time other LTC providers; (2) use nationally representative administrative data on LTC providers for which they are available (i.e., nursing homes, home health agencies, and hospices); (3) collect primary data—using nationally representative mixed-mode sample surveys of providers—only for providers for which nationally representative administrative data do not exist (i.e., residential care facilities, adult day service centers); and (4) enable comparisons across LTC provider types at the same point in time and more frequently than in the past decade. The NSLTCP survey and administrative data will be used to develop an overview report every other year on the supply and use of paid, regulated LTC in the United States. This paper reviews the reasons NCHS initiated a redesign, the goals for the redesign, differences between the redesign and NCHS' previous LTC provider surveys in key areas (e.g., data collection mode, content covered), and the challenges and opportunities in implementing the redesign.

CONCURRENT SESSION II-D:

ECONOMIC STATISTICS

Measuring the Economic Impact of Intellectual Property Reliant Industries

Brittany Bond, Jocelyn Burston, and David Langdon (U.S. Department of Commerce)

Economists in the Economics and Statistics Administration (ESA) at the U.S. Department of Commerce used publicly available data from U.S. government data agencies to estimate the economic value attributable to protected intellectual property in the U.S. economy from 2001 through 2010. Undertaking an unparalleled approach, we developed a system to select industries (at the 4 digit North American Classification System level of detail) which were most reliant on government protection of intellectual property (IP) in the forms of patents, trademarks and copyrights. We correlated this with additional IP indicators such as: IP-intensive occupations, R&D spending and investment, revenue generated by licensing rights to protected products, as well as brand intensity measurements. Finally, we include a discussion of the economic impact from the standpoint of upstream and downstream industries indirectly reliant on IP protection using Bureau of Economic Analysis (BEA) input-output, make/use, and employment requirements tables.

Our findings include revenue generated from protected IP products and their share of GDP, the number of jobs that intellectual property and reliance on protected intellectual property rights create directly or indirectly as well as the characteristics of such jobs and the people employed in them. International trade balance effects and exports directly or indirectly from IP intensive industries are evaluated, although no cross-national comparisons are made in this study. Finally, we discuss changes over time and establish an estimate of the total economic value current IP protection yields to the national economy.

This transparent report sets a standard in the examination of U.S. industries reliant upon government protections of intellectual property (IP), analyzing their characteristics in contrast to non-IP-reliant industries, and measuring their impact on the economy.

Modernization of Benchmarking Economic Time Series at the U.S. Census Bureau
Irene Brown (U.S. Census Bureau)

A common problem faced by government agencies that collect and publish time series data is to maintain a consistent set of time series. Generally, this is done by adjusting monthly or quarterly series to match the less frequent series (annual or Census) that are assumed to be of higher quality, while simultaneously preserving some characteristic of the original more frequent series. Benchmarking refers to techniques used to solve this problem of incoherence between times series data measured at different frequencies.

Economic Programs at the Census Bureau use the Causey-Trager procedure for benchmarking monthly and quarterly series to annual series and to the Economic Census every five years. This procedure, developed thirty years ago, uses an iterative, nonlinear technique known as steepest (gradient) feasible descent to obtain the benchmarked series by optimizing a measure of growth-rate preservation.

Even though the software implementation of the procedure has changed a little with each new programming language, the methodology has not changed in the last three decades. This paper describes the Causey-Trager benchmarking methodology along with current and future research to improve benchmarking methods of economic time series at the U.S. Census Bureau.

Using A Composite Index of Financial Conditions Indicators to Predict Turning Points in the U.S. Business Cycle
Gad Levanon, Jean-Claude Manini, Ataman Ozyildirim, and Jennelyn Tanchua (The Conference Board, USA)

The latest global recession highlighted the importance of the link between the financial sector and the real economy; however, the relationship between the two sectors and their timing is not always found to be consistent. Financial indicators such as yield curves and stock prices have been extensively used as leading indicators of economic activity due to their forward looking content. Over the past three decades, many new financial indicators have become available. But, since most of them have not been available for long enough periods of time, very little research has been conducted to evaluate their utility as leading indicators of recessions. While some credit market indicators appear too noisy or appear to lag the business cycle, it may be possible to identify some indicators related to the availability and cost of credit that are useful in predicting recessions and upturns. Aggregating these in a composite index may offer advantages over relying on them individually. In this paper we propose to evaluate the usefulness of a large number of indicators from financial and credit markets in predicting recessions. First, we establish the criteria which are helpful for assessing whether and when such financial indicators as interest rate swaps, credit default swaps, certain corporate-treasury spreads, the Federal Reserve's senior loan officer survey, etc. generate signals of recessions. We then choose the best ones and aggregate them into a single leading financial index. We argue that this index can be helpful to estimate recession probabilities better than those that would be provided by individual indicators. As opposed to other recent financial indexes created to measure financial instability or volatility, the purpose of ours is to signal recessions in the US economy.

Early Estimates for the Retail Trade With A Mixed Model
Pieter Vlag (Statistics Netherlands)

Like several other European countries, Statistics Netherlands increasingly uses the Value Added Tax (VAT) registration as administrative data source for small and medium (SM) sized enterprises. Yearly, quarterly and monthly turnover estimates are based on the VAT-registration. Problem when using VAT for quarterly and monthly turnover estimates is that part of the VAT-data are missing because they arrive late or are reported for another periodicity. This is especially the case for the first estimates of the retail trade – 28 days after the end of the month - when no VAT data at all are available. Hence, these early estimates have to be based on a survey for the largest enterprises only.

These early estimates are slightly biased compared to later definitive monthly estimates when part of the VAT data for the SM enterprises become available. This bias becomes even more pronounced when relating them to quarterly turnover estimates, when also data from quarterly VAT reporters can be incorporated in the estimates. Previous research revealed that this bias is mainly caused by a generally slightly lower growth rate for SM enterprises compared to large enterprises (due to economic reasons).

In this study we present a method to correct the bias in the first monthly estimates. The method is based on using observed differences in growth rates between SM and large enterprises in previous periods. During the presentation several methodological options to include the difference in growth rate for the previous periods in the first estimates will be discussed. The results show that this approach seems to be quite promising as it is quite robust for changes in the business cycle whatever the exact methodological solution chosen.

CONCURRENT SESSION III-A:

USE OF ADMINISTRATIVE RECORDS DATA IN THE 2020 CENSUS

Possible 2020 Census Designs and the Use of Administrative Records: What is the Impact on Cost and Quality?

Maryann Chapin (U.S. Census Bureau)

While administrative records have been supporting decennial censuses for decades, we are for the first time considering their use in our most large and costly operation—non-response follow-up. We believe that using administrative records to supplement non-response follow-up is the single biggest potential cost saver for the 2020 Census. Early 2020 Census planning has focused on operational design work in three areas, which are the major cost and quality components for the census: Establishing Where to Count; Enumeration Activities; and, IT and Field Operational Infrastructures. The idea will be to model, analyze, and test different approaches to conduct or support these major census functions. This paper will describe and compare alternatives based on different design options, but will focus on the alternatives that include administrative records. It will also include a discussion about potential coverage and technical issues associated with using administrative records in lieu of collecting data directly from the respondent.

2010 Census Simulation: Comparing Administrative Records and Decennial Data

Amy O'Hara (U.S. Census Bureau)

The 2010 Census Simulation measures the coverage of administrative records data compared to 2010 Census results. In this study, administrative records from federal government agencies and commercial sources are matched to census results at the unit address and at the person level. This paper presents preliminary results from the simulation, including national, state, and congressional district counts. Comparisons by person characteristics such as race, Hispanic origin, age, sex, and housing tenure are also included. The 2010 Census Simulation is the first national level evaluation of administrative records data coverage. Results from the study will be used to inform future uses of administrative records at the U.S. Census Bureau. Full results from the study will be available in 2012.

Public Attitudes Toward the Use of Administrative Records in the U.S. Census: Does Question Frame Matter?

Nancy Bates, Monica Wroblewski, and Joanne Pascale (U.S. Census Bureau)

Currently, the Census Bureau is looking to leverage administrative records housed elsewhere in the government to supplement and/or replace costly nonresponse followup operations in future Censuses. Before embarking on this new methodology, the agency must be mindful of public opinion as it poses new concerns about privacy, confidentiality and informed consent. Previous research presents a somewhat conflicting picture of the topic – on one hand, public favorability toward the use of administrative records looks to be declining (Singer, Bates, Van Hoewyk, 2011). On the other hand, a recent study of public willingness to grant informed consent to record use paints a more optimistic picture (Pascale, 2011).

In summer 2011, the Census Bureau will conduct the second iteration of a survey designed to understand public attitudes toward the decennial Census and the potential barriers and motivators to Census participation. Included in the survey is a set of questions to evaluate overall attitudes toward using administrative records and various options for communicating the use of administrative records to the public. The instrument uses a randomly assigned split-ballot questionnaire that presents three different framing contexts: framing the use of administrative records in terms of a cost savings, a decreased burden, and a control in which the ques-

tions are asked without reference to any benefits. The results will enable Census to better understand public opinion about the use of administrative records and how the agency might best go about communicating the new method. Ultimately, the Census Bureau will use these results to inform the 2020 Census communications campaign and will allow administrative records usage messaging to be tailored to different segments of the population.

CONCURRENT SESSION III-B:

ADVANCED TECHNIQUES FOR EVALUATING AND ADJUSTING FOR UNIT NONRESPONSE ACROSS DIFFERENT MODES OF DATA COLLECTION

Comparing Weighting Methods of Adjusting for Logistic Unit Nonresponse

Phillip Kott and Dan Liao (RTI International, USA)

Two common methods or adjusting for unit nonresponse fit a logistic response function to the original sample either using maximum-likelihood or pseudo-maximum likelihood methods (the latter incorporates the design weights). Then, either the fitted probabilities of response are inverted and treated like second-phase sampling weights or the fitted probabilities of response are used to sort the sample into five (say) reweighting classes.

A third method indirectly fits the logistic response model using a calibration equation as can be done with the WTADJUST procedure in SUDAAN®. When the survey variable of interest is roughly a linear function of the covariates in the logistic-response model, the third weight-adjustment method can be on stronger theoretical ground than its more commonly-used alternatives.

After reviewing some of the theory, we will compare various variants of these methods empirically using data from the Drug Awareness Warning Network (DAWN) survey of drug-related visits to hospital emergency rooms. Like other establishment surveys, DAWN features an auxiliary variable available for all hospitals on the frame. This variable can be used to improve the statistical efficiency of estimated totals because many survey variables of interest have a rough linear relationship with it.

To be fairer to the non-calibration methods, we will also evaluate nonresponse adjustment to the combined ratio estimator (so that they too can exploit this relationship) and consider the possibility that the log-odds of response is a function of the log of this auxiliary variable, not a form well suited for estimating population totals of survey variables.

A Response Propensity Based Evaluation of the Treatment of Unit Nonresponse for Selected Business Surveys

Katherine Jenny Thompson and Katrina Washington (U.S. Census Bureau)

In the world of survey design, methodologists select a “representative” sample that achieves targeted reliability from a complete frame. In the world of survey collection, not all sample units respond (complete non-response), and those that do will not always provide data on every questioned characteristic (partial non-response). Business surveys publish totals. Consequently, complete-case analyses are negatively biased, so the survey practitioner assumes a model for the nonresponse and develops adjustments to their survey estimates accordingly. Under an ignorable response mechanism, adjustment can be accomplished via weight adjustment or imputation. The choice of adjustment method and model(s) should be determined by the survey publication objectives and supported by analysis of the survey response set.

In this paper, we focus on validation of the unit nonresponse adjustment procedure used in two business surveys: the Services Annual Survey (SAS) and the Quarterly Services Survey (QSS). These programs use ratio imputation methods to account for both unit and item nonresponse. The ratio imputation model utilizes a prediction model between the missing (dependent) variable and the auxiliary (independent) variable with a covariate-dependent response mechanism within each imputation cell. Auxiliary variables differ by item. This approach allows for maximum flexibility in modeling and preserves the expected cell totals, but does not preserve multivariate relationships between items and creates variance estimation challenges. An alternative approach would be to develop a single weight adjustment procedure under either a single covariate-dependent response mechanism or using a missing-at-random (“quasi-randomization”) adjustment procedure, preserving multivariate item relationships and facilitating accurate variance estimation (Shao and Thompson, 2009) at the cost of reduced imputation model flexibility. Using historical data, we examine the ignorable response assumptions under either approach, then develop and compare the associated response propensity models. In addition, we will empirically evaluate the prediction models associated with the considered adjustment approaches.

Nonresponse and Panel Attrition in a Mobile Phone Panel Survey

Marek Fuchs (Darmstadt University of Technology, Germany)

In recent years landline telephone coverage rates have declined considerably in many industrialized countries. At the same time, mobile phone usage has increased dramatically. In reaction to these trends survey researchers have included mobile phone numbers in telephone surveys using a dual frame approach. However, methodological knowledge on how to implement mobile phone surveys is emerging slowly. In addition, panel studies using mobile phones have not been addressed at all.

Since 2009 we have build a probability-based longitudinal mobile phone panel study in Germany which is primarily concerned with methodological implications of administering surveys over a mobile phone device (N = 1,600).

In this paper we will discuss the feasibility of conducting mobile phone panel surveys. We will particularly focus on non-response and panel attrition over the course of 7 panel waves. We will assess initial non-response rates and non-response bias on the recruitment stage and subsequent panel attrition. Results indicate that initial response rates are rather low. Also, initial non-response biases are considerably high with respect to age, gender, education and employment status. Wave to wave panel attrition is knowledgeable extend, nevertheless non-response biases are only moderately modified over the course of the panel. In addition to these descriptive findings, results of two field-experiments within the panel on incentives and a mode switch for non-responders will be reported. Preliminary findings suggest that both, incentives and a mode switch have the potential to improve response rates. However, non-response and non-response bias remain on considerable high levels.

Rich List Data and Adjusting for Nonresponse Bias

Paco Martorell and David Loughran (RAND, USA) and Jacob Klerman (Abt Associates, USA)

Post-survey weighting adjustments are commonly used to mitigate nonresponse bias, but in most settings it is not possible determine to what extent such adjustments eliminate bias. Using survey records for the entire sample - respondents and nonrespondents - linked to administrative earnings records, this paper examines the degree to which reweighting eliminates nonresponse bias in the Department of Defense's Status of Forces Survey of Reserve Component Members. Unweighted tabulations of the administrative data, as well as tabulations weighting by simple demographics, show considerable unit and item nonresponse bias. However, reweighting using detailed covariates drawn from military personnel records eliminates almost all of this bias. These results provide a striking illustration of the larger point that when rich auxiliary data is available for the entire sample, reweighting using this information may be an effective way to reduce nonresponse bias.

CONCURRENT SESSION III-C:

USE OF PARADATA AND DASHBOARDS

Comparing Estimates and Item Nonresponse Rates of Interviewers Using Statistical Process Control Techniques

Robyn Sirkis (U.S. Census Bureau)

This paper discusses research using statistical process control with paradata obtained during data collection for the National Health Interview Survey (NHIS). Statistical process control (SPC) involves using statistical techniques to measure and analyze variation in operational processes. The goal is to not simply monitor, but to improve the quality of the process over time. For this paper, we group interviewers into clusters based on housing unit and demographic characteristics and produce control charts which examine the variation of the process over time for each cluster. We compare the means of interviewers within each cluster to determine if they are significantly different from the overall mean of the process, and examine some of the potential causes of process variation using selected control charts and SPC techniques. The charts displaying average estimates will include the average of all responses for each interviewer. We will compare several charts including item nonresponse rates, average estimates, and the proportion of responses for specific answers categories charts. We will compare the item response rates charts to the estimates charts to determine if the process is not in control for the same interviewers.

Paradata in the Consumer Expenditure Survey

Laura Paszkiewicz (Bureau of Labor Statistics, USA)

In 2005, the Consumer Expenditure Survey (CE) began recording information about the data collection process through the Contact History Instrument (CHI). The CHI data are entered by the interviewers after each contact attempt with respondents and potential respondents. The interviewers enter information about the attempt, including the mode of contact (e.g., by telephone or in person), the language the interview was conducted in,

and perceived reasons for non-responses. The CHI also records information about the total number of contacts required for completed interviews, and contact attempt dates and times. Collectively, with other information already collected in the Computer Assisted Personal Interview (CAPI), these data make up the paradata in the CE.

In this presentation, an overview of CE research that has been conducted using the newly available paradata will be presented. All research uses data from the CE Interview Survey. Recent research topics include investigating the potential use of interviewer observed information to adjust weights for non-response bias in surveys; determining the optimal number of contact attempts as a way to reduce costs and maintain quality; using the CHI to identify reluctant respondents; and measuring the effects of pre-paid incentives on the first occurrence of non-response in panel data, while controlling for survey participant concerns collected in the CHI.

Additionally, in 2009 the CE paradata were made available on the public-use microdata CD. A general description of the newly available paradata files and structures on the public-use microdata CD will be provided, including an overview of the available public-use paradata variables.

The Implementation of Dashboards in Governments Division Surveys

Terri Craig and Carma Hogue (U.S. Census Bureau)

Dashboards have gained popularity as a quick way to monitor metrics in manufacturing, business, healthcare, and even in election polling. After facing challenges with low unit response rates and Total Quantity Response Rates (TQRRs) for the 2007 Census of Governments: Finance, we decided to develop a dashboard that would monitor response in the very early stages of processing. The initial dashboard developed for the Local Government Finance Survey attempts to yield compliant unit response rates for each state by type of government, compliant TQRRs for each state by key item by type of government, and an adequate number of responses and acceptable response rate in each imputation cell. This dashboard has been used as a prototype for monitoring response for other surveys of governmental units. Initial dashboards have been developed for the Local Government Finance Survey, Annual Survey of Government Employment and Payroll, and Quarterly Tax Survey – Local Government Property Tax Component. This paper presents the theoretical and technical challenges that we faced in constructing each dashboard.

Development of a SAS Dashboard to Support Administrative Data Collection Processes

Marianne J. Reifschneider and Susan Harris (U.S. Energy Information Administration)

The U.S. Energy Information Administration (EIA) collects, analyzes, and disseminates independent and impartial energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment. To accomplish this mission, EIA conducts a comprehensive data collection program that covers the full spectrum of energy sources, end uses, and energy flows. As part of EIA's survey business process, the survey managers regularly monitor data collection via a series of reports on each stage of the collection process. To facilitate this oversight the authors developed a dashboard system to provide managers one central location from which to retrieve these reports.

Reports were identified for comparison across different energy data collection programs to address the survey business processes providing data on frame management, response rates, and rates of edits triggered during data collection. The focus of this paper will be the development of a subset of reports that address common data needs for frame and cycle management. The development process facilitated understanding of the challenges which occurred while designing and building this dashboard system in SAS 9.2. These challenges included the identification of commonalities and differences in information needs prepared for different survey staff, the development of common definitions for survey terms across reports, and the computation of statistics within this common framework. A future paper will comment on system efficacy, as the system is currently in a developmental phase.

CONCURRENT SESSION III-D:

INCOME

Reconcile This! Exploring a Balance Edit Approach in the Consumer Expenditure Survey

Brandon Kopp, Scott Fricker, and Nhien To (Bureau of Labor Statistics, USA)

The current Consumer Expenditure Quarterly Interview Survey (CEQ) asks an extensive battery of expenditure questions that takes 65 minutes to administer on average. The length of the interview is thought to increase respondents' perceptions of burden and contribute to underreporting. One approach that some survey organizations have used to reduce underreporting is to implement a 'balance edit' in which discrepancies between reported expenditure totals and income/ asset totals are probed to identify and reconcile possible sources of error.

This presentation reports results from a small-scale lab study (n=20) designed to examine the feasibility of implementing a balance edit in a modified CEQ survey. Study participants were administered global expenditure questions covering each of the CEQ categories, as well as income, asset, and liability questions to obtain a sense of the household cash flow. When expenditure totals differed significantly from total reported income/assets, participants were given the opportunity to review and revise their answers, and were probed about other potential sources of spending or income as necessary. Respondents' reactions to the survey were assessed in a qualitative debriefing session. The presentation will summarize the results of the study, highlighting factors that affected participants' willingness and ability to engage in the balance edit process, and its effectiveness in identifying and correcting reporting errors.

Explaining Long-term Differences Between Census and BEA Measures of the Income of Persons
Arnold Katz (Bureau of Economic Analysis, USA)

Two of the Department Commerce's featured measures of income growth adjusted for inflation and population change tell different stories about economic performance during past 40 years. The Bureau of Economic Analysis' (BEA's) measure of real personal income per capita increased 123 percent from 1969-2009 while the Census Bureau's measure of real median family income increased only 23 percent during this period. The BEA measure increased 18 percent from 1999 to 2009 while the Census measure declined 5 percent during this period. This paper explains how much of these divergent pictures of economic performance can be attributed to four key sets of factors. The first set relates to conceptual differences between the two measures, such as BEA's use of imputations. The second relates to methodological differences; specifically, BEA's makes extensive use of administrative data while Census uses sample surveys. The third relates to differences caused by changes in the distribution of income. The fourth relates to differences caused by changing demographics including the tendency for more of the income in the economy to be earned by single persons living alone.

The paper examines both cross-sectional and time-series differences, with an emphasis on the latter. Thus, while it notes how much the use of administrative data rather than that obtained by questionnaires may have contributed to the difference between the BEA and Census estimates in a given year, it focuses on whether these differences have systematically increased over time. The paper finds that changes in demographic composition and the distribution of income account for most of the differences between the BEA and Census measures from 1969 to 1999 while conceptual differences account for most of the differences afterwards.

Influential Values and Robust Estimation in the Annual Survey of Public Employment and Payroll
Joseph Barth, Yang Cheng, Mary Mulry, and John Tillinghast (U.S. Census Bureau)

Influential values occur rarely in Governments Division surveys. However, they have a large impact on survey estimates when they do occur. In order to improve the quality of estimates for the Annual Survey of Public Employment and Payroll (ASPEP), we evaluate the impact of Clark Winsorization and M-Estimation on our data in the context of both the basic expansion estimator and the generalized regression estimator. Our evaluation is based on the actual 2009 ASPEP data, as well as simulated samples using the 2002 and 2007 Census of Governments: Employment. We consider optimal Winsorization cutoff values at different levels of aggregation to determine how to minimize the impact of robust methodologies on our estimates. Repeated simulated samples of true Census data will be used to create reliable estimates of the mean squared error of our estimators. Results from the actual 2009 sample data will be used to compare with published 2009 ASPEP results.

Getting More from Survey Income Measures: Empirically-Based Recommendations for Improving Accuracy and Efficiency
John Czajka (Mathematica Policy Research, USA) and Gabrielle Denmead (Consultant, USA)

Recent comparative analyses of income measurement in major household surveys found substantial variation in the sources and distribution of income, and noted areas such as retirement income where survey data lag reality. The comparisons also found similar aggregates of income, earnings and transfers, despite very large differences in the number and detail of survey income questions. These findings suggested a range of opportunities for improvements, and this paper reports on follow-on research related to the Current Population Survey Annual Social and Economic Supplement (CPS ASEC), with recommendations for improving its estimates of income while reducing the overall number of questions.

A key finding is that CPS ASEC interview time and question detail for individual income sources has little relationship to their relative importance in the aggregate or for those receiving that source. For example, earnings represent about 83 percent of total income, yet CPS ASEC earnings questions get detail only on the longest job held during the reference year, not on multiple jobs. Thus they are ill suited to persons who change employment, have multiple employers or businesses, or have intermittent attachment to the labor force. At the

opposite extreme, much of the CPS ASEC income battery is devoted to detail and to sources that, collectively, represent very little income in any part of the income distribution.

A second key finding relates to the dramatic shift in the nature of private retirement systems. Income in the CPS ASEC currently excludes benefits from defined contribution retirement plans such as 401(k)s, or from IRAs, except for insignificant amounts of regular withdrawals, yet these forms of retirement income have been replacing traditional pension benefits for decades. Until the CPS ASEC captures the newer sources of retirement income, estimates of the income and well-being of retirees and the elderly will not be accurate.

CONCURRENT SESSION IV-A:
SMALL AREA ESTIMATION

Establishing a Feasibility of Small Area Estimation from the National Ambulatory Medical Care Survey (NAMCS) Data
Vladislav Beresovsky and Donald Malec (National Center for Health Statistics, USA)

To be able to produce reliable state level estimates from the National Ambulatory Medical Care Survey (NAMCS) data, its design was recently modified and sample size substantially increased. However, small area estimation from stratified multilevel surveys is well known to be challenging because of extreme variability of survey weights and the high level of data clustering. Different methods of estimation may be more or less effective depending on the sample size within small area, sampling fractions, intra cluster correlation and the type of small area variability. We conducted a simulation study using the new NAMCS design to simulate binary outcome variables from two-stage logistic linear mixed models. Final population proportions of the outcome variable in states and associated mean squared error and confidence intervals were estimated using hierarchical and empirical Bayesian methods. Results of the conducted simulation study allow us to specify the conditions when state level estimates are reliable and identify the optimal estimation strategy depending on the characteristics of the sample.

Small Area Estimation for Governments Surveys
Bac Tran and Yang Cheng (U.S. Census Bureau)

In the past three years, we developed decision-based estimation to estimate survey totals for the Annual Survey of Public Employment and Payroll and the Annual Finance Survey. In this paper, we discuss some small area challenges when we estimate survey totals at a function level, such as employees working in airports, public welfare, hospitals, education, etc. First, in this paper we introduce synthetic estimation and modified direct estimation. Then, we modify the composite estimator as a weighted average of the modified direct estimator and synthetic estimator. We also apply the empirical Bayes to estimate the survey total, and then compare it to the modified composite using 2007 Census of Governments data. Finally, we study the methods of estimating the mean square error for the small area.

Small Area Prediction for a Unit Level Lognormal Model
Emily Berg (National Agricultural Statistics Service, USA), Hukum Chandra (Indian Agricultural Statistics Research Institute, India), and Ray Chambers (University of Wollongong, Australia)

When sample sizes for estimation domains are too small to support reliable direct estimators, model based small area estimation (SAE) methods are often used. Many variables of interest in business and agricultural surveys have skewed distributions. An example from the National Agricultural Statistics Service is the number of acres harvested for a particular crop at the county level. Standard small area estimators based on linear mixed models with constant variances can be inefficient for skewed data. SAE methods for skewed data that satisfy a lognormal model are investigated. A minimum mean squared error (MMSE) predictor for small area means and a corresponding mean squared error (MSE) estimator are proposed. Estimators of the MMSE predictor and alternative predictors as well as MSE estimators are evaluated through simulation studies. Empirical results show that the proposed predictor can lead to efficiency gains for skewed data.

CONCURRENT SESSION IV-B:
MULTI-MODE AND WEB DATA COLLECTION

Evaluating Data Collection Mode Options in the 2010 National Survey of College Graduates
John Finamore (National Science Foundation, USA)

In recent years, the use of a web-based self-administered questionnaire (SAQ) over the Internet has become an increasingly popular mode of data collection. While research was conducted as part of the 1993 National

Survey of College Graduates (NSCG) to examine the impact on response associated with the mail SAQ and telephone data collection modes, no research has been conducted in recent years and no research has been conducted to examine the impact the option of a web SAQ has on the NSCG. In response to these research deficiencies, the National Science Foundation and the Census Bureau included a mode effects study as part of the 2010 NSCG to examine the impact data collection modes has on response and cost.

In the 2010 NSCG data collection effort, we offered three data collection modes (mail SAQ, Computer Assisted Telephone Interview (CATI), and web SAQ). These modes were first offered on an individual basis to three randomly assigned treatment groups to examine the impact a single-mode data collection approach has on survey response. Then, after an eight-week period, each single-mode treatment group was offered the other two data collection modes in a sequential manner. This approach allowed us to evaluate numerous data collection mode combinations with the goal of identifying a data collection methodology that produced the best balance of increased response, increased data quality, and reduced cost. This paper includes results from the eight-week single-mode evaluation period and results from the full survey cycle where treatment groups were offered all three data collection modes. In addition, the paper discusses how the results from this evaluation will be used to help determine the data collection strategy for the 2012 NSCG survey cycle.

Developing a Multi-Mode, Longitudinal Study to Understand College Student Outcomes Using Becker's Human Capital Framework

Matthew Soldner and Tracy Hunt-White (National Center for Education Statistics, USA) and Jennifer Wine and Natasha Janson (RTI International, USA)

A key research purpose of the Beginning Postsecondary Students Longitudinal Study (BPS), sponsored by the National Center for Education Statistics (NCES), is to understand what student and institutional factors are related to college completion. Beginning in 2010, NCES and its data collection contractor for BPS, RTI International (RTI) engaged a panel of subject-matter and survey research experts in the retooling of BPS, guided by an extensible and respected econometric model of education decision-making: Becker's Human Capital (HC) model. Although information about the entire scope of the redesign will be presented, we focus on two challenges unique to ED's first-ever implementation of a nationally-representative, HC-informed survey using methods amenable to phone and Web administration

First, the study sought to parsimoniously measure respondents' discount rates (that is, the present value of future money) free of confounding priming effects. Based on preliminary findings from work with students in cognitive labs, field-test experiments in which we systematically varied time horizons (e.g., delayed receipt in six months vs. one year) and base amounts (e.g., receipts of $750, $1125, or $1500) were developed. Those results, and their implications for practice, will be presented.

Second, the study sought to employ web-based Visual Analog Scales (VASs)to measure a wide range of constructs where increased response variability was deemed desirable, such as perceived likelihood of completing college. Based on preliminary cognitive lab results, additional field test experiments were designed to ascertain whether the initial placement of the sliding button used by respondents to indicate their response along the VAS was related to item means or response variability. Findings, and their consequences for the use of VASs in web-based administrations, will be shared.

Considering a Digital 2020 Census

Michael Thieme (U.S. Census Bureau)

As we look forward to the U.S. 2020 Census, we are faced with an unprecedented opportunity to leverage the American society's increasing willingness to incorporate technology into their lives in ways we could not have anticipated only a few years ago. With this in mind, this paper reviews the effectiveness of continuing to use paper for data collection in a census environment.

Although the 2010 Census originally planned to capture data both on paper (through mail out/mail back operations) and digitally (through the Internet and handheld devices), in the end, we captured virtually all respondent data using paper. The initial plans for the 2020 Census propose a number of different combinations of technologies for data collection, including paper, Internet, and administrative records, with the intent to give respondents maximum flexibility, encourage self-response, and reduce costs. While this is an admirable goal, the increase in available modes of response also comes with increased complexity, as each of the options will require a greater level of integration between groups and operations.

Though a complex process is not inherently "good" or "bad", overly complex processes can unnecessarily increase costs, lower agility, cause unnecessary confusion among respondents, and present a greater level of

risk. We are faced with an opportunity here at the beginning of 2020 planning cycle to make a decisive move in the direction of simplicity. Though such a move will bring the Census Bureau out of its familiar comfort zone, it is the position of this paper that the Census Bureau should consider performing a virtually "All Digital" response Census and do away with the use of paper for data response in 2020.

CONCURRENT SESSION IV-C:

REDUCING MEASUREMENT ERROR

Exploring Underreporting and Respondent Records Usage in the Consumer Expenditure Survey
Neil Tseng, Brandon Kopp, Janel Brattland, and Jeanette Davis (Bureau of Labor Statistics, USA), Emily Geisen (RTI International, USA), and M. Christopher Stringer (U.S. Census Bureau)

The Consumer Expenditure Survey (CEQ) presents a number of challenges for both interviewers and respondents. The interview is long, the questions detailed, and the experience can be perceived as burdensome. In part, because of these challenges, there is a widespread belief that some CEQ data are underreported. Underreporting has been variously attributed to recall error, panel conditioning, respondent fatigue, and other causes. One common suggestion to combat underreporting is to increase the use of respondent records, having respondents refer to external sources such as receipts or other documentation to aid their response process.

This qualitative study was developed as part of a comprehensive and ongoing effort to explore measurement error and gain insight into the experience of a CEQ respondent. An earlier field study revealed discrepancies in the availability, type and quality of respondent records, as well as respondent perceptions of the interview experience. However, the previous study was too small to draw many meaningful conclusions about records and measurement error. The current study, consisting of a sample size of 100 field interviews, seeks to add to the information on these topics. Specifically, this qualitative study has three objectives:

Gain an understanding of the availability respondent records and the extent to which the information in those records aligns with the time frame, level of detail, and categories of expenditures asked in CEQ.

Evaluate the direction and magnitude of measurement error by comparing information on available records to the expenditures reported by the respondent based only on recall.

Develop insight into the respondent's interview experience during the CEQ, including perception of burden, task difficulty, expectations of required accuracy, and reactions to the advance letter and Information Book.

The information gained from this research will inform future research as well as potential survey redesign decisions.

Statistical Issues in the Financial Well-Being of the Aged
Chris Anguelov, Howard Iams, and Patrick Purcell (Social Security Administration, USA)

The project reviews the shift of retirement income sources from the traditional defined benefit pensions to defined contribution pensions and individual accounts. The project presents recent data on the pension participation trend from the 1998, 2004, and 2009 panels of the Census Bureau's Survey of Income and Program Participation and from the Bureau of Labor Statistic's National Compensation Survey of employers. The project assesses the impact of correcting survey reporting error on investment accounts in SIPP data with Social Security Administrative data and compares the results to the National Compensation Survey. Using the Federal Reserve's Survey of Consumer Finance from 2992 through 2007, the project also presents recent data on the trend in the aged population prevalence and the relative wealth in these investment style accounts by major demographic groups. These data sources document the shifting participation and holdings in investment style retirement accounts. The project then reviews recent analyses on the spending of money from these accounts to document that the majority of the funds are withdrawn when mandated by the Internal Revenue tax law. The implications of these trends are discussed for the measurement of the financial well-being of the aged in surveys and for the financial well-being of the aged.

Evaluation of Gross Vacancy Rates From the Decennial Census Versus Current Surveys: Early Findings From Comparisons With the 2010 Census and the 2010 ACS 1-Year Estimates
Arthur Cresce, Jr. (U.S. Census Bureau)

Federal, state, and local agencies, private sector market analysts, homebuilders and real estate agents use data on vacant units to evaluate the overall viability of local housing markets. Record number of foreclosures, high levels of unemployment, and borrowers owing more on their mortgages than their homes are worth continue to depress housing markets evident by increasing proportions of vacant units.

The decennial Census of Population and Housing provides detailed data on the occupancy status of housing units down to the block level. We will explore changes in the total vacancy rates between the 2000 and 2010 censuses and compare the 2010 Census total vacancy, homeowner, and rental vacancy rates to those from the 2010 American Community Survey (ACS) and the 2010 Housing and Vacancy Survey (HVS).

In this paper, we intend to examine the vacancy rates for different geographies, including the nation, states, and selected metropolitan areas, counties, and places. We will highlight the subtle differences in the reference dates, residency rules, and sampling frames among the various data sources which could suggest reasons for the differences in the vacancy rates from the census and current housing surveys. Further, we will compare several vacant statuses (i.e. seasonal units, vacant-for-rent, vacant-for-sale, and other vacant) at the national and state level.

Evaluating an Alternative Data Source for Editing MEPS Drug Prices

Marc Zodet, Steven Hill, and Samuel Zuvekas (Agency for Healthcare Research and Quality, USA)

The Medical Expenditure Panel Survey (MEPS) is a unique, nationally representative, source of micro data on health care use and expenditures, including drugs, for all payers. Household respondents report drugs and the number of times each drug was obtained, while a follow-back survey of pharmacies is the primary source of price data. Data quality is critical, because the MEPS is used for national estimates, behavioral modeling, and policy simulations, including analyses of prescription drugs.

We investigate a potential method of improving data editing. Specifically, one quality control check in the editing process compares the retail unit price (RUP) paid for each drug to its average wholesale unit price (AWUP). The AWUP, despite its name, is a list price (not an average of transaction prices). The AWUP may be subject to manipulation and strategic behavior that could alter the AWUP-RUP relationship, and the correlation between AWUP and RUP is weak for generic drugs. We assess the feasibility of replacing AWUPs with median prices estimated from the MarketScan Outpatient Pharmaceutical Claims database, which contains adjudicated claims for millions of people with private employment-related insurance.

The analysis has three phases. First, we assess the reasonableness of the MarketScan data as a proxy for Medicare payers by comparing the MarketScan median unit prices with median prices calculated from Medicare prescription drug event files, obtained from the CMS. Second, we compare the distribution of prices in the MEPS with those in the MarketScan data. Specifically, for each purchase, we calculate the ratio of the unit price to the median for the drug. Then we compare the distributions of this price dispersion separately for brand name and generic drugs. Third, we use matched MEPS-Medicare validation data to determine thresholds in the ratios that suggest outlier unit prices in the MEPS that require editing.

CONCURRENT SESSION IV-D: TECHNICAL DEMONSTRATIONS

UTILIZING TECHNOLOGY TO IMPROVE PROCESSES, PRODUCTS, AND COMMUNICATION

Cropscape: The New Visualization, Querying, and Dissemination Web Portal for the NASS Cropland Data Layer

Audra Zakzeski (National Agricultural Statistics Service, USA)

For more than a decade the National Agricultural Statistics Service (NASS) has been producing the widely utilized crop specific geospatial land cover classification called the Cropland Data Layer (CDL) detailing more than 100 different crop categories over the United States. In 2009 NASS first debuted the coast to coast CDL and announced intentions to produce a national CDL on an annual basis. To facilitate easier distribution and a more effective way to analyze the CDL for customers across the world, NASS partnered with developers at George Mason University to create a web-based interactive mapping and querying system called Cropscape. With history dating back more than 10 years in some areas Cropscape provides a comprehensive method for analyzing agricultural patterns over a state, county, agricultural district, or any user defined region. Analysis options include the creation of tables detailing pixel counts and acreage totals as well as histograms, pie charts and change analysis tables. An additional function provided by Cropscape is the ability to download the CDL data directly into other mapping software such as ArcGIS Explorer or Google Earth which can be used to create mashups with other data sources. Without the need for expensive mapping software Cropscape will allow anyone with an internet connection and a web browser to explore the agricultural footprint of the United States over many years with just the click of a button.

Visualizing Agricultural Data via the Web
Irwin Anolik (National Agricultural Statistics Service, USA)

This demonstration focuses on different ways that the National Agricultural Statistics Service (NASS) currently disseminates data via the Web, so that data customers can freely and effectively view and analyze NASS data using a web browser. Specifically, we demonstrate web-based solutions that enable viewing, analyzing, and dynamically interacting with summary data at the state and county level.

The demonstration presents concepts and technologies implemented on the NASS web site to enhance the ability to find data of interest, and see structure and patterns inherent in the data.

We display historical data through the use of animated maps and charts, as well as “sparklines” - a term proposed by Edward Tufte for "small, high resolution graphics embedded in a context of words, numbers, images".

We also display current data using dynamic maps and charts, and make use of “small multiples”, galleries, and multivariate column views; all meant to give more context and power to the data customer.

While this demonstration focuses on Agricultural Data, the principles and methods can be applied to other sources of survey and census data.

Testing Production Data Capture Quality
K. Bradley Paxton, Steven Spiwak, Douglass Huang, and James McGarity (ADI, LLC, USA)

In the 2010 U. S. Decennial Census, we at ADI, LLC supplied a subsystem called Production Data Quality (PDQ) that ran during Decennial Response Integration System (DRIS) production data capture and delivered daily data quality results to Census management. The availability of precise, cost-effective and current data quality metrics enabled early detection of pockets of error. The PDQ subsystem sampled 845,000 Census forms from March through September 2010. A feature of PDQ is the easy-to-use dashboard, with which one can instantly see where errors are occurring in the data capture system, and readily drill down step-wise (even to the original image snippet) to determine the root cause of the errors. Using this root cause analysis capability, more than a dozen processing errors were discovered and disclosed that led to timely corrective actions.

The PDQ subsystem employed unique automated independent data capture technology, assisted by a small team of data capture analysts to determine the hard match truth of the sampled production data at the field level. Once this truth is created, the production data was scored in detail using a version of the Jaro-Winkler algorithm as well as a hard match.

Census had set overall data capture quality requirements, called Service Level Agreements or SLAs, which applied to three types of data capture: Optical Character Recognition (OCR), Keying From Image (KFI) and Optical Mark Recognition (OMR). By the use of PDQ results delivered to Census management, it was demonstrated that the SLAs for all three data capture were successfully met.

The talk will include a description of the statistical design of the PDQ subsystem, a review of the associated probability model indicating the efficiency and precision of the test results, and some actual screen shots showing the data dashboard.

Statipedia: A Wiki Workspace for Federal Statistical Staff
Peter Meyer (Bureau of Labor Statistics, USA)

A wiki for the federal statistical staff was proposed at the FCSM conference in 2009. It is now up and running and serves as a workspace and a reference work. We will show how it is used to store definitions; training materials; and information about relevant activities, and methods.

The EPA hosts the new service on its extranet, where it runs on the same software as Wikipedia does.

Its founders invite other federal statisticians and other specialist scientific staff to contact us and join in.

CONCURRENT SESSION V-A:
WEIGHTING COMPLEX SURVEY DATA

Comparing the Selection of One Person per Household to the Selection of All Household Members: Can Less be More?
Vincent Iannacchione and Bonnie Shook-Sa (RTI International, USA)

The decision to select one or all eligible members of a sampled household (HH) is influenced by a number of factors including burden on the HH, cost, and the sampling variance of survey estimates. Design effects quantify the influence of a complex sampling design on the variance of survey estimates. Restricting a sample to one eligible person per sampled HH can have counteracting impacts on design effects. A one-person per HH sample requires subsampling in multi-person HHs which increases the design effects attributable to unequal weighting. Conversely, selecting one person from each sampled HH could reduce the design effects attributable to clustering because the potential intra-household correlation among respondents in the same HH is avoided. If this reduction is larger than the increase caused by subsampling, a one-person per HH sample can achieve the same sampling variance as a multi-person sample with less cost and burden.

We present the results of a simulation study that evaluated the design effects associated with the selection of one person per HH on personal victimization rates based on the 2008 National Crime Victimization Survey (NCVS) which currently selects all persons 12 and older for the survey. We selected replicate samples of one respondent from each HH from the 2008 NCVS public-use database and then compared the design effects associated with victimization rates for a one-person per HH sample to those of the existing multi-person sample. We found that the increase in design effects caused by unequal weighting associated with a one-person sample was significantly greater than the decrease caused by the elimination of intra-household correlation. We discuss the pros and cons of a one-person per HH design, and estimate the number of HHs that would be needed to equalize the sampling variances of the current multi-person sample design and a one-person design.

Methods for Incorporating an Undersampled Cell Phone Frame When Weighting a Dual-Frame Telephone Survey
A. Elizabeth Ormson, Kennon Copeland, Kirk Wolter, and Kathleen Santos (NORC at the University of Chicago, USA) and Stephen Blumberg (National Center for Health Statistics, USA)

In December 2010, the National Center for Health Statistics (NCHS) reported that the percentage of wireless-only households had grown to 26.6%, up from 13.6% in 2007 (Blumberg and Luke, 2010). This finding has substantial implications for any telephone survey designer who wants to interview a representative sample of the population and create unbiased estimates. Until recently, most telephone surveys were fielded using a random-digit-dial (RDD) sample drawn from working banks of only landline telephones.

A dual frame design including RDD samples of both landline and cell telephones requires careful thought as to the weighting approach that both reduces coverage bias and controls variance. Cell telephone samples are more costly to field and therefore are typically undersampled compared to the landline telephone sample. This leads to larger weights in the cell telephone sample compared to weights associated with the landline telephone sample and a resultant increase in variance.

This paper will use the National Survey of Children with Special Health Care Needs (NS-CSHCN) as a basis for considering alternative weighting approaches for a combined landline and cell phone sample. The NS-CSHCN is sponsored by the Maternal and Child Health Bureau and conducted by NCHS. The most recent NS-CSHCN fielded a landline telephone RDD sample during six quarters (Q3/2009 – Q4/2010), supplemented with a cell telephone RDD sample fielded for two quarters (Q3/2010 – Q4/2010). Cases from the cell telephone sample represented 9% of all household completes. This dual frame design resulted in different sampling rates from the landline and cell telephone frames. Moreover, sampling rates from the two frames differed across states.

In this paper, we discuss various approaches considered for weighting the NS-CSHCN sample and the resultant impact on bias, variance, and mean squared error of estimates at both the state and national level.

Longitudinal Survey Weight Calibration Applied to the NSF Survey of Doctorate Recipients
Michael Larsen, Siyu Qing, Beilei Zhou, and Mary Foulkes (The George Washington University, USA)

The National Science Foundation's Survey of Doctorate Recipients is conducted every two or three years and collects detailed information on individuals receiving PhDs in science and engineering in the U.S. and some others with PhDs from abroad in these areas. Survey weights adjust for oversampling and nonresponse on a cross sectional basis. A significant portion of the sample (e.g., 60% on 3 or more surveys from 1993-2006) appears in multiple survey years and can be linked across time. No longitudinal weight exists that would enable esti-

mation of statistical models or comparison of finite population characteristics using data from multiple survey waves together. This paper applies calibration estimation for construction of such a longitudinal weight for this survey. Previous results studied the process of weight construction through simulation. Here we report on applications to NSF survey data. Choices of initial weights and multivariate calibration targets are compared in a series of analyses.

Methods for Adjusting Survey Weights When Estimating a Total

Richard Valliant (University of Michigan) and Kimberly Henry (Internal Revenue Service, USA),

Several design-based, model-based, and model-assisted methods have been developed to adjust survey weights for nonresponse or coverage errors, to reduce variances through the use of auxiliary data or by restricting the range of the weights themselves. Some methods directly change the weights, like nonresponse adjustments, calibration weighting, and design-based ad hoc weight trimming methods. Other methods, like the robust superpopulation modeling approaches, implicitly adjust the associated case weights. Another alternative is a newly developed method called generalized design-based inference involves modeling the weights as a function of the survey response variables and using the smoothed weights predicted from the model to estimate finite population totals.

This paper first provides a review of the alternative weighting adjustment methods proposed in the literature. We then illustrate the effect of applying several example methods on complex survey weights and discuss the implications for sample-based inference.

CONCURRENT SESSION V-B:

EVALUATING THE SIPP RE-ENGINEERING, COMPARING SIPP AND SIPP-EHC

A Preliminary Evaluation of the Residence History Data in the 2010 and 2011 Field Tests of the Re-Engineered Survey of Income and Program Participation

Matthew Marlay and Peter Mateyka (U.S. Census Bureau)

The Survey of Income and Program Participation (SIPP) collects longitudinal data on a variety of topics including employment and program participation. The strength of SIPP is that respondents are followed for two to four year periods, even when they change residences, enabling researchers to perform analyses not possible with cross-sectional Census surveys. The longitudinal design of SIPP means collecting accurate residence history information is vital to the success of the survey.

The Bureau is redesigning the SIPP survey instrument, and this redesign will have implications for how the survey collects residence history information. The current in production SIPP has a respondent recall period of four months and respondents are only allowed one move during each four month period, or wave. The Re-Engineered SIPP instrument uses an Event History Calendar (EHC) format with a respondent recall period of one year. Respondents are allowed to report living in up to five residences during the reference year of the survey and detailed address information is collected for each residence reported.

We analyze residence history data collected in the 2010 and 2011 field tests of the Re-Engineered Survey of Income and Program Participation (Re-SIPP). We use two methods to evaluate the quality of the Re-SIPP residence history data. First, we match the addresses collected in the 2010 and 2011 field tests of Re-SIPP to a national address database. We report the percentage of addresses successfully matched at the place, county, and state level. Second, we compare the monthly mover rates reported in the test 2011 Re-SIPP (covering calendar year 2010) to the 2010 Re-SIPP (covering calendar year 2009) to the 2008 SIPP (covering calendar year 2009). We primarily focus on whether there are time trends in the likelihood of respondents to report moving in the two Re-SIPP field tests, indicating memory decay or misreporting.

Measuring School Enrollment in the 2011 Re-SIPP Field Test

Stephanie Ewert and Sarah Crissey (U.S. Census Bureau)

Accurate data on school enrollment is important to help scholars, educators and policy makers understand how students progress through the educational system. The Re-Engineered Survey of Income and Program Participation (Re-SIPP) captures monthly longitudinal school enrollment information for children and adults using Event History Calendar (EHC) interviewing techniques to improve recall over a longer period. School enrollment is unique among EHC topics because respondents enrolled throughout an entire calendar year should have breaks in attendance for summer vacation or grade level changes.

With the increase from a 4-month to over 12-month reference period from the 2008 SIPP Panel to Re-SIPP, we included tools to reduce underreporting of attendance breaks and grade changes. The first was an instrument

check for full year spells that instructed interviewers to change spell dates for attendance breaks or grade changes. The second was a grade repetition item for respondents who reported enrollment in the same grade across a calendar year.

To assess measurement of school enrollment in the 2011 Re-SIPP Field Test compared to SIPP 2008, we first compare monthly school enrollment rates, particularly focusing on summer months when enrollment rates should be lower. We also examine enrollee characteristics such as grade level, school type, part-time status, and age in Re-SIPP and SIPP 2008 to determine whether observed differences in enrollment rates between surveys result from the responses of particular subgroups.

Next, we compare rates of suggested grade repetition in each survey by looking at respondents with the same level of enrollment across a calendar year. Using the Re-SIPP data, we then compare respondents who confirmed that they repeated a grade to those whose data suggested grade repetition to examine the usefulness of the grade repetition item. Finally, we explore variation in the characteristics of students who repeated a grade.

"I Don't Remember": Effects of Recall Period on Reported Job and Program Participation Duration

Rebecca Chenevert and Renee Ellis (U.S. Census Bureau)

In 2006 an initiative was taken to re-design the Survey of Income and Program Participation (SIPP), a major longitudinal survey. This was done in order to reduce costs and respondent burden as well as improve data collection. As part of this re-design effort, the reference period was extended from four months to one year. A major concern in making this change was that the longer recall period would result in fewer reports of spells of program participation and employment, especially during the early months of the recall period. Several studies are ongoing to evaluate the effectiveness of the re-design; this paper seeks to study the effect of the recall period on reported job duration, WIC spells and SNAP (Food Stamps) spells. As a result of testing the new survey instrument, an experiment has been created where households were randomly assigned to the current production version of the survey or the re-designed survey. We will exploit the experimental nature of this design to look at the effect of the recall period on the reported job duration and program participation. We will also examine differences by several demographic characteristics, such as age, sex, race and family composition. We use a proportional hazards model to estimate whether the recall period influences reported duration, and we find that there is very little evidence of differences in recall over the two periods.

An Evaluation of Field Test Data From Re-engineered SIPP using Administrative Records and 2008 SIPP

Graton Gathright, Martha Stinson, and Lori Reeder (U.S. Census Bureau)

The U.S. Census Bureau is re-engineering the Survey of Income and Program Participation (SIPP). Multiple field tests of the new re-engineered SIPP (SIPP-EHC) survey have now been conducted. In this paper, we evaluate the quality of data from two SIPP-EHC field tests. For the 2010 and 2011 field tests and using 2008 SIPP as a benchmark, we evaluate rates of reported employment and program participation and distributions of reported earnings and program receipt across the two surveys. For the 2010 SIPP-EHC field test and using linked administrative records on employment and program participation (SNAP, TANF, Medicare, Medicaid, OASDI, and SSI), we evaluate the accuracy of survey reports at the person-level including the accuracy of reported transitions and the possibility of memory decay in SIPP-EHC.

CONCURRENT SESSION V-C:

MONITORING AND USING PARADATA AND METADATA

Designing a Paradata Application in a CAPI Environment

Barbara O'Hare and Matthew Jans (U.S. Census Bureau)

The analysis of continuously-collected survey paradata provides new opportunities to systematically monitor and manage survey cost and data quality during a data collection period. In recent years, many survey organizations have begun to develop paradata-based field management applications to inform decisions while cases are still being worked. This helps to deliver surveys on time and within budget, while retaining data quality. Whether in the context of responsive design or continuous quality improvement methodologists are beginning to share their experiences.

During 2011, the Census Bureau has been exploring and developing the system infrastructure and the analytic tools that will enable better use of real-time paradata in a large-scale CAPI environment. This paper will describe the project, including the challenges of integrating paradata across operational systems, developing useful analytic tools for use by methodologists and practitioners, and introducing the paradata application to a

CAPI field force. The purpose of this presentation is to elucidate operational and analytic challenges that might be shared across survey organizations.

We will discuss the issues of database access, data structures, units of analysis and merging data across systems that we have experienced integrating data from case management, case characteristics, and financial systems to create a system analyzable for survey management with paradata. Many decisions in this process had important implications for the paradata analyses.

We will highlight our decision process in developing descriptive charts and graphs that are useful in a CAPI field environment. The propensity and productivity models that we developed will be explained.

Last, we will discuss the particular challenges of introducing responsive survey design in a large-scale CAPI field environment where interviewer hours and geography present logistical limitations. Experiences in bridging research methodologists' and field operation managers' perspectives will also be shared.

Using Paradata to Monitor Survey Production, Cost, and Quality Within an Adaptive Total Design Framework
David Cunningham Hunter, Susan Mitchell, and Lisa Carley-Baxter (RTI International, USA)

Adaptive Total Design (ATD) aims to improve the use of paradata for monitoring survey production, cost, and quality. ATD: (a) identifies major contributors to total survey error; (b) provides means for monitoring costs and errors during implementation; (c) allocates resources to maximally control errors and costs; and (d) monitors results of embedded experiments to inform later phases or survey rounds. The paradata used include record of calls data, cost and productivity data, sample frame and sample information, data quality measures, interviewer characteristics, and interviewer observations.

A cross-disciplinary team within RTI has created an ATD approach that collects a universal set of paradata across surveys. This approach allows for conservation of cost by minimizing the need for custom programming and allows for comparability across surveys. An ATD "dashboard" presents the paradata for review and evaluation by project staff members. This paper will present highlights of RTI's ATD approach, considerations when choosing which paradata to examine and how to best utilize them, lessons learned during dashboard implementation, and practical considerations for developing and implementing an ATD approach including the education of end users.

Producing Control Charts to Monitor Response Rates for Business Surveys In the Economic Directorate of the U.S. Census Bureau
Yarissa Gonzalez and Broderick Oliver (U.S. Census Bureau)

Control charts are powerful tools for monitoring ongoing processes, providing evidence of the process's stability and capability. We applied control chart techniques to the unit response rate (URR), a commonly used business survey quality indicator. At the U.S. Census Bureau this measure is computed as part of survey processing and is subject to targeted benchmark. Control charts will provide program managers a broader view of the process and will help them to assess if their current procedures are capable of meeting this mandated benchmark.

The usage of control charts originated in a manufacturing setting, where processes are repeatable and continuous, and data about the product being manufactured are collected frequently. In contrast, survey response processes have a human element, where measurement errors arise from both the respondents and the inspecting analysts. Survey response processes may also suffer from small sample sizes within a given domain, making it difficult to produce reliable control limits. In addition, the response rates are produced less frequently (e.g., monthly, quarterly, annually) making it more challenging to determine a reliable process average.

This paper presents our recommended methods for producing a p-chart for the URR. We employed simulation studies to address the challenges pertaining to this particular response indicator. Our objective was to find methods that could produce control limits that were reasonably precise and relatively easy to calculate.

A Web-Based Approach for Combining Metadata, Search, and Data Profiling
Jeff Butler (Internal Revenue Service, USA)

With nearly one petabyte of data serving hundreds of IRS research analysts, the Compliance Data Warehouse (CDW) brings together data and analytics for the world's largest tax agency on a massive scale. Metadata are available for more than 25,000 unique database columns and over 500,000 separate attributes through the CDW website and represents the largest database-driven repository of metadata in the IRS. Data profiling capabilities are also available through the CDW website, letting users quickly explore patterns through frequency tables, statistical distributions, trends, and geographic maps—often on billions of rows of data.

This paper discusses a web-based approach used by CDW for combining metadata, search, and data profiling into a single experience. Rules for standardizing metadata are outlined—including column definition templates, lookup reference tables, data types, and other artifacts. A database-oriented format is proposed for organizing and displaying search results for data stored in relational databases. Finally, a solution is presented for adding data profiling capabilities to both metadata and search results to let users dynamically explore and visualize patterns in data using a set of flexible ad-hoc database queries. Since it was implemented in 2009, this dynamic data profiling capability now accounts for roughly one-third of the 3,000 average daily database queries executed in CDW.

CONCURRENT SESSION V-D:

STATISTICAL METHODS

An Exploratory Components Procedure to Uncover the Q Matrix in Cognitive Diagnostic Testing

Catherine Close, Mark Davison, and Ernest Davenport, Jr. (University of Minnesota, USA)

Cognitive diagnostic assessment (CDA) involves diagnosing students' academic strengths and weaknesses for purposes of individualizing instruction. CDA combines both cognitive psychology and psychometrics. For application of CDA models to test items, one must hypothesize a Q matrix describing the required skills and the items requiring each skill. Factor (or components) analysis has been suggested as an exploratory method that could supplement theory in the derivation of a Q matrix for a set of items, but there is little description of exactly how factor analysis could be applied. This paper will evaluate the application of factor/components analysis to derivation of a Q matrix. First, the components representation will be presented for items which satisfy any one of three common CDA models: the reduced RUM (Reparameterized Unified Model), NIDA (Noisy Input Deterministic And), and DINA (Deterministic Input Noisy And). The dimensions in the Q-matrix correspond to single skills necessary, but generally not sufficient, for solution of items requiring the skill. The components in the component representation correspond to skill sets composed of one or more skills such that the skill set is necessary and sufficient for solution of items loading on the component. While components analysis is not ideal, since there is no one-to-one relationship between skills and components, nevertheless the relationship between skills and components suggests a components based analysis to recover Q matrix skills. The analysis consists of three steps: a components analysis of items with rotation to simple structure; identification of the skill set associated with each component; and construction of the Q matrix from the skills that appear in the component skill sets. Illustrations with real (Tatsuoka's) and simulated data will be presented. Alternative approaches involving linear and nonlinear factor solutions will be discussed. Using the analysis to enhance reporting of assessment results will also be discussed.

The Multivariate Differential Effects Value-Added Model

Robert Meyer and Emin Dokumaci (University of Wisconsin-Madison, USA)

A value-added model (VAM) is a quasi-experimental model that yields estimates of the contribution of educational units to student achievement, controlling for non-school sources of student achievement, including prior student achievement, student characteristics, and (typically) latent student growth trajectories. The objective is to facilitate valid and fair comparisons of productivity with respect to student outcomes, given that schools may serve very different student populations. The conventional value-added model imposes the restriction that a high-performing educational unit is identically high performing for all types of students. If this assumption is approximately true, teachers can validly be compared on the basis of a single performance indicator. In fact, a given teacher could be very effective for students with low prior achievement, for example, but less so with high achieving students.

We present a generalized value-added model, which we refer to as a multivariate differential effects VAM. This model captures differences in value-added productivity by student subgroups, defined by prior achievement and student demographic variables. Multivariate shrinkage is used to produce effect estimates with minimum mean squared error and to eliminate differential effects in cases where the variance in differential effects (with respect to a given subgroup) is essentially zero. A major challenge in estimating the model is that differential effects for a given subgroup are estimated with limited precision for educational units with limited (or no) variance in the subgroup. For example, it is not possible to estimate a differential effect with respect to student poverty if all students in a classroom are all poor (or all non-poor). We explore several different estimation strategies, including maximum likelihood estimation, for estimating model parameters and the variance-covariance of differential effects (required for multivariate shrinkage estimation). We illustrate the utility of these models using data from Chicago, Los Angeles, Milwaukee, and New York City.

Genuine Exact Two-Stage Methodologies for Producing Assigned Accuracy Estimators for a Gamma Mean

Kevin Tolliver (U.S. Census Bureau)

When sampling from largely right-skewed populations it is better to assume the population is Gamma rather than Normal. The Gamma distribution is often assumed when modeling mean-time-to-failure in the biological field survival analysis and the engineering field reliability analysis. The sequential methods in this paper are used to determine what sample size is required to attain an accurate estimator assuming the data comes from a Gamma population. This paper proposes two methods for finding estimators with pre-assigned accuracy: (1) point estimator and (2) an interval estimator. It implements a genuine two-stage sampling procedure. The term genuine refers to the fact that, in contrast to previous methods, the procedures proposed herein are based on the combined samples from both stages, rather than ignoring the data from the first-stage sample. All the results are exact. At no point was an asymptotic or large sampling approximation used and all the derivations assumed an underlying distribution of Gamma. Results are found for when shape is both known and unknown.

Quantification of Profile Matches

Ernest Davenport, Jr. and Mark Davison (University of Minnesota, USA)

The use of profiles to determine disease states, apprehend criminals, identity successful candidates for school or work, etc. has been a staple in many clinical and applied fields (e.g. medicine, psychology, college admissions, human resources). Regardless of discipline, multiple measures are taken and the pattern of scores on these measures are compared to a prototypical profile, where prototypical profiles show the pattern of scores desired for a particular clinical diagnosis. While the measures provide quantitative data, the final decision is often based on expert opinion regarding the closeness of an individual's pattern of scores to that of the prototypical profile. Thus, the final decision is oftentimes as much art as objective. This paper provides a procedure to quantify the match of score information for an individual with a prototypical profile. The procedure need not supplant clinical judgment, but can be used as one of the elements that enter into the clinician's decision.

There has been much activity since the 1950's to quantify score profiles. Most of the earlier methods may not have criterion-related validity. Indeed, cluster analysis, modal profile analysis, and profile analysis via multidimensional scaling rely solely on the subtest variables within each profile, and no external criterion is used to identify the core profile types. Thus, one cannot be sure the resulting profiles have criterion validity. Davison and Davenport suggest a procedure that quantifies the relationship of individual profiles to an external criterion (prototypical profile). It makes use of Cronbach and Glesser's decomposition of a set of scores into three components: elevation, scatter, and shape. Specifically, an index that combines scatter and shape allows for a criterion match statistic that quantifies a subject score profile with an empirical and/or theoretical prototypical profile. This statistic will be demonstrated across the realms of mathematics course taking patterns and college admissions.

CONCURRENT SESSION VI-A:

VARIANCE ESTIMATION

Can Post-Imputation Calibration Mitigate Possible Bias in the Multiple Imputation Variance Estimator?

Benjamin Reist (U.S. Census Bureau) and Michael Larsen (The George Washington University, USA)

One criticism of multiple imputation for use in sample surveys is that it produces biased variance estimates under any of the following conditions: 1) when key subpopulations are ignored in the imputation model, 2) when survey weights are not used in the imputation model, and 3) when prior distributions other than single observation unbiased priors (SOUPs) are not used in the imputation model. We examine the effect of post-imputation calibration on the bias in variance estimation under these conditions. Calibration weighting and its familiar forms, including raking and post-stratification, are often used in sample surveys to adjust sample estimates to match control total values. Here, we explore possibilities for using calibration weighting in combination with multiple imputation when there are missing data. Methods could apply to both sample survey and more general study design contexts.

Condition Indexes and Variance Decompositions for Diagnosing Multicollinearity in Linear Model and Generalized Linear Model Analysis of Survey Data

Dan Liao (RTI International, USA)

Survey data are often used to fit linear or generalized linear regression models. The values of covariates used in modeling are not controlled as they might be in an experiment. Thus, multicollinearity among the covariates is an inevitable problem in the analysis of survey data.

Multicollinearities among explanatory variables in linear and generalized linear regression models affect estimates from survey data just as they do in non-survey data. Undesirable effects are unnecessarily inflated standard errors, spuriously low or high t-statistics, and parameter estimates with illogical signs. However, the available multicollinearity diagnostics are not generally appropriate for survey data because the variance estimators they incorporate do not properly account for stratification, clustering, and survey weights.

In this study, we derive condition indexes and variance decompositions to diagnose multicollinearity problems in complex survey data. The adapted diagnostics are illustrated with data based on a survey of health characteristics.

Within PSU Sort Research to Reduce Variances for the Survey of Income and Program Participation (SIPP)

Sarah Tekansik and Stephen Mack (U.S. Census Bureau)

Every ten years, following the Decennial Census, the Census Bureau demographic surveys redesign their sample selection process to determine ways to improve each survey's efficiency. One of the areas of research is the sort used prior to selecting a systematic sample of households within primary sampling units (PSUs). This research attempted to find the optimal sort scheme for the SIPP that would minimize the within PSU variance of an estimate of persons in poverty using a method of successive differences. Also it addressed some of the new challenges that came with this redesign. These included using the American Community Survey (ACS) for possible sort variables and using the Master Address File (MAF) as a frame instead of the Decennial Census.

Weighting and Variance Estimation Under Responsive Designs and Related Forms of Sample-Driven Resource Allocation for Complex Surveys

John Eltinge (Bureau of Labor Statistics, USA)

With some notable exceptions, formal inference from complex sample data has traditionally been based on the premise that randomization mechanisms for resource allocation are determined a priori. This resource allocation may include several components, including sample selection; assignment of collection mode and interviewers; nonresponse follow-up; and re-interviews to evaluate measurement error properties. In practice, however, field-level decisions in the abovementioned areas often involve decisions that are not entirely determined a priori, and that may be based on paradata or other preliminary information available on some of the sample units. For the case of nonresponse follow-up, Groves and Heeringa

(2006) and subsequent publications explored some of these decision processes under a "responsive design" framework, and suggested a number of ways in which to use this framework to improve the balance between data quality and cost.

The current paper explores issues of weighting and variance estimation under a "responsive design" framework and related forms of sample-driven resource allocation. It places primary emphasis on methods based on extensions of standard analyses of data collected through two-phase or multi-phase sample designs. Three complementary approaches receive primary attention. These involve methods based on, respectively: (1) a set of randomized resource-allocation rules determined a priori; (2) conditioning on the observed paradata; or (3) integration with respect to the distributions induced by a superpopulation model for the paradata, as well as the randomization distribution induced by the original sample design. A simulation study illustrates some practical distinctions in results obtained through approaches (1) through (3), respectively.

CONCURRENT SESSION VI-B:

RESPONSE AND NONRESPONSE IN ESTABLISHMENT SURVEYS

Survey Data Collection Methods in Establishment Surveys

Brad Edwards (Westat, USA) and David DesRoches (Mathematica Policy Research, USA)

Establishments present a number of challenges for survey data collection, including some that are unique to establishments. Who is the respondent (the receptionist, the president, or someone in between)? What is the best mode (web, telephone, interactive voice response, text, fax, mail, face-to-face)? When are multi-mode designs appropriate? What level of the establishment is best for collecting the data? What is the best approach for establishments that are part of chains? Questions like these need careful consideration in developing the study design, and their answers form the basis for the data collection plan and survey management system. This paper will review current data collection practices in establishment surveys, offer some ideas for best practices, and outline future research needs. Examples will be drawn from major federal establishment surveys and from the authors' data collection experiences with commercial enterprises, medical establishments and schools.

The framework for the paper is total survey error (TSE), and specifically nonsampling error. Current practices in data collection mode choice (including multi-modes), in respondent selection, in minimizing non-response, and in integration of administrative data with interview data will be discussed in the TSE context. Tradeoffs between quality and cost and between different error sources (e.g., interviewer effects, mode effects) will be addressed.

Response burden is a major consideration in establishment surveys. How can we measure perceived burden as well as time or resource burden? What is the relationship between perceived burden and survey participation, and how does it vary by establishment type or size? What methods are available to control response burden?

Although nonsampling errors in cross-sectional national surveys of U.S. establishments will be the focus, the paper will include some discussion of comparative error in multi-national studies of establishments and conditioning error and other time-related errors in longitudinal studies.

A Comparison of Establishment Collection Mailing Methods

Chris Ellis (RTI International, USA), Margaret Noonan, Tim Flanigan, and Scott Ginder (Bureau of Justice Statistics, USA), and Hope Smiley McDonald (RTI International, USA)

The Deaths in Custody Reporting Program (DCRP), the primary vehicle for collecting mortality statistics among inmates within correctional facilities, began full scale data collection in 2001. Authorized by Congress and funded by the Bureau of Justice Statistics (BJS), the DCRP collects data on deaths in state prisons and correctional facilities operated by local governments. It is the single, unified collection in the United States that captures the circumstances of death for inmates who died while in custody.

RTI International, who began serving as the data collection agent for DCRP in 2009, implemented several procedural changes in order to further enhance the response rates and data quality of the study. One such change, drawn from the best practices in survey methods, was to send all forms packages to agency respondents via FedEx. The cumulative results of the 2009 protocol were a 100% response rate among state Departments of Corrections (DOCs) and a 95% response rate among local jail jurisdictions. Anecdotal feedback from agency contacts, however, indicated FedEx being perceived as excessive for a study a) already known to them, and b) being conducted in a time of national economic upheaval.

As a result, RTI and BJS embed a methodological experiment within the 2010 annual forms mailing to test the variable effects of different mailing types. Namely, two new treatments (USPS Priority Mail and USPS 1st Class) were tested alongside of the control group of FedEx, the current method. Our paper will examine the results of the experiment, including timing and rate of response associated with different mailing options (e.g., FedEx) in the context of a longitudinal study of establishments. Additionally, the costs associated with the mailing types will be compared, including any savings achieved as a result of early responders who did not require one or more nonresponse prompts.

Does it Pay to Try Again? Using Production Metrics From the Recruiting Process on an Establishment Survey to Design Recruiting Protocols

Sara Zuckerbraun, John Loft, and Patricia LeBaron (RTI International, USA) and Manisha Sengupta (National Center for Health Statistics, USA)

Survey researchers conducting establishment surveys face challenges in gaining cooperation. There are various industry-accepted methods to address these challenges, e.g., endorsement letters, financial incentives, explaining the societal benefits derived from the survey data, and lowering respondent burden. Our presentation assumes that survey researchers are familiar with and implement the industry-accepted strategies for gaining cooperation which are most appropriate for their surveys. We focus instead on two key production metrics within the recruiting process. Our research question asks what effect these may have on the establishment's likelihood of participating.

Background: Our data come from the recent National Survey of Residential Care Facilities (sample size = 2,664 establishments). Field staff needed to telephone the director, explain the study, and complete a subsequent in-person multi-hour interview. Our metrics are: 1) the number of contacts required until the establishment reached its final decision (i.e., whether or not to do the interview), and 2) elapsed calendar time from when the case was first contacted until the establishment reached its final decision. A new contact is counted each time field staff contacted a case, e.g., sent a letter, spoke to the establishment, or left voicemail. Methods: We examine the distribution, mean, and standard deviation of time and number of contacts required to reach particular response rates. We analyze how these statistics vary by different outcomes (unreachable, considering whether to do the interview, refused, completed interview) and by sample characteristics (e.g., establishment size, region, metropolitan statistical area), and we consider whether a "cut point" number of contacts, types of contacts, or elapsed time exists after which the likelihood of obtaining participation approaches zero. The conclusions reached may be useful to researchers planning the schedule, budget, and recruiting protocol for gaining participation and for reducing the potential for non-response bias on establishment surveys.

Developing Response Metrics for the Economic Census

Joanna Fane Lineback, Broderick Oliver, and Diane Willimack (U.S. Census Bureau)

The U.S. Census Bureau conducts the Economic Census every five years. The Economic Census is a survey of U.S. business establishments in the eight major business trade areas: manufacturing, construction, mining, retail, services, wholesale, finance-insurance-real estate (FIRE), and utilities-transportation. Check-in has long been the proxy indicator of response for the Economic Census, where a valid check-in indicates a mailed form was returned, but not necessarily that item level information was received. Efforts are under way at the Census Bureau to develop additional response metrics for the Economic Census, including unit and item response rates. The reasons for developing new metrics are three-fold: to comply with Census Bureau and Office of Management and Budget standards, to produce response metrics similar to those used by other economic survey programs, and to produce more meaningful information about the "quality" of Economic Census data. The Economic Census, however, is not like other survey programs with respect to design or magnitude of information collected. Therefore, introducing new methodology introduces challenges that are noteworthy to methodologists and users of the data. We discuss the Economic Census survey design and our approach to developing response metrics, highlighting both our challenges and successes thus far.

CONCURRENT SESSION VI-C:

MEASURING, MONITORING, AND IMPROVING THE QUALITY OF SURVEY DATA

Survey Quality Indicator Measures

Donsig Jang (Mathematica Policy Research, USA) and Flora Lan (National Science Foundation, USA

Response rate is often used as an indication of measuring the quality of the survey response. However, it only tells one side of the survey story; the other side about the association between respondents and non-respondents are unknown. Researchers continue to seek the tools to assess and compare the quality of the response to different surveys. Introduced by Schouten et al. (2009), R-indicators are used to measure how well a respondent set represents the sample or population from which it was drawn. This measure may be a better indicator of survey nonresponse bias than response rates for survey outcomes closely related to auxiliary variables used for R-indicator calculation. In this study, we will examine R-indicators as a possible alternative method in measuring survey quality.

Using the data from the 2008 National Survey of Recent College Graduates (NSRCG), we will examine R-indicator and response-rate trends over a six-month data collection period to observe the relationship between the two measures. To understand the relationship between the two measures and potential nonresponse bias, we will examine any linkages between trends in key survey estimates during data collection and the two survey quality measures. Specifically, each week during the last eight weeks of the study, we will produce survey estimates that only include sample members who have responded by that week. We will then compare these estimates with the estimates from all sample members.

Through this empirical investigation, we expect to identify the relationship between survey quality measures and nonresponse bias. Our findings will help us to identify an alternative quality measures to use for next round of the NSRCG and to control the survey process in such a way that a representative sample is obtained and decide when to end the data collection period.

A Quality Control Program for NASS' National Operations Center

Jeffrey Boone and Joseph Parsons (National Agricultural Statistics Service, USA)

The National Agricultural Statistics Service (NASS) is centralizing much of its survey data collection activities at its new National Operations Center (NOC) beginning in August 2011. The NOC Division will be responsible for telephone data collection, the processing of paper questionnaires, the maintenance of NASS' list frame, and other survey related activities. The motivation for the creation of the NOC was to reduce the source of error inherent in data collection activities, improve data quality, and reduce operational costs. The business case for the NOC articulated the need for a comprehensive quality control program at the new center to drive performance excellence. This paper examines the current state of the quality control program at the NOC, including metrics and monitoring methods. The process of implementing quality control measures into various functions is discussed as well as future opportunities for enhancement and lessons learned.

Quality Assurance for EPA's National Coastal Survey

Marla Smith, Sarah Lehmann, and Treda Grayson (U.S. Environmental Protection Agency)

The U.S. Environmental Protection Agency maintains a Quality System to ensure that its environmental data have scientific integrity. This presentation will describe EPA's experiences and challenges in defining and monitoring quality during the National Coastal Condition Assessment survey. The survey is a coordinated effort

among the EPA, the National Oceanic and Atmospheric Administration (NOAA), the U.S. Geological Survey, the U.S. Fish and Wildlife Service, coastal states, and the National Estuary Programs. During the summer of 2010, EPA and its partner teams sampled more than 1300 sites within U.S. coastal waters and Great Lakes. Quality was considered at every phase of the survey and started with the creation of a Quality Assurance Project Plan. The QAPP includes requirements for field sampling, laboratory analysis, auditing, database development, and the data analysis.

Achieving Information Quality via Continuous Quality Improvement

Shawna Waugh (Energy Information Administration, USA)

EIA's Petroleum and Biofuels Program has effectively utilized the process of continuous quality improvement to achieve information quality in our energy programs. Data from surveys in this program are integrated into several agency publications, the Monthly Energy Review, the Annual Energy Review and other web publications. These data are utilized by internal and external data users and stakeholders to assess aspects of the petroleum industry, including the relationship between demand and price, and the differences in geospatial trends in the United States

EIA publishes relevant and reliable data – weekly, monthly, and annually. We continuously seek compliance with agency and OMB standards, and have introduced lessons learned and best practices to improve the dimensions of information quality, including relevance of the petroleum data for our users. We continuously seek to improve accuracy, timeliness, accessibility, interpretability and coherence of our data. The challenges is finding survey methods capable of balancing cost, timeliness and quality. Some of the best options are technology-driven and costly which make planning and implementation challenging given budget constraints.

One unique aspect of the PBS energy programs is integration of surveys which utilize alternative survey methods – sample and census, alternative validation, quality control and quality assurance methods, and mixed mode data collection for weekly, monthly and annual surveys. These alternative methods are utilized based on balancing the triple constraints.

System thinking and survey research are essential in providing guidance and developing methods for designing a more efficient and effective program, one capable of integrating survey and auxiliary data using an approach which will ensure continued data quality.

CONCURRENT SESSION VI-D:

PROTECTING CONFIDENTIAL DATA AND PREVENTING DISCLOSURE

Applicability of Basic Separability Principles To Enhance the Operational Efficiency of Synthetic Tabular Data Generation Procedures in Multi Dimensional Table Structures

Ramesh Dandekar (Energy Information Administration, USA)

Dandekar2001 proposes using synthetic tabular data generation (i.e. controlled adjustments to tabular data - CTA) as an alternative to complementary cell suppression procedures. Dandekar2009 addresses quality aspects of CTA protected tabular data with an objective to completely replace conventional complementary cell suppression procedures with a new tabular data protection method. The proposed method is a hybrid of three different tabular data protection methods. CTA has already been demonstrated to be equally effective on multi-dimensional counts data and magnitude data containing complex hierarchies and linked table structures.

In this paper we go a step further to demonstrate how the operational efficiency of the CTA procedures could be enhanced significantly by applying basic separability principles to complex table structures. The proposed enhancements have significant potential to cut down on computational resources by reducing the problem size in terms of the number of variables and associated mathematical constraints.

A Stochastic/Deterministic Hybrid Search Method for Solving the 3-Dimensional Cell Suppression Problem for Deeply Hierarchical-Structured Tables

Matt Fetter (National Agricultural Statistics Service, USA)

In an earlier paper (Fetter, 2010), the author presented a stochastic search algorithm for solving the 3-dimensional cell suppression problem for simple hierarchical tables. The current paper describes subsequent development of a deterministic component that can be combined with the stochastic search, providing a method that solves the 3-dimensional cell suppression problem for tables with deep hierarchical row structure. Deeply structured tables (super-tables) can be thought of as a collection of linked, simple hierarchical 'sub-tables'. The stochastic portion of the algorithm is restricted to work within the simple sub-table where the sensitive cell resides. At each iteration, the stochastic search produces a closed, 3-dimensional path of complementary cell suppressions within the sub-table. These paths, however, might not provide any real protection to the sensitive

cell due to the additive nature of the aggregate rows in the other sub-tables. Such paths are considered "not closed" with respect to the super-table.

It is therefore necessary to suppress additional cells in other sub-tables to insure that the candidate solution path is closed with respect to the super-table, meeting the single most important requirement for protecting the sensitive cell. One way to accomplish this is finishing the path by working upward to the super-table's grand total row. This would be expensive in terms of cell value suppressed. The deterministic portion of the algorithm avoids this solution if possible. Instead, it attempts to find a feasible set of additional cell suppressions that has minimum cost by moving downward to the base level in the hierarchy, splitting flow across multiple paths if necessary. The result is a set of 3-dimensional paths that is closed with respect to each sub-table and is also closed with respect to the super-table.

k-Anonymization May Be NP-complete, but Can It Be Practical?

David Wilson (RTI International, USA)

k-Anonymity is a well-established statistical disclosure control method that is designed to ensure that every record in a microdata file is indistinguishable from at least k-1 other records in terms of the values of a set of variables One implementation of k-anonymity seeks to find the minimum number of record level data suppressions that produces k-anonymity. Substantial research has shown that the process of determining the minimum number of suppressions is a NP-complete problem. Given the increase in computational power of modern computers, this paper examines the degree to which k-anonymization is practical by applying the methodology to public-use data from a study conducted by the National Center for Education Statistics.

This paper summarizes the process of 2-anonymizing data from the base-year student-level public-use file of the Education Longitudinal Study of 2002. Nineteen quasi-identifying variables and records for approximately 11,000 students were included in this study.

The impact of 2-anonymization on statistical inference was examined by comparing univariate and multivariate estimates calculated from the original data and the 2-anonymized data. Weighted hot-deck imputation was used to impute the suppressed values while ensuring that imputed values differed from the original values. Estimates using the imputed data set were calculated and compared and contrasted with estimates derived from the original data set and the 2-anonymized data set.

Establishing a Secure Data Center with Remote Access

Jeffrey Gonder and Evan Burton (National Renewable Energy Laboratory, USA) and Elaine Murakami (Federal Highway Administration, USA)

The Transportation Secure Data Center (TSDC) is a newly established center for securely archiving and accessing detailed transportation data. Data containing high-resolution location or other information, such as the growing amount collected with global positioning system (GPS) devices, contains tremendous research value but presents a privacy challenge because the latitude/longitude details can link to a specific address. With support from the Federal Highway Administration, the TSDC enables secure, controlled access to detailed data for legitimate research in a way that preserves participant anonymity.

As host for the TSDC, the National Renewable Energy Laboratory (NREL) created a secure enclave for preserving existing datasets from multiple agencies around the U.S. The datasets, originally collected for particular travel surveys/studies, contain driving profiles for thousands of person-days of travel. NREL processing adds metadata, corrects errors, and prepares the data for two different levels of user access.

For the first level, NREL makes "cleansed" versions of the data freely available for download. Cleansing involves removing geographic positioning and other details that could be used to identify an individual. For many analyses, such as vehicle performance and emissions modeling, the remaining details of trip distance, time, second-by-second vehicle speed, etc. are sufficient.

The TSDC also includes an application process for research requiring access to spatial and other data details. After signing a confidentiality agreement, approved users receive an account for remotely connecting to a restricted virtual environment. The data remains protected behind NREL firewalls and removal is prohibited by disabling functions such as drive sharing, copy/paste and external internet access. The environment includes several software packages for data query, statistical analysis and GIS visualization. To support researchers' analysis efforts, NREL also provides dummy datasets, sample code, and a process through which to bring in externally-developed files and take out aggregated analysis results from the secure environment.

CONCURRENT SESSION VII-A:
LINKING DATA TO ASSESS NONRESPONSE AND MEASUREMENT

Nonresponse Bias Analysis of BMI Data in the ATUS Eating & Health Module
Karen Hamrick (Economic Research Service, USA)

Obesity is the most common food and nutrition-related health problem in America. In order to provide a new avenue for obesity research, ERS developed the Eating & Health Module, a supplement to the Bureau of Labor Statistics American Time Use Survey (ATUS), which was fielded over 2006-08. The Module includes questions on height and weight, such that the respondent's Body Mass Index (BMI) can be analyzed with the ATUS time diary information. Less than 5 percent of respondents did not report height and/or weight resulting in missing BMIs. Despite the high cooperation rate, there was concern that there may be nonresponse bias, that is, bias in the time use estimates due to the missing BMI observations.

This research analyzes the respondents who did not report height and/or weight. Nonresponse bias would occur if those who did not report height and/or weight had different characteristics, and in particular, different time use patterns than those who reported both height and weight. The ATUS and Eating & Health Module data for 2006-08 were used. In addition, paradata—information about the survey process, such as the number of call attempts—were also used to evaluate the quality of the missing BMI respondents' time diaries and questionnaire responses. Descriptive statistics, multivariate analysis, and a dissimilarity index were used to determine whether time use patterns were different for those with missing BMI than for others.

Findings included that respondents who did not report height and/or weight had disproportionately higher indicators of being reluctant or uncooperative respondents. In addition, the time use patterns of those with missing BMIs closely resembled those with Normal weight for men and with those who were Overweight for women, which suggests that any BMI item nonresponse bias is likely to be small.

Estimating Mental Illness in an Ongoing National Survey
Joe Gfroerer, Sarra Hedden, Peggy Barker, and Jonaki Bose (Substance Abuse and Mental Health Services Administration, USA), and Jeremy Aldworth (RTI International, USA)

Up-to-date estimates of the prevalence of mental illness in the population are needed by policymakers and program officials. However, instruments assessing the large number of mental disorders specified in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) take too long and are too complicated to administer within ongoing large-scale multipurpose surveys. To address the need for these data, a new approach was implemented in the 2008 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian non-institutionalized population age 12 and older, conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA). This method allows for the estimation of mental disorders for all sampled adults without directly administering diagnostic clinical interviews to each respondent. Short scales measuring distress and impairment are administered to all adult survey respondents. A small sample of NSDUH respondents (500 to 1500 each year) receives a full diagnostic interview administered by trained clinicians over the telephone. The clinical interview data provide a basis for developing models predicting mental illness, by linking the short scale data with the clinical interview data in the subsample. A number of challenges were encountered during the development of this new mental illness estimation system, including the subsample design; determination of optimal short scales and best model to predict mental illness; weights; identifying methods for estimating prevalence rates, trends, variances; and nonresponse bias. We highlight some of the methodological issues associated with this new estimation system, discuss the pros and cons of approaches that were explored, explain the rationale for the final design, and describe the results. A general discussion of the advantages and limitations of this methodology, compared with alternative methods for estimating mental illness in the general population will be included.

Telling Truth From Y's: Accuracy of Self-Reported Condom Use Assessed by a Semen Y-Chromosome Biomarker for Unprotected Sex
Janet Rosenbaum (University of Maryland, USA), Johan Melendez (Johns Hopkins University, USA), Eve Rose, Gina Wingood, and Ralph DiClemente (Emory University, USA) Jonathan Zenilman (Johns Hopkins University, USA)

Many adolescents use condoms inconsistently but may over-report condom use. Biomarkers offer a potential solution to the over-report, allowing interventions to target adolescents at risk from inconsistent condom use. This study uses semen Y-chromosome biomarker from a vaginal swab, sensitive for 14 days post-coitus with specificity of 92%. Respondents are 715 sexually active African-American adolescent women ages 15-21, not trying to become pregnant, participating in a safe sex intervention in urban Atlanta in 2002-04. Respondents

were surveyed by ACASI at baseline, 6 and 12 months and tested for sexually transmitted infections, pregnancy, and Y-chromosome DNA. Among the 29% of respondents who claimed perfect condom use in the past 14 days at wave 1, 30% tested positive for Y-chromosome, so they are suspected over-reporters of condom use; 20% and 15% of respondents are suspected of over-reporting condom use at waves 2 and 3, respectively. Suspected over-reporters of condom use were more likely to be pregnant at the following wave than any other category of condom users, including never users: 19% of wave 1 suspected over-reporters were pregnant at wave 2, and 16% of wave 2 suspected over-reporters were pregnant at wave 3 (Fisher's exact p=0.001, p=0.08). Suspected over-reporters were more likely to be pregnant than any other category of contraceptive users at last sex, including those reporting no contraception (Fisher's exact p=0.02, p=0.05). The higher pregnancy rate among suspected over-reporters was despite them reporting fewer episodes of coitus, and few never-users reporting use of other contraceptive methods. Many adolescent women in a reproductive health intervention are suspected to have over-reported condom use, and this group may be at particularly high risk for pregnancy.

Linking Children From the National Health and Nutrition Examination Survey to Medicaid Enrollment and Claims Data

Lisa Mirel (National Center for Health Statistics, USA), Gloria Wheatcroft (Center for Medicare and Medicaid Services, USA), and Cordell Golden and Kenneth Schoendorf (National Center for Health Statistics, USA)

The National Health and Nutrition Examination Survey (NHANES) is a nationally representative survey designed to assess the health and nutritional status of adults and children in the United States. The NHANES household interview includes a section on health insurance coverage. Recently the 1999-2004 NHANES was linked to Centers for Medicare and Medicaid Services (CMS) Medicaid enrollment and claims data from 1999-2007, using two different linkage algorithms. The linkage algorithms differed based on the personal identifier information used for matching. From 1999-2004, NHANES included 14,065 children less than 18 years of age; 11,502 (82%) were linkage eligible, meaning they had sufficient personal identifier information (PII) and/or had not refused to provide their Social Security Number (SSN) and had a SSN verified by the Social Security Administration Enumeration Verification System. Concordance in Medicaid reporting was defined as having Medicaid reported as a health insurance source in NHANES and having a CMS administrative record indicating Medicaid enrollment within a month of the NHANES interview. Another form of concordance was no report of Medicaid coverage in NHANES and no match to the CMS administrative records for Medicaid enrollment. Discordance was defined as a report of Medicaid in NHANES without CMS evidence of enrollment, or, conversely, no report of Medicaid in NHANES while actually being enrolled. We will compare the prevalence and characteristics of the concordant and discordant pairs according to the linkage algorithm. These results will be discussed in the context of challenges encountered in linkage, including the linkage algorithm and accurately collecting this information in a self-reported survey.

CONCURRENT SESSION VII-B:

QUESTIONNAIRE AND SURVEY DESIGN

Measurement Strategies for Identifying Holders of Certificates and Certifications

Sharon Boivin and Isaiah O'Rear (National Center for Education Statistics, USA)

Attaining a postsecondary credential has become increasingly important for securing jobs in the United States. However, degrees awarded through credit-bearing instruction in traditional higher education institutions comprise only a portion of subbaccalaureate education. Data limitations have precluded any analysis of the economic returns of alternative credentials such as educational certificates and industry-recognized certifications.

These data limitations have long been recognized by the federal statistical community. In early 2000, an interagency committee was formed to consider improving the measurement of educational attainment. Because of the technical complexity of this topic and the relative lack of policy interest during that period of time, no further action was taken until recently.

In 2009, the Council of Economic Advisors and the Office of Management and Budget asked NCES to take the lead on developing new survey items for federal data collections that measure education and training. NCES convened a federal interagency working group with representatives from Census and BLS that has been working to develop a parsimonious set of items that can accurately identify adults with certificates and certifications.

This paper will describe the process of item development: conducting a review of research literature and data collections; developing a bank of existing survey items; conducting focus groups to learn how potential respondents talk about certificates and certifications; developing and testing in cognitive interviews a draft questionnaire designed to identify respondents with certificates and certifications; and conducting a pilot study with

a nationally-representative address-based sample to evaluate questionnaire items used to identify adults with certificates and certifications.
The pilot study also included a seeded sample of known holders of certificates and certifications to help evaluate false negatives. The paper will include our analysis of pilot study survey item accuracy and effectiveness based on comparisons with the seeded sample and existing data sources.

Evaluation of the 2008 American Community Survey Employment Status Question Change
Alfred Gottschalck, Braedyn Kromer, David Howard, and David Hedengren (U.S. Census Bureau)

For the 2008 American Community Survey (ACS), changes were made to the questionnaire that modified and improved existing questions for several subject areas.1 In particular, an improved series of labor force questions was introduced to better capture data on employment status. Prior research and analysis of employment data from the ACS revealed that employment levels were underestimated and unemployment levels were overestimated relative to benchmark data from the Current Population Survey (CPS) or Local Area Unemployment Statistics (LAUS) program at the Bureau of Labor Statistics. The CPS and LAUS are used to produce the official estimates of employment and unemployment for the nation and states.

We provide a summary of the rationale for the question change and present the specific changes made to the series of labor force questions capturing employment status data. Second, we discuss the anticipated impact on employment status data from the questionnaire change, in addition to discussing the characteristics of those respondents answering the revised question sequence. Third, benchmark comparisons are made to both CPS and LAUS data for the years 2007, 2008, and 2009. Fourth, we present employment status data by mode of collection to examine the role changing mode proportions may have played in explaining the differences between 2007 and 2008. Lastly, we provide a straightforward methodology to "bridge" the 2007 and 2008 employment status data.

We believe the modifications and improvements to the 2007 ACS series of labor force questions had the effect of increasing the number of employed persons captured in the 2008 and 2009 ACS data. Given the decreased prevalence of statistical differences between ACS and CPS/LAUS employment status data for the years 2008 and 2009, we believe that the 2008 and 2009 ACS data represent an improvement in the estimates of employment and unemployment data compared to prior ACS data.

A summary of the changes made to the 2008 ACS can be found at http://www.census.gov/acs/www/AdvMeth/content_test/summary_results.htm. The 2007 and 2008 ACS questionnaires are available at http://www.census.gov/acs/www/Downloads/SQuest07.pdf and http://www.census.gov/acs/www/Downloads/SQuest08.pdf, respectively.

National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR) Cell Phone and Debit Card Test
Elke McLaren, Aniekan Okon, and Denise Pepe (U.S. Census Bureau)

The FHWAR is an address-based sample with interviews conducted by Computer-Assisted Telephone Interviewing (CATI) and Computer-Assisted Personal Interviewing (CAPI). The 2011 FHWAR budget will only fund 5,154 CAPI interviews out of approximately 34,064 households with no available telephone number. Decreasing our sample from 34,064 to 5,154 introduces additional variance in our survey data.

The purpose of the Cell Phone and Debit Card Test is to research alternative survey designs that could increase the number of CATI interviews while reducing the variance associated with conducting fewer CAPI interviews.

This test contains three panels of 500 households with no available telephone number. Each advance letter, although unique to the panel, requests that a respondent calls the telephone center to conduct a telephone interview. The panels are described below.

1. Advance Letter and Cell Phone – For Communication Between the Household and the Census Bureau
2. Advance Letter with a $25 Debit Card Incentive – PIN Received Upon Completion of Interview
3. Advance Letter Only – Wording Impresses that CATI interviews Save Government Dollars.

We will compare the results of our test panels to our production CATI and CAPI samples to evaluate whether any of the test options are viable alternatives to collecting data through a personal visit when we have an address but no telephone number. If this study proves successful, it will provide an option for future surveys to decrease interviewing costs while equalizing the difference between households with a telephone number and households without a telephone number.

At the conference, the authors plan to present study logistics, results, and suggested improvements for further study.

The Missing Link: From Concepts to Questions in Economic Surveys

Diane Willimack (U.S. Census Bureau) and Ger Snijkers (Statistics Netherlands)

Data collected in business surveys typically rely on technical definitions and well-specified measurements. Moreover, the desired data are expected to be available in business records. However, underlying concepts may not be as clear-cut or as measurable as expected. Even a concept as seemingly straightforward as "employment" has different dimensions. E.g., a straight "head count" or number of Full-Time-Equivalents? Should part-time workers be included? What about "temps," leased employees, or contractors? Should employees on paid or unpaid leave be counted? And so on...

Questions in business surveys are intended to provide valid measurements of underlying economic concepts (e.g., construct validity) that often have many attributes, some measurable, some not, some know and some unknown. In survey practice, the variety of attributes may lead to mismatches with a respondent's interpretation or with available data, resulting in measurement error, as collected data fail to meet the intent of the survey question, the underlying concept, or the needs of data users. Survey designers are often unable to identify this ambiguity in the questions or concepts until cognitive pretesting or even data collection is complete, demonstrating that the missing link from concepts to questions and data is often overlooked in business surveys.

The aim of this paper is to put a spotlight on this missing link, not only for survey questions but also for evaluating the validity of register data. We will discuss research methods that can be used to investigate concepts, ascertain attributes, identify measurements, and specify questions, to achieve construct validity at the design stage. We distinguish between top-down theory-driven and bottom-up data-driven approaches. These methods are, among others: dimension/indicator analysis, feasibility or exploratory studies, early stage scoping, record keeping studies, focus groups, concept mapping, and factor analysis. A number of these methods will be described and illustrated with examples.

CONCURRENT SESSION VII-C:

CHALLENGING ISSUES FOR TODAY AND TOMORROW: INNOVATIVE APPROACHES

Consequences of Social Security Number Randomization

Bert Kestenbaum (Social Security Administration, USA)

On June 25, 2011, the Social Security Administration (SSA) changed the method of assigning social security number (ssn) from a structured method to a random method. The SSA believes that the random assignment of ssns will help protect the integrity of the ssn and also extend the longevity of the nine-digit configuration.

In this paper we report and elaborate on the consequences of this change. These consequences arise because of the loss of place and time information embedded in the structure of the ssn and because the assignment method allowed the identification of particular populations of interest.

More specifically, the first 3 digits of the ssn generally identify the State of residence of the applicant for the number. Therefore, for example, the social security numbers of persons currently residing in some area provide information on the extent of migration into that area. The first 5 digits of the ssn identify when the ssn was assigned. Then, for example, this information taken together with the numberholder's year of birth, might serve as a means for identifying immigrants.

Two examples of how particular populations of interest can be identified stand out. We identify twins from the proximity of their social security numbers. Especially now that the very great majority of infants are enumerated at birth, it happens frequently that their ssns follow one another in sequence. Indochinese refugees in the mid-1970s were assigned a distinct set of numbers, and thus the ssn can be used to study their economic situation.

Taxable Property Values: Exploring the Feasibility of a Survey

Peter Schilling and Brian Zamperini (U.S. Census Bureau)

In 2007, the Committee on National Statistics evaluated Governments Division programs and recommended that the Division "carry out a program of research and testing to explore conceptually sound and cost-effective means of collecting data on taxable property values". The Governments Division of the U.S. Census Bureau initiated a study to determine the feasibility of collecting and producing national statistics on taxable property values. One of the first aspects of the feasibility study was to examine options for data collection that meet statistical standards. We evaluated the availability and quality of data from a third party vendor, direct collection from local jurisdictions, and local jurisdiction websites. To evaluate the quality and availability of data from direct collection and local jurisdiction websites, a sample of local jurisdictions representing a small, medium, and large government from each of the 50 states was selected for both assessment and recorder of deeds

offices. The insights provided by the jurisdictions, along with the evaluation of the other data collection methods helped us to understand the complexity of the multiple facets of developing a survey of taxable property values. This paper presents the analysis and summarizes our findings on the feasibility study of data collection methods.

Measurement Error in the Relationship Status of Same-Sex Couples in the 2009 American Community Survey
Matthew Virgile (U.S. Census Bureau)

In the American Community Survey, same-sex couples living together may describe their relationship as either "husband or wife" or "unmarried partner." Some of these couples report their relationship differently from others, despite living in the same objective conditions. Past literature finds that how same-sex couples define and report their relationship depends on both legal and sociological factors; as a result, far more same-sex couples report their relationship as "husband or wife" than are legally married. Using 2009 ACS 1-year internal data and logistic regression we test three hypotheses: same-sex couples are more likely to report as "husband or wife" (as opposed to "unmarried partner") when they (1) reside in a state recognizing same-sex marriage, (2) live with their own children, and (3) respond to the initial mailout form.

Statistics on Temporal and Spatial Activity Based on Mobile Phone Network Data
Edwin de Jonge, Merijn van Pelt, and Marko Roos (Statistics Netherlands)

We explore a mobile phone dataset for its possible use in official statistics, for example dynamic population density and movement statistics. The dataset acquired from a telecommunication company contains all speech and text call-events on their network in the Netherlands for a period of two weeks. Each record contains the time, serving cell tower and a scrambled identification key of the connected phone. The data set contains over 600 million call-events, 20,000 different cell tower antennas with their geo-location and 5 million different phones.

We describe our method which transforms the cell tower locations into a Voronoi tessellation appropriate for geo-spatial analysis.

We made a geo-spatial animation that clearly shows that high call intensity coincides with high population density and activity. The patterns in the time series of call activity are explored using k-means clustering, which suggest a correlation between call activity and economic activity. Our (simplified) analysis of phone mobility gives an underestimation of the official average travelling distance. "High mobility" phone analysis clearly reveals the main infrastructure of the Netherlands.

Future research will contain a detailed look at tourism in the Netherlands, which could lead to locations visited by tourists and their length of stay. These results would improve the official statistic.

CONCURRENT SESSION VIII-A:
MULTIPLE IMPUTATION

MI Double Feature: Multiple Imputation to Address Nonresponse and Rounding Errors in Income Questions Simultaneously
Jörg Drechsler (Institute for Employment Research, Germany) and Hans Kiesl (University of Applied Sciences, Germany)

Questions on income in surveys are prone to two sources of errors that can cause bias if not addressed adequately at the analysis stage. On the one hand, income is considered sensitive information and response rates on income questions in general tend to be lower than response rates for other non-sensitive questions. If the missing data mechanism for income is not missing completely at random (MCAR) – and there is ample of research indicating that it isn't – results based only on the observed income values will be biased. On the other hand respondents tend to round their income. Depending on the respondent and on the value of income, the magnitude of rounding can range from rounding to the closest 5 dollar value to rounding to the closest 5000 dollar value. While this kind of measurement error will not affect statistics like the mean of the income variable if the rounding is random, analysis regarding the distribution of income such as the median income will be biased. This can be especially problematic if income thresholds are used to establish program eligibility.

In this paper we propose a two stage imputation strategy that estimates the posterior probability for rounding given the observed values on the first stage and re-imputes the observed income values given the rounding

probabilities on the second stage. Missing values are also imputed at this stage. We provide a simulation study that illustrates that the proposed imputation model can help overcome the possible negative effects of rounding. We also present results based on the household income variable from the German panel study "Labour Market and Social Security".

Multiple Imputation: Limitations Born From Missing Data Among Covariates

Kenneth Steve and Li Leung (Bureau of Transportation Statistics, USA).

The Bureau of Transportation Statistics (BTS) has conducted the National Census of Ferry Operators (NCFO) biennially since 2006. Data is collected from approximately 260 ferry operators currently operating in the United States. This data is used to maintain the national ferry database containing information regarding routes, vessels, passengers and vehicles carried, funding sources, etc.

As with many surveys or censuses of businesses, ferry operators have shown a reluctance to provide certain information. More specifically, some operators consider passenger boarding data to be business sensitive information. While a significant number of operators simply don't provide the information, others may ask that it not be made public. This presents BTS with challenges in regards to producing accurate estimates of ferry passenger boardings.

In an effort to generate a more useful picture of the true number of passenger boardings for the 2006 NCFO, a SAS macro for multiple imputation was employed (Giesbrecht, 2008). The current paper will discuss some of the difficulties in trying to reproduce that effort within the 2008 NCFO dataset. The missing data patterns for both years were non-monotone; therefore a Markov chain Monte Carlo method was used to estimate missing data. Using this simultaneous inference approach with the 2008 data led to problems with negative (i.e., non-plausible) values being imputed for the number of passengers in many cases.

Attempts to restrict the range of imputed values to non-negative integers led to a failure of the imputation model to converge. Further attempts to resolve the issue included the removing covariates with inverse relationships to passenger counts, removing data points with extreme influence, and removing covariates with missing data. In the end, a SAS macro for nearest neighbor imputation was used to produce plausible imputation values that were within range of what one would expect given the previous census data.

Effect of Different Imputation Methods on Factor Analyses of CAHPS Nursing Home Survey

Robert Baskin, Judy Sangl, and Marc Zodet (Agency for Healthcare Research and Quality, USA)

The Consumer Assessment of Healthcare Providers and Systems (CAHPS) program is an initiative to develop standardized surveys of patients' experiences with health care. CAHPS designed questionnaires for the two main types of resident populations (long-stay (usually >3 months) chronic care and short-stay post-acute care). These questionnaires were tested in 13 nursing homes in several New England states in 2005 but the short-stay sample was insufficient to determine the efficacy of CAHPS for this population. In 2009 Maryland conducted a large-scale survey (N=1828) of short-stay patients discharged from Maryland nursing homes using the CAHPS Nursing Home questionnaire. The primary purpose was to confirm the factor analysis from the 2005 long-stay population, composed of five factors: environment, care, communication, autonomy and activities. During the confirmatory factor analysis, which requires nonmissing values on all observations for all variables, issues arose about the methodology for imputing for the large number of missing items primarily due to survey skip patterns for some items. Different approaches to dealing with the missing values were explored including multiple imputation with Proc MI in SAS and MICE and multiple hotdeck in R. The results (such as various goodness of fit statistics) under the different imputations are reported including the effect of imputation on Cronbach alpha for the different factors. The multiple hotdeck is proposed as the preferred solution.

CONCURRENT SESSION VIII-B:

ISSUES IN DESIGNING AND REDESIGNING ESTABLISHMENT SURVEYS

Closing the Feedback Loop in Survey Design–How Public Sector Research Needs Inform Evolution of the Census of Agriculture

Luanne Lohr (Economic Research Service, USA) and Donald Buysse (National Agricultural Statistics Service, USA)

The Census of Agriculture is an enumerative and descriptive survey of American farm and ranch operators that is conducted every five years by the U.S. Department of Agriculture's National Agricultural Statistics Service (USDA-NASS). Response is mandated by law. Data are collected on land use, demographics, production practices, farm economics, and other topics.

The Census of Agriculture provides the only source of uniform, comprehensive agricultural data that is tabulated for every county in the nation. Census data are used by public sector researchers to discover trends,

predict program outcomes, evaluate program effectiveness, and respond to legislative requests for information prior to policy formulation. Two goals for data development are stability – obtaining the same statistical data over the years to establish a time series – and responsiveness – including questions to elicit information on emerging topics that are of interest to consumers of the data products.

To address these goals, the USDA-NASS may add or reformulate questions in the Ag Census to identify key subgroups from which additional information may be collected in Special Studies, known as census follow-on surveys. The 2008 Organic Production Survey and 2009 On-Farm Renewable Energy Production Survey were follow-ons conducted in response to the need for detailed information about these subgroups. The survey populations were established based on responses to identifier questions in the 2007 Census of Agriculture.

Given the expense of designing, testing and conducting the Ag Census, changes require strong justification. This paper explores the feedback loop between public sector researchers and the design of the Census of Agriculture. We describe the process of design evolution resulting from policy and research needs, and offer two examples of feedback loops that resulted in changes to the Ag Census questionnaire – organic production and local foods marketing. We discuss the challenges of meeting data user needs and pose creative solutions.

Survey Redesign Process for a Business Survey: Lessons Learned From the Business R&D and Innovation Survey
Richard Hough (U.S. Census Bureau)

Over the last five years, the National Science Foundation and the U.S. Census Bureau's Economic Directorate have collaborated on a redesign of the Survey of Industrial Research and Development. In 2002, NSF commissioned a review of its R&D surveys by the Committee on National Statistics (CNSTAT) of the National Research Council. CNSTAT appointed a panel of experts representing the fields of statistics, survey research, economics, data analysis, research policy, and R&D to review current data coverage and relevance, collection systems, and survey and statistical methodologies. The report, Measuring Research and Development Expenditures in the U.S. Economy, was published in 2005. As its core recommendation, the panel concluded that it was time to implement a major redesign of the Survey of Industrial Research and Development. During the period 2005 through 2008, NSF and Census staff engaged in a number of activities to inform the redesign. Many of the steps were guided by the CNSTAT recommendations. The resulting 2008 Business R&D and Innovation Survey (BRDIS) was a 56-page instrument to collect data on R&D and other innovation activities, which in January 2009 was sent to approximately 39,500 businesses in the United States. This paper briefly describes the process for redesigning the R&D survey, and the lessons learned from the process.

Comparison of the Data Reliability of Nationally Representative Estimates From the Full (2007) and Half (2008) Samples of the National Hospital Discharge Survey –Implications for Trend Analyses
Margaret Jean Hall and Maria Owings (National Center for Health Statistics, USA)

The threat of budget cuts and their impact on reliability of survey estimates is a problem generalizable across statistical agencies. This paper describes effects of sample size cuts on the National Hospital Discharge Survey (NHDS), conducted annually since 1965 by the National Center for Health Statistics, Centers for Disease Control and Prevention. NHDS is a national probability sample survey of discharges from nonfederal, short-stay hospitals. Until 2007, data were gathered from about 400 hospitals and 360,000 discharges. In 2008-2010, data were gathered from about 200 hospitals and 166,000 discharges. The decision to halve the sample was made after projecting the anticipated effects this would have on reliability of estimates of inpatient care released annually.

This paper discusses the results of the half sample on the weighted estimates and their standard errors. It first compares the 2007 and 2008 data in large categories like hospital discharges' age, sex, region, source of payment and discharge disposition. Next, smaller estimates for diagnoses, and surgical and nonsurgical procedures, of particular interest are compared. As projected, many of the one-year 2008 estimates from the NHDS continued to be reliable. For other estimates it was necessary to combine data over 2 years to obtain reliable estimates. There were factors, unrelated to the half sample, which had unexpected effects on the reliability of the estimates. The information in this paper is particularly important when, as in the case of the NHDS, survey data are used for tracking trends over time.

Integration of the National Hospital Discharge Survey and the National Hospital Ambulatory Medical Care Survey Into the National Hospital Care Survey
Carol DeFrances and David Woodwell (National Center for Health Statistics, USA)

The Centers for Disease Control and Prevention National Center for Health Statistics (NCHS) gathers statistics on the use, access, quality, and cost of health care provided in the United States. However, the growing number

and range of settings in which health care is delivered presents operational challenges. Historically, NCHS has conducted three national surveys annually across six ambulatory and hospital-based settings. NCHS is launching a new survey, the National Hospital Care Survey (NHCS), which integrates the National Hospital Discharge Survey (NHDS) and the National Hospital Ambulatory Medical Care Survey (NHAMCS). NHDS, conducted continuously since 1965, produces nationally representative hospital inpatient utilization estimates. NHAMCS, conducted since 1992, collects data on use of hospital emergency and outpatient departments (EDs and OPDs) and ambulatory surgery centers (ASCs).

There will be two phases to the initiation of NHCS. In the first phase, starting in 2011, NCHS will recruit a new sample of 500 hospitals. Participating hospitals will be asked to submit Uniform Bill (UB)-04 administrative claims data for all inpatients. The second phase will occur in 2013 and will integrate NHAMCS with the inpatient component.

NHCS will continue to provide the nationally representative estimates and trends that NHDS and NHAMCS currently provide. However, the new survey has some strategic advantages. NHCS will collect protected health information to allow linkage across hospital settings and to other data sources such as the National Death Index and Medicare data. In addition, NHCS will gather UB-04 data electronically and as hospitals adopt electronic health records, NCHS will be poised to link these data to UB-04 data. By collecting UB-04 data for all inpatients, it will be possible to conduct special studies that focus on specific research and policy relevant questions.

This paper will describe the process and challenges of integrating two surveys with similar frames but different objectives.

CONCURRENT SESSION VIII-C:

DISCLOSURE

Generalized Domain Size Threshold for Analysis Restrictions for Remote Analysis Servers

Avinash C. Singh, Joshua M. Borton, and Allan M. Crego (NORC at the University of Chicago, USA)

We consider output treatment of analyses based on remote analysis servers where the user is not granted access to the microdata housed behind firewalls. This is currently a very active research area with important contributions from government agencies such as MAS by the Census Bureau, ANDREW by NCHS, and DAS by NCES. We propose an alternative system of generalized domain size threshold for analysis restrictions (GDSTAR) which relies on a list of screened analytic variables satisfying various aggregate level analysis restrictions besides usual restrictions with respect to allowable variable transformations and unit level analysis output such as quantile estimates and residual plots for model diagnostics. Therefore, unlike other remote analysis servers, there is no need for perturbation of estimates and additional restrictions for aggregate level output. Moreover, even for advanced analysis, there are no deemed requirements of any kind of user authorization and approval for exporting analysis results. The main contribution of this paper is the characterization of various aggregate level analysis restrictions in term of a new concept of minimum generalized domain (or g-domain) sample size derived from estimating functions. This leads to the creation of a simple check list of screened analytic variables that greatly simplifies the analyst's job to identify suitable analytic variables before submitting queries. The condition of minimum g-domain sample size is not really restrictive because it is also often used to address concerns about reliability or precision of estimates besides confidentiality. The proposed method GDSTAR leads to a new type of PUF (referred to as a query-based PUF) which has only indirect access through queries along with analysis output treatment. The proposed method was partially developed in the context of creating PUFs from CMS Medicare claims data files for which its application in providing a query-based PUF is discussed as an illustration.

Logistic Regression With Variables Subject to Post Randomization Method

Yong Ming Jeffrey Woo and Aleksandra Slavkovi (The Pennsylvania State University, USA)

An increase in quality and detail of publicly available databases increases the risk of disclosure of sensitive personal information contained in such databases. The goal of Statistical Disclosure Control (SDC) is to provide data in such a way that individual information is sufficiently protected against disclosure, while providing society with as much data as possible needed for valid statistical inference. One such SDC method is the Post Randomization Method (PRAM), where values of categorical variables are perturbed via some known probability mechanism, and only the perturbed data are released. To address issues of disclosure risk and in particular data utility with PRAM, we propose a number of EM-type algorithms to obtain unbiased estimates of the logistic regression model after accounting for the effect of PRAM. We also derive estimates of standard errors, evaluate the effect of the level of perturbation and sample size on the estimates, and discuss extensions to generalized linear models.

Disclosure Avoidance Through Noise Infusion and Synthetic Data: The Case of the Quarterly Workforce Indicators
John Abowd and Lars Vilhuber (Cornell and U.S. Census Bureau), Kaj Gittings (Louisiana State University, USA), Kevin McKinney (U.S. Census Bureau), Bryce Stephens (Bates White Economic Consulting, USA), and Simon Woodcock (Simon Fraser University, USA)

The Census Bureau's Quarterly Workforce Indicators (QWI) provide detailed quarterly statistics on employment, accessions, layoffs, hires, separations, full-quarter employment and flows, job creations, job destructions, and earnings, tabulated by worker characteristics such as age, gender, and more recently (since 2011) race, ethnicity, and education in various combinations. The data are released for detailed NAICS industries and for several levels of geography, the lowest aggregation of which are counties. OnTheMap, another Census Bureau product, provides a subset of these tabulations at the tract level. Disclosure avoidance methods are required to protect the information about individuals and businesses that contribute to the UI wage records, the QCEW/ES-202 quarterly reports, and the Census Bureau demographic data that have been integrated with these sources.

The QWI disclosure avoidance mechanism relies heavily on the use of noise infusion through a permanent multiplicative noise distortion factor, used for magnitudes, counts, differences and ratios. There is minimal suppression and no complementary suppressions. To our knowledge, the release in 2003 of the QWI was the first large-scale use of noise infusion in any official statistical product.

As we show, the released statistics are analytically valid along several critical dimensions – measures are unbiased and time series properties are preserved. We provide an analysis of the degree to which confidentiality is protected. Furthermore, we show how the judicious use of synthetic data, injected into the tabulation process, can completely eliminate suppressions, maintain analytical validity, and increase the protection of the underlying confidential data.

CONCURRENT SESSION VIII-D: TECHNICAL DEMONSTRATIONS
DATA COLLECTION AND MONITORING TECHNOLOGY

Technology Advancing Data Collection: Implementing Thin Client Computer Assisted Personal Interviewing Into National Agricultural Statistics Services' 2011 Field Data Collection Program
Michael Gerling, Pam Hird, and James Harris (National Agricultural Statistics Service, USA)

The National Agricultural Statistics Service (NASS) surveys farmers and ranchers across the United States and Puerto Rico in order to estimate crop production, number of livestock, assess production practices, and identify economic trends.

Computer Assisted Personal Interviewing (CAPI) involves an interviewer going to the respondent with the questionnaire on a computer (generally a laptop) and recording all responses onto the laptop. Although, CAPI is been around since the 1980s, all forms of CAPI have been centered around data and the data collection instrument residing on the laptop. Also, over the past three years, there have been several accounts of agencies' and businesses' losing or having stolen laptops which contained clients' personal information. While the possibility of data actually being compromised is small due to the encryption of the data on the device, it still exists. This has caused the general public to be more hesitant in participating in surveys which is impacting the entire survey industry.

NASS' CAPI solution is an innovative thin-client approach which leverages wireless broadband technology and a recently completed Web based data collection system named Electronic Data Reporting System.

The term Thin Client CAPI (CAPITC) was coined. In general, a field interviewer using a tablet computer (Apple iPad) accesses the Internet using a broadband wireless mobile hotspot. The field interviewer then accesses NASS' Web data collection site and brings up the questionnaires that need to be completed by that particular agricultural operator. The interview is completed real time over the Internet with no data ever residing on the tablet.

This demonstration/presentation will show how CAPITC can be used to be an efficient and cost effective way of collecting field data.

Data Quality Record—A Desktop Application
Judy Lieberman, Joseph Greenblott, and Patricia Mundy (U.S. Environmental Protection Agency) and Michael Crow (Crow Environmental, USA)

The US Environmental Protection Agency has developed a unique Lotus Notes application to manage data quality records for performance data. The format of the record was modeled after the federal government's

Open Government metadata form. It captures the quality procedures and management controls associated with environmental or administrative data, starting from data collection, through storage in an IT system, and the calculation of the performance result. In another area of the database, program managers are required to certify by electronic signature that the quality procedures described in the record were followed. In this way, the public can be confident of the accuracy of the published information, as well as be informed of the methods for data handling and manipulation.

Federal agencies like the US Environmental Protection Agency use performance measures to track progress toward strategic and annual goals. Performance targets are published prospectively in the Consolidated Budget, and results achieved relative to the targets are disseminated in the agency's Annual Performance Plan. Often the data used for measuring performance were collected for a different original purpose. Maybe they were collected to track compliance, as prescribed by statute, and now, for example, track "percent of population receiving drinking water that meets all applicable health-based drinking water standards." As a result, the data may have to be highly aggregated to calculate a performance result.

The Data Quality Record was developed to capture the information critical for reproducibility and transparency of the performance result. It includes measure term definitions, calculations, assumptions, and data limitations. It also includes information on applicable EPA quality assurance standards, and risk management procedures. For internal use only, it identifies individuals and their responsibilities for quality assurance and control.

The database could have wide applicability in the federal government as a desk-top tool for tracking management controls, and accountability for any type of data used for a specific purpose. For the EPA, certification by the programs that quality procedures were adhered to before publishing performance results provides the basis for the Administrator's annual statement to the President on the completeness and reliability of the Agency's performance data in its Annual Performance Plan.

CARI Interactive Data Access System
Carl Fisher (RTI International, USA)

The CARI Interactive Data Access System is a CARI monitoring system developed for the US Census Bureau by RTI. It provides an easy-to-use platform for management, coding, and reporting of audio and image files associated with survey cases. It is a multi-faceted web-based system that allows CARI files to be reviewed for different purposes, including behavior coding, quality assurance, and interviewer coaching. It provides a consistent framework where CARI files can be reviewed and coded according to the goal of the analysis, using a standardized or customized set of codes; this allows for comparison of data across interviewers, coders, surveys, and even survey modes.

In addition to providing a standardized coding system, the system offers a wide range of configuration, workflow, and reporting tools needed to deploy the system for production use at a field office or telephone center, or across many field offices and telephone centers. It provides multiple levels of security and is certified to operate in FIPS-Moderate environments.

Our demonstration is for those who are interested in using CARI as a method for improving interviewer performance, data quality, or questionnaire performance, for those who are coding interviews using manual or subjective methods, and for those who are interested in seeing the Census Bureau's state-of-the-art system.

RTI Quality Evaluation System (QUEST)
Susan Kinsey, Derek Stone, and Curry Spain (RTI International, USA)

QUEST (Quality Evaluation System) is a survey mode independent system and set of protocols for evaluating interviewer performance either real-time (while the interview is being performed) or post survey administration (through a review of the recorded interview). RTI built QUEST to support all phases of quality monitoring including selecting interviewers, observing their work, evaluating performance, providing feedback, and analyzing performance data across interviewers to identify cross-cutting instrument or performance issues. QUEST protocols can be customized to meet specific project quality requirements, such as the sampling algorithm for selecting interviewers to monitor, the selection of specific sections within the survey instrument to evaluate, or the inclusion or exclusion of evaluation form modules based on the type of survey being monitored. QUEST supports monitoring of field interviews using computer audio-recorded interviewing (CARI) as well as both live and recorded telephone interviews.

QUEST offers a standardized coding scheme and a wide range of configuration and reporting tools to support production monitoring of telephone and field studies. It also provides secure capabilities for internal and external clients to perform remote monitoring of interviews.

This demonstration is aimed at those who are interested in using CARI for interviewer performance and data quality improvements, and implementing standardized protocols for interview quality assessment and performance feedback. We will review some of the key features of the web-based system, including the standardized evaluation form and scoring protocol, and discuss differences in the monitoring protocol for live and recorded interviews.

CONCURRENT SESSION IX-A:

BESTING THE ODDS: OPTIMAL REPORTING OF LOGISTIC REGRESSION

Understanding Regression Risk Analysis in Multinomial and Ordered Logistic Regressions

Morgen Miller and Edward Norton (University of Michigan, USA), Lawrence Kleinman, Jason Wang, and Kasey Coyne (Mount Sinai School of Medicine, USA)

Health care researchers frequently address questions where the outcome of interest is polytomous. The analysis of such outcomes is challenged by the ability to translate regression results into clear-cut conclusions. The interpretation of multinomial or ordered logistic regressions in particular, like logistic regressions more generally, commonly relies on odds ratios to describe results. In light of limitations of the odds ratio, this study extends regression risk analysis to multinomial and ordered logistic regressions, allowing results to be reported as adjusted risk ratios (ARR) and adjusted risk differences (ARD). The study describes how to apply regression risk analysis to multinomial and ordered logistic regressions.

We extend the methods of regression risk analysis to multinomial and ordered logistic regression and apply this procedure to a publicly-available, nationally-representative dataset. Data from the Medical Expenditure Panel Survey (MEPS) is used to predict determinants of insurance coverage among working-age adults. The risk of a particular insurance coverage will be illustrated for covariates such as race and ethnicity, sex, education and marital status. The benefits of ARR and ARD over adjusted odds ratio will be described with examples.

This study extends the work of Kleinman and Norton (HSR 2009) to analyze multinomial and ordered logistic regressions using regression risk analysis. We validate our methods using Monte Carlo simulations and illustrate our approach in a complex panel survey using MEPS data. We will illustrate user-friendly Stata code.

Health care researchers are not confined to reporting odds ratios after running multinomial or ordered logistic regressions. We show that it is easy to compute risk ratios and risk differences, either for the overall sample or for subsamples of interest, with standard errors. These methods should improve best practice and the translation of results for policymakers.

Estimating Variance for Regression Risk Analysis

Jason Wang (Mount Sinai School of Medicine, USA), Edward Norton (University of Michigan, USA), Kasey Coyne (Mount Sinai School of Medicine, USA), Morgen Miller (University of Michigan, USA) and Lawrence Kleinman (Mount Sinai School of Medicine, USA)

Logistic regression is the most popular method researchers use to model a dichotomous outcome. Although the literature has identified a few standard ways to present results from logistic regression, as odds ratios or risk ratios, it has not settled on a way to calculate the variance associated risk ratio estimates. There exist several approaches to estimate the Adjusted Risk Ratio (ARR) from the modeling results. Regression risk analysis is one such method with desirable performance characteristics. The current study describes our investigation of the relative advantages of the two most common approaches to measuring variance as it applies to regression risk analysis: the Delta method, which models a first order Taylor series expansion, and bootstrapping, a numeric (non model based) approach to estimating variance.

The Delta method explicitly models the variation that results from the two distinct sources that contribute to variance: differences in the distribution of covariates across the sample and uncertainty in the estimates of the coefficients. Standard errors are computed and symmetric confidence intervals may be estimated. Although the computations are complex and may represent a challenge to program for complex models, the computer time required to run the program is typically modest. In contrast, bootstrapping is relatively easy to program but can take a long time to run, particularly for large data sets. Confidence intervals are estimated directly from the bootstrap and may be asymmetrical; standard errors may be estimated from the width of the confidence interval.

We use Monte Carlo simulations that incorporate variation and confounding to demonstrate and specify the advantages and disadvantages of these two approaches to estimating uncertainty and variance for adjusted risk measures from regression risk analysis. We will propose decision rules to help investigators choose how to calculate and describe uncertainty in specified circumstances.

Practitioner's Guide to Regression Risk Analysis Using SAS and Stata
Edward Norton (University of Michigan, USA), Lawrence Kleinman (Mount Sinai School of Medicine, USA), Morgen Miller (University of Michigan, USA), Kasey Coyne (Mount Sinai School of Medicine, USA), and Jason Wang (Mount Sinai School of Medicine, USA)

How should one quantify the results from a logistic regression? Kleinman and Norton (HSR 2009) propose a general method for estimating risk ratios and risk differences from non-linear multiple regression analysis, using the example of logistic regression. The simple-to-measure odds ratio can deviate greatly from the more intuitive risk ratio. It also can suggest a less effective treatment is more effective. Current software does not automatically produce risk ratios and risk differences for easy interpretation. We intend to fill that gap.

In this session we will demonstrate how to calculate risk ratios and risk differences from logistic regression results using user-friendly SAS and Stata software. In particular, we will show how to compute risk ratios, risk differences (and odds ratios for comparison) for individual observations, the full sample, and for any interesting sub-sample. We will discuss the interpretation and how to present the results in a research paper.

Part of making predictions is calculating confidence intervals for those predictions. We will show how to calculate standard errors and confidence intervals using both delta method and bootstrapping, with examples.

The session will use examples from the well-known MEPS data. The session will be interactive and demonstrate the use of software in real time. Therefore, this session will be of great value to practitioners. All software will be available on a public website.

Rank Reversal in Comparative Effectiveness Research
Lawrence Kleinman (Mount Sinai School of Medicine, USA), Edward Norton (University of Michigan, USA), Kasey Coyne (Mount Sinai School of Medicine, USA), Morgen Miller (University of Michigan, USA), and Jason Wang (Mount Sinai School of Medicine, USA)

Following up our recent description of regression risk analysis, which estimates adjusted risk ratio (ARR) and adjusted risk difference (ARD) from logistic regression and other nonlinear models, we ask whether the use of ARR vs the more common adjusted odds ratio (AOR) could change rank order when comparing effectiveness of various interventions.

Both ARR and AOR measure effect size relative to a baseline outcome, while ARD measures absolute effect size, in all cases adjusting for confounders included in the model. We find that the preference for one treatment over another often changes based upon the choice of OR or RR to rank the treatments. We demonstrate this both with mathematical modeling and through re-analysis of well published meta-analyses. When the baseline differs (as in the typical comparison or integration of multiple studies, such as meta-analysis), rank order often changes. Only when comparing multiple therapies to a single baseline can one be assured that the rank order of interventions will not change based upon the choice of measure

While it has been well described that the magnitude of an OR always exceeds that of a RR, this is the first study to describe that the rank order of various therapies can differ depending upon the use of RR or OR. The relationship between AOR and ARR is identical to that between RR and OR.

In comparative effectiveness studies that incorporate more than one baseline measurement, ARR and ARD provide complementary information and AOR may confuse. The use of OR or AOR may suggest the wrong choice is preferable: choice of measure impacts rank order of interventions. Whenever using comparative effectiveness research for clinical or policy decisions, one should insist on results being reported in terms of risk.

CONCURRENT SESSION IX-B:
RESPONDENT BURDEN: MEASUREMENT AND REDUCTION

Reducing Respondent Burden: Evaluating the Progress of the 2011 SIPP-EHC
Rachael Walsh and Jason Fields (U.S. Census Bureau)

Under the current redesign plan, the Survey of Income and Program Participation is incorporating an event history calendar (SIPP-EHC), one goal of which is to reduce respondent burden by interviewing annually. Several key changes were implemented between the 2010 and 2011 tests, one of which was extending the interview period from December of the previous calendar year to the interview month. This key change was intended to

improve the flow associated with creating spells and completing topics within the calendar, though it extended the reference period from 12 to a possible 15 months. The flow of the collection of spells was changed to script more of the probing and transitions as topics are completed, reducing some of the keystrokes and burden associated with routine conversational transitions. In addition, an income screener was introduced to take questions off path that were not applicable based on income.

Using data from both the 2010 and the 2011 SIPP-EHC, comparisons with respect to respondent burden are evaluated based on the number of question asked as well as interview length. Section timers based on keystroke audit trail files are used to evaluate topical sections and sequences where significant changes were made. The design of both the 2010 and 2011 instruments produce longer individual interview lengths for some respondents given the first person interviewed provides the demographic and coverage information for the entire household, while the last person interviewed provides feedback and future contact information. In approximately 60% of the households interviewed, the first and last respondent is the same individual thus creating multiple differences in interview length among and between respondents that will be accounted for in the evaluation and detailed through a thorough description of the sample. The implications of this research will be used to guide the development of the SIPP-EHC production instrument.

Differing Person Nonresponse Rate by Interview Length and Respondent Characteristic: Results From the 2010 and 2011 SIPP-EHC Field Tests

Daniel Perez-Lopez and Rachael Walsh (U.S. Census Bureau)

The Census Bureau is re-engineering the Survey of Income and Program Participation (SIPP). Beginning in 2014, the re-engineered SIPP (SIPP-EHC) will collect data from households once a year, utilizing an event history calendar (EHC) to assist respondent recall and maintain data quality. Current SIPP has a core set of data items which are collected in every wave, and several topical modules in each wave, some of which are asked once or twice each panel. In SIPP-EHC, some of these topical module questions will become part of the core instrument, and be asked every wave.

As part of the re-engineering, tests of the SIPP-EHC instrument were carried out in 2010 and 2011, with 7,982 and 4,051 households interviewed, respectively. Paradata were generated on interview length and will be made available to Census staff in early summer of 2011.

We propose to study person non-response in the SIPP-EHC. We'll test the hypothesis that longer interviews lead to higher person non-response rates, as respondents become fatigued and consequently decline to provide information about additional household members. We will also investigate whether this time effect varies by section. Demographic data on the entire household are provided by the first interviewed adult, allowing us to explore the relationship between respondent characteristics and non-response.

The analysis plan has two components. First, we produce descriptive statistics for select demographic characteristics for the first interviewed adult as well as non-respondents, and data on the length of household interviews. Second, we model the probability of a non-response in a household as a function of the length of the first interview, characteristics of the first interviewed adult, and characteristics of the non-respondents. This model will allow Census staff to restructure the interview instrument as necessary to ensure the lowest person non-response and the highest quality data is collected.

Does Length Really Matter? Exploring the Effects of a Shorter Interview on Data Quality, Nonresponse, and Respondent Burden

Scott Fricker, Brett Creech, Jeanette Davis, Jeffrey Gonzalez, Lucilla Tan, and Nhien To (Bureau of Labor Statistics, USA)

The Consumer Expenditure Quarterly Interview Survey (CEQ) is an ongoing monthly survey conducted by the U.S. Bureau of Labor Statistics (BLS) that collects expenditure information from American households. Sample households are interviewed five times over the course of thirteen consecutive months; the interviews are long, the questions detailed, and the experience can be perceived as burdensome. This presentation reports the results of a small-scale field test which examined the effects of administering a shorter CEQ instrument on data quality, nonresponse error, and respondent burden. One-third of study participants were interviewed monthly using a 1-month reference period, and one-third received a modified CEQ which replaced half of the detailed expenditure questions with global expenditure items. Data from these two treatment groups were compared against a control group that was administered a quarterly CEQ interview consisting only of detailed expenditure questions. We present results from data quality analyses that examined both direct measures (e.g., number of expenditure reports, expenditure amounts) and indirect measures (e.g., response rates, measures of perceived burden, item nonresponse, etc.), nonresponse bias analyses (e.g., comparisons of response rates, sample composition, and expenditure estimates across treatment conditions), and discuss limitations and implications of the study.

CONCURRENT SESSION IX-C:

USING ADMINISTRATIVE DATA FOR SURVEY DATA QUALITY EVALUATIONS

Linking Medicaid Administrative Records Over Time and Space: Methods and Findings, 2005 to 2007

Shinu Verghese and John Czajka (Mathematica Policy Research, USA)

To provide access to Medicaid administrative data in a form suitable for research, the Centers for Medicare & Medicaid Services (CMS) has funded and overseen the development of an annual Medicaid Analytic Extract (MAX) containing enrollment and claims information for each person enrolled in Medicaid in a calendar year. In 2010, CMS contracted with Mathematica Policy Research to design and implement a MAX Enrollee Master file (MAXEM) that would identify unique Medicaid enrollees both within and across states and provide a reliable means of linking their records over time.

The Medicaid records that states submit to CMS lack names and addresses, so the linkage variables consisted of numeric identifiers and demographic variables. Two records were linked and assigned a common MAXEM ID if they satisfied pre-specified, deterministic linkage criteria. Through these pair-wise links, all records that appeared to represent the same individual were ultimately assigned a common MAXEM ID.

The common MAXEM ID makes it possible to generate unduplicated counts of unique individuals enrolled in Medicaid in any year or across multiple years. Over the three years our unduplicated counts of unique Medicaid enrollees nationally range from 97.35 to 97.76 percent of the total records in MAX. Of those who were ever enrolled in 2005, 82.2 percent were enrolled in the same state in 2006, and 66.8 percent were enrolled in the same state in 2007 as well. In 2005, annual-ever enrollment was 26 percent higher than the average monthly enrollment, although this varied by age.

The MAXEM file is not available to the public at this time but is part of an initiative at CMS to develop a unique ID for each Medicare and Medicaid beneficiary as part of overall efforts to improve data dissemination. CMS continues to work on this at present.

Misreporting of Food Stamp Participation in the NHANES: Results from a Pilot Study Linking Administrative Records to Survey Data

John Kirlin (Economic Research Service, USA), Michael Wiseman (George Washington University, USA), Daniel Schroeder (The University of Texas at Austin, USA), and Lisa Mirel (National Center for Health Statistics, USA).

There are widely recognized benefits to linking administrative records to other administrative records or to survey records. One benefit is to assess the validity of a survey respondent's report of program participation status. There are also barriers (confidentiality, privacy, logistic) and costs in data linkage. One "cost" not often discussed is the consequences for data analysis and policy formulation if the match process generates false positives (apparent matches that are false) or false negatives (not matching records for the same individual). Because it is often difficult to recognize when match errors occur, their likely impacts on analysis results are uncertain.

As part of the Texas NHANES Pilot, the Ray Marshall Center (RMC) in Texas has linked administrative records from the Food Stamp Program (FSP) in Texas to survey records from the National Health and Nutrition Examination Survey (NHANES). The proposed paper will provide a detailed description of the method used to match individual records from NHANES to the FSP's household records. This method evolved out of frequent discussions among the three organizations and review of interim results. The paper will then summarize what has been learned about mis-reporting of food stamp participation by NHANES respondents—one of the primary goals of the original research that has revealed an overall mis-reporting rate of at least 17 percent.

If the match process itself contains errors, then estimated rates of under- and over-reporting of FSP participation will be biased. The paper will present results from an ancillary process used to assess the apparent validity of each of the pilot's 920 original FSP-NHANES record matches. This information is used to examine the potential error in our original estimates of mis-reporting and to revise the match protocol.

This ancillary process may prove useful in assessing the accuracy of other data matches.

How Good Are ASEC Earnings Data? A Comparison to SSA Detailed Earning Records

Joan Turek, Kendall Swenson, and Bula Ghose (Department of Health and Human Services, USA) Fritz Scheuren (NORC at the University of Chicago, USA), and Charles Nelson and Edward Welniak (U.S. Census Bureau)

At the 2009 FCSM Conference, research jointly conducted by the Office of the Assistant Secretary for Planning and Evaluation/HHS and the Census (Effect of Imputation on CPS Poverty Series: 1987-2007) was presented. That work highlighted systematic differences over time in Current Population Survey (CPS) poverty rates

among persons reporting all of their income items; those reporting some information on income but missing one or more amounts (item imputes); and, those who were imputed all CPS/ASEC supplement items including their income (whole imputes). The Annual Social and Demographic Supplement (ASEC) to the CPS is the data source.

This research, however, cannot determine to what degree Census methods for imputing income bias reported poverty rates. Systematic differences in poverty rates of persons by imputation status can only be addressed by matching survey data with administrative data which provide independent estimates of income. Unfortunately, administrative records permitting evaluation of all income sources do not exist. Using funding from an interagency transfer from ASPE, the Census Bureau is matching CPS/ASEC respondents to the Social Security Administration detailed/summary earnings records by imputation status. These independent amounts will be compared to the reported/imputed CPS responses. Additional findings from matching Temporary Assistance to Needy Family and Supplemental Security Income administrative records to the CPS/ASEC will also be reported, if available in time.

Evaluating Job Data in the Redesigned SIPP Using Administrative Records

Graton Gathright, Jeremy Skog, and Martha Stinson (U.S. Census Bureau)

In January 2010, the Census Bureau conducted a field test of a new re-engineered SIPP instrument and collected information about jobs held during calendar year 2009 using an experimental event history calendar method. This paper seeks to evaluate these data by comparing answers to survey questions to information on the same firms found in the Census Bureau Business Register. The link between our demographic survey and the Business Register was performed by first using Social Security numbers to match individuals to their 2009 W-2 records as filed by employers with the IRS. We then used employer tax identifiers, names and addresses to match individual W-2 records to specific job reports in the new SIPP. Thus we are able to compare the number of employers in 2009 for matched individuals, as well as industry, firm size, type of employer (government or private sector), and earnings for specific jobs. Our results shed light on what types of jobs are not reported in surveys as well as what types of jobs are not captured by administrative data. We also highlight the differences between industry coding and firm size classification based on respondent descriptions of their employers and the same characteristics as assigned to firms in the Census Business Register. We compare the distribution of self-reported earnings to the distribution of employer-reported taxable earnings and classify individuals by degree of difference between the two earnings sources.

CONCURRENT SESSION IX-D:

SAMPLING FOR EVERY OCCASION

Expanding the Community Address Updating System Universe

Bryan Schar, James Lawrence, Star Ying, and Jim Hartman (U.S. Census Bureau)

The Community Address Updating System (CAUS) conducts targeted block listings to improve the coverage of the American Community Survey's sample frame. Blocks with certain address characteristics have traditionally been excluded from listing by the CAUS program because it was believed that data provided by other sources kept them current. However, results from the nationwide 2010 Census Address Canvassing operation suggested that some of these excluded blocks may benefit from being listed by the CAUS program. Therefore, a study was conducted to determine if the universe of CAUS blocks should be expanded. This paper will discuss the methods being used to determine if the CAUS universe should be expanded, some preliminary results of the investigation, and plans for future research.

Estimation and Sampling of Longitudinal Person-Firm Data

Kevin McKinney (U.S. Census Bureau)

Linked employer-employee universe files present unique estimation challenges (and opportunities) compared with typical cross-section or longitudinal survey data. Unlike most datasets that contain information for a single population, longitudinal person-firm data is composed of several populations. The datasets contain information on workers, firms, as well as the jobs created when a worker is employed. This information allows for the estimation of a rich set of statistical models. However, given the large scale of administrative data such as those produced by the Census Bureau's Longitudinal Employer Household Dynamics (LEHD) program, the development of suitable sampling methods reduces the computational burden and for some complex estimators may be required.

With minor modifications, traditional random sampling methods can be applied to longitudinal person-firm data. I explore four different approaches to producing a random sample of persons, firms, and jobs, quantify-

ing the advantages and disadvantages of each method. For example, using LEHD data I show that quarterly average earnings or employment totals can be estimated most precisely by taking a random sample of jobs each quarter. However, traditional sampling methods have a major drawback, they do not preserve a useful representation of worker-firm mobility patterns.

More complex estimators often require the connected firm graph generated by worker mobility. Using graph theory combined with a Metropolis-Hastings random walk I produce connected subgraphs that are themselves random samples of the entire dataset. To get an estimate of the efficiency of the new sampling algorithm, I compare it with the four traditional sampling approaches mentioned above. In addition, I estimate a fixed person firm effect earnings model and compare the performance of the connected subgraph(s) estimates with the estimates produced using the full dataset.

Estimates of External Bias in Impact Evaluations That Select Sites Purposively
Stephen Bell and Robert Olsen (Abt Associates, USA) and Larry Orr and Elizabeth Stuart (Johns Hopkins University, USA)

Substantial bias may result when evaluations of government programs purposively—rather than randomly—select the sites from which data will be collected and estimates derived (see Olsen, Orr, Bell, and Stuart, 2010, for a formal expression of that bias). To estimate "external validity" bias of this sort, we use data from a recent evaluation of the Reading First program covering all school districts in 15 states. From these data, we compute a benchmark estimate of Reading First's impact on student achievement in all districts in all 15 states. We then compare this benchmark to estimates of impact from the subset of districts that participated in a different education impact evaluation which selected districts purposively. The difference between the purposive sample estimate and the benchmark estimate is an estimate of the external bias that would have resulted if Reading First had been evaluated in the purposively-chosen sites alone. Measures of this bias are computed for 12 different purposive-sample evaluations taken from the education literature. The result is robust empirical evidence on the question of whether policy impact evaluations based on purposive samples have adequate external validity.

CONCURRENT SESSION X-A:
BAYESIAN STATISTICAL METHODS

Calibrated Bayes: Spanning the Divide Between Frequentist and Bayesian Inference
Roderick Little (U.S. Census Bureau and University of Michigan-Ann Arbor, USA)

The lack of an agreed inferential basis for statistics has negative implications for the practice and status of statistics in industry, science and government. Some illustrations and implications of the Bayes/frequentist rift, specific and general, are offered, focusing particularly on the extreme case of sample survey inference. An analysis of strengths and weaknesses suggests that the "calibrated Bayes" approach of Box, Rubin and others combines the best features of Bayesian and frequentist statistics. In this approach inferences are Bayesian and hence model-based, but model formation and assessment can and should involve frequentist ideas. Some implications for the analysis of complex survey data are offered.

A Semi-Parametric Approach to Account for Complex Designs in Multiple Imputation
Hanzhi Zhou, Trivellore Raghunathan, and Michael Elliott (University of Michigan, USA)

Multiple imputation (MI) has become one of leading approaches in dealing with missing data in survey research. However, existing software packages and procedures typically do not incorporate complex sample design features in the imputation process. Researcher has demonstrated that implementation of MI based on simple random sampling (SRS) assumption can cause severe bias in estimation and hence invalid inferences, especially when the design features are highly related to survey variables of interest (Reiter et al. 2006). Recent work to accommodate complex sample designs in imputation has focused on model-based methods which directly model the complex design features in the formulation of the imputation model (Schenker et al. 2006). In this paper, we propose a semi-parametric procedure as an alternative approach to incorporate complex sampling designs in MI. Specifically, we divide the imputation process into two stages: the complex feature of the survey design (including weights and clusters) is fully accounted for at the first stage, which is accomplished by applying a nonparametric method to generate a series of synthetic datasets; we then perform conventional parametric MI for missing data at the second stage using readily available imputation software designed for an SRS sample. A new combining rule for the point and variance estimates is derived to make valid inferences based on the two-stage procedure. Using health survey data from the Behavior Risk Factor Surveillance System, we evaluate the proposed method with a simulation study and compare it with the model-based method with respect to complete data analysis. Results show that the proposed method has better confidence interval coverage and is more efficient than the model-based method.

Nonresponse Adjustment Based on Auxiliary Variables Subject to Error

Brady West (University of Michigan-Ann Arbor, USA) and Roderick Little (U.S. Census Bureau and University of Michigan-Ann Arbor, USA)

Effective unit nonresponse adjustments require auxiliary variables that are associated with both key survey variables and the propensity to respond. Such auxiliary variables are rare in survey practice. Promising candidates include interviewer observations on sample units and linked auxiliary variables from commercially available household databases. These variables are prone to measurement error, and as a result, the assumption of missing at random (MAR) that underlies standard weighting or imputation nonresponse adjustments is violated when missingness depends on the true values of these variables, leading to biased estimates of survey quantities.

This study proposes estimators of means for this situation based on pattern-mixture models (PMMs). The set of auxiliary variables, measured for respondents and nonrespondents, includes one (say) that is an error-prone proxy for a survey covariate (). Values of and survey variables are recorded for survey respondents but missing for nonrespondents. Means of the survey variables are estimated under a PMM that assumes that nonresponse depends on C and, rather than the less plausible MAR assumption that nonresponse depends on C and. Bayesian and multiple imputation estimates are developed, and compare favorably in simulation studies with imputation and weighting estimates assuming MAR, and estimates based on complete case analysis. Applications of the method to data from two real surveys are presented, together with R code for the proposed method.

CONCURRENT SESSION X-B:

LINKING AND QUALITY OF ADMINISTRATIVE DATA

Pulling Together: Linking Unemployment Insurance and Supplemental Nutrition Assistance Program Administrative Data to Study Effects of the Great Recession

Theresa Anderson (George Washington University, USA and Economic Research Service), John Kirlin (Economic Research Service, USA), and Michael Wiseman (George Washington University, USA)

The “Great Recession” of 2007-2009 created record levels of unemployment and record durations of joblessness for many workers. Many newly unemployed, experienced workers turn to the Unemployment Insurance program (UI) as a first fallback against economic hardship. As UI benefits become exhausted, workers may increasingly turn to other safety net elements for support, including the Supplemental Nutrition Assistance Program (SNAP). The sequence of use and interaction of the SNAP and UI programs are of great interest to policymakers. Survey data are often untimely and are known to suffer from misreporting of receipt of SNAP or UI benefits. To gain a better understanding of SNAP/UI relationships, the USDA Economic Research Service (ERS) sponsored a project in which researchers in seven states linked the states’ SNAP and UI administrative data for 2006-2009; databases for both programs are maintained at the state level. The project examines the cross-sectional and longitudinal relationships between these two programs that provide crucial income support in times of economic challenge.

This paper describes the SNAP-UI data linkage project and the many barriers and challenges confronted by ERS and its partnering state researchers throughout the process. These challenges include dealing with the diversity of state administrative systems, gaining access to the data while ensuring security and confidentiality, and developing a common analysis framework. This paper focuses on organizational and institutional issues. Sample tabulations of the merged SNAP/UI dataset illustrate outcomes. The experience chronicled and precedents established in this project should prove useful to policy analysts, policy makers, program administrators, and others who seek to use state administrative data for improving program achievement in circumstances in which the interaction of programs is likely to have significant consequences for outcomes.

Study of Factors Affecting Record Linkage in Federal Statistical Databases

Yuan Zhao and Michael Larsen (The George Washington University, USA)

Record linkage, or exact file matching, consists of bringing together records in two or more files on the same population. Files are linked for the purposes of creating a larger database, enabling analyses that would otherwise not be possible, and counting the population. When unique, error-free identification codes are not available on both files, then record linkage can be accomplished through probabilistic methods. When implementing matching algorithms, one must choose matching variables, define for each variable what it means to agree or disagree, choose blocking factors that restrict the space of comparison pairs, and decide the level of evidence required to declare that a pair of record is probable match. The National Center for Health Statistics (NCHS) uses record linkage to match surveys, such as the National Health Interview Survey (NHIS) or the National

Health and Nutrition Examination Survey (NHANES) to the National Death Index (NDI) for studies of mortality and morbidity. Based on files simulated to be similar to the NDI and a NCHS health survey, choices that affect the performance of probabilistic record linkage are studied. The impact of several choices as well as file sizes, recording errors, and analysis methods (e.g., Bayesian versus likelihood) are compared. The work has direct relevance for improving and evaluating record linkage operations in the federal statistical system.

Qualitative Assessment of Administrative Records: The Case of State Prisons' Records
Anna Chan (U.S. Census Bureau)

The use of administrative records to collect survey and census data is common when the target sample resides in group quarters (GQs). These are places where people live or stay, in a group living arrangement that is owned or managed by an entity or organization providing housing and/or services for residents.

In 2000, almost 52% of decennial census data collected from GQ population came from administrative records (Jonas 2003). Although records have been used for data collection for quite some time, knowledge regarding the accuracy of administrative data and its impact on survey statistics is limited. Recent research (Belton, Smith & Encarnación 2010) has shown that surveys completed in GQs by using administrative records resulted in higher item allocation rates (can be as high as 71.5 %) than data obtained from self responses. In this study, we examined the data accuracy and completeness of administrative records obtained from correctional facilities, which comprised of 25 percent of all GQ population in the United States.

Findings for this study are based on data collected from qualitative research studies conducted in 2010 and 2011 in three large state prisons in two western states. Inmate rosters were obtained from the prisons and about 200 inmates were randomly selected from the lists to participate in the studies. Inmates were given a self-administered questionnaire containing items that were in the Census 2010 forms. Administrative records containing the name, date of birth, race and or ethnicity were obtained and were compared to the self responses of the inmates.

Our preliminary results suggest that data provided by administrative records are not always accurate or complete even when we were only collecting basic demographics information with our survey instrument. Based on our findings, we will also discuss issues with collecting survey data using self-administered questionnaire versus using administrative records.

Methods of Assigning Race and Hispanic Origin to Births from Vital Statistics Data
Christine Guarneri and Christopher Dick (U.S. Census Bureau)

Birth certificates collect information on the race and Hispanic origin of the mother and father, but this information is not collected for the child. In order to utilize birth certificate data in research by these demographic characteristics, a method must be developed for assigning race and Hispanic origin to each birth. In the past, assignments have been made solely based on the race or Hispanic origin of the mother or father. However, with the number of multiple race/Hispanic origin marriages believed to be increasing over time, the validity of these methods is called into question. The "Kid Link File" (KLF) was developed to address this issue and, using Census or survey data, links characteristics of children aged 0 to 17 with the characteristics of their parents. This format allows the reported characteristics of the child—such as race and Hispanic origin—to be connected to parents' reporting. Patterns of reporting may then be used to proportionally assign race and Hispanic origin to the birth data. This paper describes the process of developing this file based on various sources of data, including Census 2000, the American Community Survey (ACS), and the 2010 Census. Developing a KLF based on multiple datasets makes it possible to evaluate differences in reporting and potentially calculate changes in race and Hispanic origin identification over time.

The application of kid link proportions in the production of estimates and projections is also explored. Specifically, kid link proportions are used to assign race and Hispanic origin to projected births. Additionally, kid link proportions were recently used in the 2010 Demographic Analysis to estimate the population according to the race categories of Black and non-Black and ethnicity categories of Hispanic and non-Hispanic.

CONCURRENT SESSION X-C:

DETECTING AND CORRECTING NONRESPONSE ERROR AND BIAS IN LARGE SAMPLE GOVERNMENT DATA COLLECTIONS IN THE UNITED STATES AND CANADA

Total Survey Error in the American Time Use Survey

John Dixon and Brian Meekins (Bureau of Labor Statistics, USA)

The American Time Use Survey (ATUS) asks respondents to list all activities and their duration of the day immediately prior to the day of interview. The ATUS sampling frame consists of respondents who complete the final wave of the Current Population Survey (CPS). Using CPS responses and information from the CPS contact history instrument (CHI) together with variables from the ATUS telephone contact history, the authors develop a model that incorporates both nonresponse bias and measurement error in ATUS time estimates.

Experimental Design for Nonresponse Follow-up of Electronic Questionnaire Survey

Joanne Leung, Jeannine Claveau, and Claude Turmelle (Statistics Canada)

At Statistics Canada, business survey collection consists of many steps and uses more than one collection mode. Many business surveys continue to use mail questionnaires for initial data gathering. Also, telephone follow-up is conducted to resolve edit problems with mailed back questionnaires and to collect data from respondents who have not returned the questionnaires after a pre-specified period.

Currently, Statistics Canada is undertaking a general restructuring of its business statistics programs. One of its goals is to let electronic data collection become the principal mode of collection for business surveys. Until now, follow-up methods used for electronic questionnaire were based on paper collection methods. International and Statistics Canada experiences showed that electronic respondents have different follow-up patterns.

In order to establish standard collection follow-up strategy for electronic questionnaire, Statistics Canada builds an experimental design to compare different non-response follow-up methods (combining fax, phone, and e-mail reminder). An embedded balanced factorial design is used for this experiment. Analysis of paradata outputs such as response rates, time spent on follow-up and time of response would permit us to compare the different methods. We hope to bring some conclusion about the importance of having phone contact versus e-mail reminder, to know how far we can get if we use only e-mail reminders and to establish a follow-up strategy which is the most efficient. This presentation provides results of this experiment.

Comparison of the American Community Survey Voluntary Versus Mandatory Estimates

Karen King, Michael Starsinic, and Alfredo Navarro (U.S. Census Bureau)

The American Community Survey (ACS) collects essentially the same detailed demographic, housing, and socio-economic data as were collected on the decennial census long form questionnaire using similar mandatory collection methods.

We were requested by Congress in 2002 to conduct research to determine if the ACS could be implemented as a voluntary collection survey. In 2003, we designed and developed a test to assess the effect of a switch from a mandatory to a voluntary ACS on

feasibility – by looking at costs and workloads

public reaction – by analyzing response rates by mode

quality – in terms of sampling error and levels of unit and item response

The 2003 test of the voluntary data collection method was not designed to assess if estimates produced from data collected using a voluntary method would differ from estimates from data collected using the mandatory collection method in the ACS. Additional analysis of the data that were collected in the 2003 collection year might shed some light on the question of whether a change to a voluntary ACS would result in different ACS estimates.

Our technique to answer this important question was to create two sets of weighted annualized 2003 estimates – one based on data collected using the voluntary collection methods, and the other based on data collected using the mandatory collection methods.

We selected a subset of characteristics to focus on, compared the estimates by collection method, and then assessed if the resultant estimates differ beyond sampling variability. We also implemented quantitative analysis to assess the relative and absolute change in the distributions. This paper shows some of the results from this analysis.

FCSM RESEARCH CONFERENCE AUTHOR AND COAUTHOR INDEX

NAME	SESSION
John Finamore	IV-B
Carl Fisher	VIII-D
Tim Flanigan	VI-B
Anne Flatness	I-D
Mary Foulkes	V-A
Scott Fricker	III-D, IX-B
Marek Fuchs	III-B
Graton Gathright	V-B, IX-C
Emily Geisen	IV-C
Michael Gerling	VIII-D
Joe Gfroerer	VII-A
Bula Ghose	IX-C
Scott Ginder	VI-B
Kaj Gittings	VIII-C
Cordell Golden	VII-A
Jeffrey Gonder	VI-D
Jeffrey Gonzalez	IX-B
Yarissa Gonzalez	V-C
Alfred Gottschalck	VII-B
Treda Grayson	VI-C
Joseph Greenblott	VIII-D
Karen Grigorian	I-A
Christine Guarneri	X-B
Thomas Hager	II-B
Margaret Jean Hall	VIII-B
Karen Hamrick	VII-A
James Harris	VIII-D
Susan Harris	III-C
Lauren Harris-Kojetin	II-C
Jim Hartman	IX-D
Sarah Hastedt	I-A
Sarra Hedden	VII-A
David Hedengren	VII-B
Kimberly Henry	V-A
Steven Hill	IV-C
Pam Hird	VIII-D
Carma Hogue	III-C
Richard Hough	II-C, VIII-B
David Howard	VII-B
Douglass Huang	IV-D
David Cunningham Hunter	V-C
Tracy Hunt-White	I-A, IV-B
Howard Iams	IV-C
Vincent Iannacchione	V-A
Donsig Jang	VI-C
Matthew Jans	V-C
Natasha Janson	IV-B
Eric Jensen	I-D
Dean Judson	I-D
Alan Karr	II-A
Anthony Kassekert	I-D
Arnold Katz	III-D
Andrew Keller	I-C
Bert Kestenbaum	VII-C
Hans Kiesl	VIII-A
Karen King	X-C
Susan Kinsey	VIII-D
John Kirlin	IX-C, X-B
Lawrence Kleinman	IX-A
Jacob Klerman	III-B
Brandon Kopp	III-D, IV-C
Phillip Kott	III-B
Braedyn Kromer	VII-B
Flora Lan	VI-C
David Langdon	II-D
Michael Larsen	V-A, VI-A, X-B
James Lawrence	IX-D
Patricia LeBaron	VI-B
Sarah Lehmann	VI-C
Joanne Leung	X-C
Li Leung	VIII-A
Gad Levanon	II-D
Jianzhu Li	I-B
Dan Liao	III-B, VI-A
Judy Lieberman	VIII-D
Joanna Fane Lineback	VI-B
Roderick Little	X-A
John Loft	VI-B

NAME	SESSION
Luanne Lohr	VIII-B
Sharon Lohr	I-B
Sharon Long	I-D
David Loughran	III-B
Stephen Mack	VI-A
Donald Malec	IV-A
Saida Mamedova	II-C
Jean-Claude Manini	II-D
Matthew Marlay	V-B
Paco Martorell	III-B
Peter Mateyka	V-B
Hope Smiley McDonald	VI-B
James McGarity	IV-D
Kevin McKinney	VIII-C, IX-D
Elke McLaren	VII-B
Cameron McPhee	I-A, II-C
Brian Meekins	X-C
Johan Melendez	VII-A
Peter Meyer	IV-D
Robert Meyer	V-D
Lynn Milan	I-A
Morgen Miller	IX-A
Lisa Mirel	VII-A, IX-C
Susan Mitchell	V-C
Mary Mulry	III-D
Patricia Mundy	VIII-D
Elaine Murakami	VI-D
Alfredo Navarro	I-C, X-C
Charles Nelson	IX-C
Margaret Noonan	VI-B
Edward Norton	IX-A
Amy O'Hara	III-A
Barbara O'Hare	V-C
Aniekan Okon	VII-B
Broderick Oliver	V-C, VI-B
Robert Olsen	IX-D
Isaiah O'Rear	VII-B
A. Elizabeth Ormson	V-A
Larry Orr	IX-D
Jane Osburn	I-C
Maria Owings	VIII-B
Ataman Ozyildirim	II-D
Joseph Parsons	VI-C
Joanne Pascale	III-A
Laura Paszkiewicz	III-C
K. Bradley Paxton	II-B, IV-D
Denise Pepe	VII-B
Daniel Perez-Lopez	IX-B
Michael Planty	I-B
Patrick Purcell	IV-C
Siyu Qing	V-A
Trivellore Raghunathan	X-A
Lori Reeder	V-B
Marianne Reifschneider	III-C
Benjamin Reist	VI-A
Marko Roos	VII-C
Eve Rose	VII-A
Janet Rosenbaum	VII-A
Judy Sangl	VIII-A
Kathleen Santos	V-A
Bryan Schar	IX-D
Fritz Scheuren	IX-C
Peter Schilling	VII-C
Kenneth Schoendorf	VII-A
Daniel Schroeder	IX-C
Frederic Selck	II-B
Manisha Sengupta	VI-B
Bonnie Shook-Sa	V-A
Peter Siegel	I-A
Avinash Singh	VIII-C
Robyn Sirkis	III-C
Jeremy Skog	IX-C
Aleksandra Slavkovi	VIII-C
Marla Smith	VI-C
Ger Snijkers	VII-B
Matthew Soldner	IV-B
Rachel Soloveichik	I-D
Curry Spain	VIII-D

NAME	SESSION
Steven Spiwak	IV-D
Michael Starsinic	X-C
Bryce Stephens	VIII-C
Kenneth Steve	VIII-A
Martha Stinson	V-B, IX-C
Derek Stone	VIII-D
M. Christopher Stringer	IV-C
Elizabeth Stuart	IX-D
Kendall Swenson	IX-C
Lucilla Tan	IX-B
Jennelyn Tanchua	II-D
Sarah Tekansik	VI-A
Michael Thieme	IV-B
Katherine Jenny Thompson	III-B
John Tillinghast	III-D
Nhien To	III-D, IX-B
Kevin Tolliver	V-D
Bac Tran	IV-A
Neil Tseng	IV-C
Joan Turek	IX-C
Claude Turmelle	X-C
Joanna Turner	I-C
Richard Valliant	V-A
Merijn van Pelt	VII-C
Shinu Verghese	IX-C
Lars Vilhuber	VIII-C
Matthew Virgile	VII-C
Pieter Vlag	II-D

NAME	SESSION
Rachael Walsh	IX-B
Jason Wang	IX-A
Katrina Washington	III-B
Shawna Waugh	VI-C
Lynn Weidman	I-C
Edward Welniak	IX-C
Brady West	X-A
Gloria Wheatcroft	VII-A
Diane Willimack	VI-B, VII-B
David Wilson	VI-D
Jennifer Wine	I-A, IV-B
Gina Wingood	VII-A
Michael Wiseman	IX-C, X-B
Kirk Wolter	V-A
Yong Ming Jeffrey Woo	VIII-C
Simon Woodcock	VIII-C
David Woodwell	VIII-B
Monica Wroblewski	III-A
Star Ying	IX-D
Audra Zakzeski	IV-D
Brian Zamperini	VII-C
Jonathan Zenilman	VII-A
Yuan Zhao	X-B
Beilei Zhou	V-A
Hanzhi Zhou	X-A
Marc Zodet	IV-C, VIII-A
Sara Zuckerbraun	VI-B
Samuel Zuvekas	IV-C

Reports Available in the Statistical Policy Working Paper Series

1. Report on Statistics for Allocation of Funds, 1978 (NTIS PB86-211521/AS)
2. Report on Statistical Disclosure and Disclosure-Avoidance Techniques, 1978 (NTIS PB86-211539/AS)
3. An Error Profile: Employment as Measured by the Current Population Survey, 1978 (NTIS PB86-214269/AS)
4. Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics, 1978 (NTIS PB86-211547/AS)
5. Report on Exact and Statistical Matching Techniques, 1980 (NTIS PB86-215829/AS)
6. Report on Statistical Uses of Administrative Records, 1980 (NTIS PB86-214285/AS)
7. An Interagency Review of Time-Series Revision Policies, 1982 (NTIS PB86-232451/AS)
8. Statistical Interagency Agreements, 1982 (NTIS PB86-230570/AS)
9. Contracting for Surveys, 1983 (NTIS PB83-233148)
10. Approaches to Developing Questionnaires, 1983 (NTIS PB84-105055)
11. A Review of Industry Coding Systems, 1984 (NTIS PB84-135276)
12. The Role of Telephone Data Collection in Federal Statistics, 1984 (NTIS PB85-105971)
13. Federal Longitudinal Surveys, 1986 (NTIS PB86-139730)
14. Workshop on Statistical Uses of Microcomputers in Federal Agencies, 1987 (NTIS PB87-166393)
15. Quality in Establishment Surveys, 1988 (NTIS PB88-232921)
16. A Comparative Study of Reporting Units in Selected Employer Data Systems, 1990 (NTIS PB90-205238)
17. Survey Coverage, 1990 (NTIS PB90-205246)
18. Data Editing in Federal Statistical Agencies, 1990 (NTIS PB90-205253)
19. Computer Assisted Survey Information Collection, 1990 (NTIS PB90-205261)
20. Seminar on Quality of Federal Data, 1991 (NTIS PB91-142414)
21. Indirect Estimators in Federal Programs, 1993 (NTIS PB93-209294)
22. Report on Statistical Disclosure Limitation Methodology, 1994 (NTIS PB94-165305)
23. Seminar on New Directions in Statistical Methodology, 1995 (NTIS PB95-182978)
24. Electronic Dissemination of Statistical Data, 1995 (NTIS PB96-121629)
25. Data Editing Workshop and Exposition, 1996 (NTIS PB97-104624)
26. Seminar on Statistical Methodology in the Public Service, 1997 (NTIS PB97-162580)
27. Training for the Future: Addressing Tomorrows Survey Tasks, 1998 (NTIS PB99-102576)
28. Seminar on Interagency Coordination and Cooperation, 1999 (NTIS PB99-132029)
29. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 1999 (NTIS PB99-166795)
30. 1999 Federal Committee on Statistical Methodology Research Conference: Complete Proceedings, 2000 (NTIS PB2000-105886)
31. Measuring and Reporting Sources of Error in Surveys, 2001 (NTIS PB2001-104329)
32. Seminar on Integrating Federal Statistical Information and Processes, 2001 (NTIS PB2001-104626)
33. Seminar on the Funding Opportunity in Survey Research, 2001 (NTIS PB2001-108851)
34. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 2001 (NTIS PB2002-100103)
35. Seminar on Challenges to the Federal Statistical System in Fostering Access to Statistics, 2002
36. Seminar on the Funding Opportunity in Survey Research, 2003
37. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 2003
38. Summary Report of the FCSM-GSS Workshop on Web-Based Data Collection, 2004
39. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 200540.
40. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 2007
41. Federal Committee on Statistical Methodology Research Conference (Conference Papers), 2009

Copies of these working papers may be ordered from NTIS Document Sales, 5285 Port Royal Road, Springfield, VA 22161; Telephone: 1-800-553-6847. The Statistical Policy Working Paper series is also available electronically from FCSM's Web site <http://www.fcsm.gov>.

NOTES

NOTES