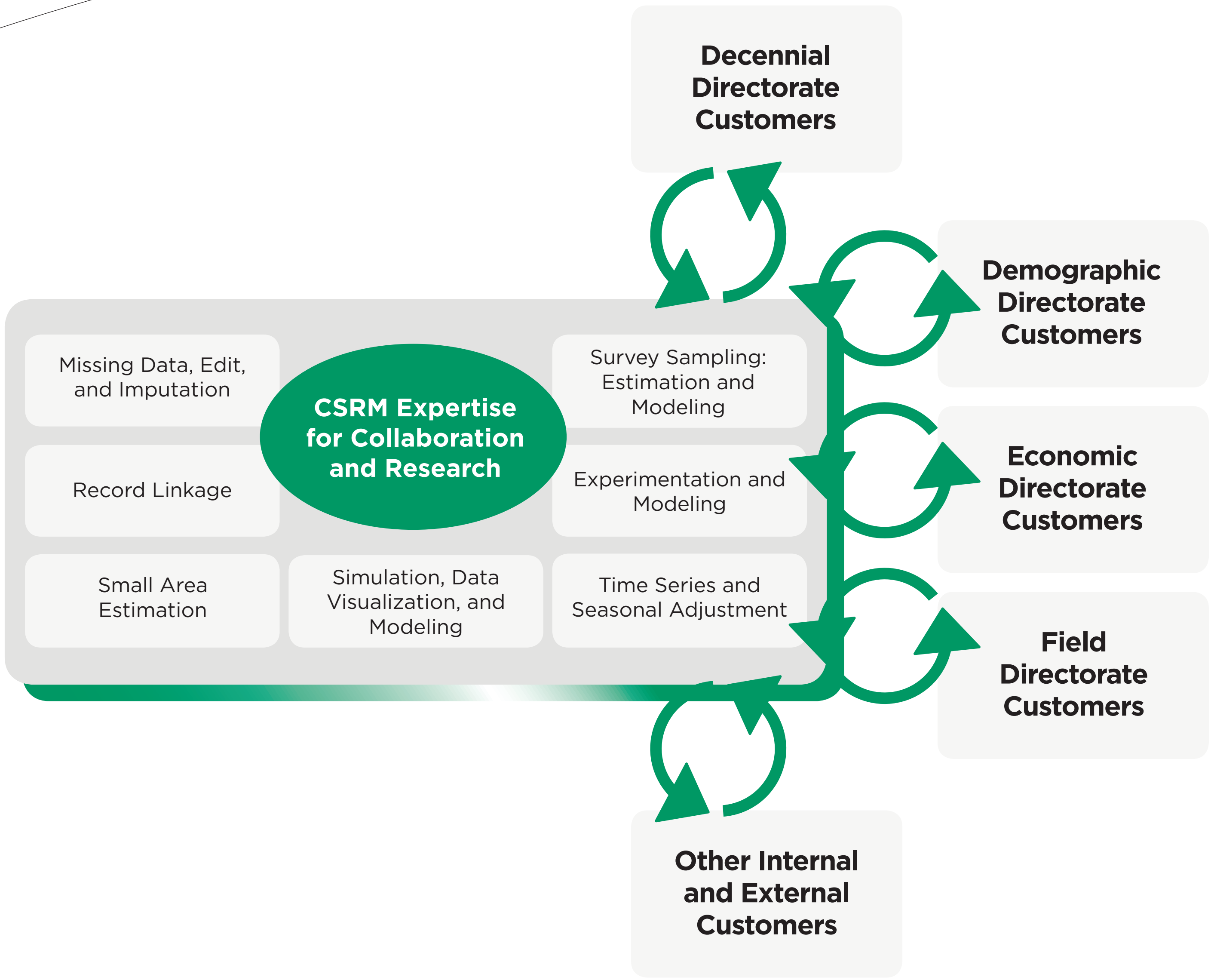
Annual Report of the Center for Statistical Research and Methodology

Research and Methodology Directorate

Fiscal Year 2017



Decennial Directorate Customers

Demographic Directorate Customers

Economic Directorate Customers

Field Directorate Customers

Other Internal and External Customers

CSRM Expertise for Collaboration and Research

Missing Data, Edit, and Imputation

Record Linkage

Small Area Estimation

Simulation, Data Visualization, and Modeling

Survey Sampling: Estimation and Modeling

Experimentation and Modeling

Time Series and Seasonal Adjustment

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Since August 1, 1933—

"… As the major figures from the American Statistical Association (ASA), Social Science Research Council, and new Roosevelt academic advisors discussed the statistical needs of the nation in the spring of 1933, it became clear that the new programs—in particular the National Recovery Administration—would require substantial amounts of data and coordination among statistical programs. Thus in June of 1933, the ASA and the Social Science Research Council officially created the Committee on Government Statistics and Information Services (COGSIS) to serve the statistical needs of the Agriculture, Commerce, Labor, and Interior departments … COGSIS set … goals in the field of federal statistics … (It) wanted new statistical programs—for example, to measure unemployment and address the needs of the unemployed … (It) wanted a coordinating agency to oversee all statistical programs, and (it) wanted to see statistical research and experimentation organized within the federal government … In August 1933 Stuart A. Rice, President of the ASA and acting chair of COGSIS, … (became) assistant director of the (Census) Bureau. Joseph Hill (who had been at the Census Bureau since 1900 and who provided the concepts and early theory for what is now the methodology for apportioning the seats in the U.S. House of Representatives) … became the head of the new Division of Statistical Research … Hill could use his considerable expertise to achieve (a) COGSIS goal: the creation of a research arm within the Bureau …"

Source: Anderson, M. (1988), *The American Census: A Social History,* New Haven: Yale University Press.

Among others and since August 1, 1933, the Statistical Research Division has been a key catalyst for improvements in census taking and sample survey methodology through research at the U.S. Census Bureau. The introduction of major themes for some of this methodological research and development, where staff of the Statistical Research Division[1] played significant roles, began roughly as noted—

- **Early Years (1933–1960s):** sampling (measurement of unemployment and 1940 Census); probability sampling theory; nonsampling error research; computing; and data capture.
- **1960s–1980s:** self-enumeration; social and behavioral sciences (questionnaire design, measurement error, interviewer selection and training, nonresponse, etc.); undercount measurement, especially at small levels of geography; time series; and seasonal adjustment.
- **1980s–Early 1990s:** undercount measurement and adjustment; ethnography; record linkage; and confidentiality and disclosure avoidance.
- **Mid 1990s–Present:** small area estimation; missing data and imputation; usability (human-computer interaction); and linguistics, languages, and translations.

At the beginning of FY 2011, most of the Statistical Research Division became known as the Center for Statistical Research and Methodology. In particular, with the establishment of the Research and Methodology Directorate, the Center for Survey Measurement and the Center for Disclosure Avoidance Research were separated from the Statistical Research Division, and the remaining unit's name became the Center for Statistical Research and Methodology.

[1] The Research Center for Measurement Methods joined the Statistical Research Division in 1980. In addition to a strong interest in sampling and estimation methodology, research largely carried out by mathematical statisticians, the division also has a long tradition of nonsampling error research, largely led by social scientists. Until the late 1970s, research in this domain (e.g., questionnaire design, measurement error, interviewer selection and training, and nonresponse) was carried out in the division's Response Research Staff. Around 1979 this staff split off from the division and became the Center for Human Factors Research. The new center underwent two name changes—first, to the Center for Social Science Research in 1980, and then, in 1983, to the Center for Survey Methods Research before rejoining the division in 1994.

U.S. Census Bureau
Center for Statistical Research and Methodology
Room 5K108
4600 Silver Hill Road
Washington, DC 20233
301-763-1702

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

We help the Census Bureau improve its processes and products. For fiscal year 2017, this report is an accounting of our work and our results.

Center for Statistical Research & Methodology
www.census.gov/srd/csrm

Highlights of What We Did...

As a technical resource for the Census Bureau, each researcher in our center is asked to do three things: *collaboration/consulting*, *research*, and *professional activities and development*. We serve as members on teams for a variety of Census Bureau projects and/or subprojects.

Highlights of a selected sampling of the many activities and results in which the Center for Statistical Research and Methodology staff members made contributions during FY 2017 follow, and more details are provided within subsequent pages of this report:

- *Missing Data, Edit, and Imputation*: (1) showed how to use log-linear models to improve the efficiency (reducing the sampling error) of longitudinal estimators of quarterly changes in labor force status and healthcare coverage and showed how these estimators can be implemented for labor force and healthcare coverage measurement using data from the Survey of Income and Program Participation; (2) researched, developed, and evaluated methods for raking balance complexes in the Quarterly Financial Report (QFR) when items are negative or there is subtraction in the balance complexes.

- *Record Linkage:* (1) applied and made updates to record linkage software.

- *Small Area Estimation*: (1) developed a unit-level small area projection model to estimate state level disability rates from the Survey of Income and Program Participation augmented by the American Community Survey; (2) developed a small area model for tracts using a generalized Poisson distribution.

- *Survey Sampling-Estimation and Modeling*: (1) analyzed the potential of market segmentation from an external source to provide useful information to the 2020 Census communications campaign; (2) developed alternative methods of ranking census tracts for their potential to be influenced by decennial census outreach operations; (3) developed a conceptual framework for designing, testing, and evaluating policies for curtailment of follow-up in household sample surveys; modelled on research done with and by American Community Survey staff on CATI and CAPI contact history paradata; (4) completed theory and visualizations (CRAN) in comparing several populations with overlapping and non-overlapping confidence intervals.

- *Time Series and Seasonal Adjustment*: (1) completed work on fitting vector moving averages by establishing stability properties and proving asymptotic properties; (2) constructed weather regressors, and incorporated these into seasonal adjustment methodology, with application to regional construction series; (3) developed new algorithms for forecasting and signal extraction of multivariate time series; (4) investigated the causation of residual seasonality in indirect seasonal adjustments of time series, and developed a methodology for reconciliation of direct adjustments.

- *Experimentation and Statistical Modeling:* (1) released an updated COMPoissonReg package on CRAN, which supports both zero inflated and standard COM-Poisson regression modeling; (2) developed a prototype R package for spatio-temporal change of support modeling (Public American Community Survey data were used to demonstrate that estimates on non-standard geographies and lookback periods, i.e., not released by the Census Bureau, could be produced by data users).

- *Simulation and Statistical Modeling:* (1) developed new methodology that uses the principle of sufficiency to create synthetic data whose distribution is identical to the distribution of the original data under the normal linear regression model; (2) applied small area estimation methodology to compute state and county level estimates based on the Tobacco Use Supplement to the Current Population Survey for the National Cancer Institute.

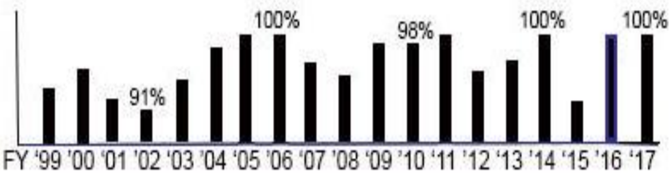
- *SUMMER AT CENSUS:* Sponsored, with divisions around the Census Bureau, scholarly, short-term visits by 31 researchers/leaders who collaborated extensively with us and presented seminars on their research. For a list of the 2017 *SUMMER AT CENSUS* scholars, see *http://www.census.gov/research/summer_at_census/*.

How Did We[1] Do...

For the 19th year, we received feedback from our sponsors. Near the end of fiscal year 2017, our efforts on 29 of our program (Decennial, Demographic, Economic, Administration, External) sponsored projects/subprojects with substantial activity and progress and sponsor feedback (Appendix A) were measured by use of a Project Performance Measurement Questionnaire (Appendix B). Responses to all 29 questionnaires were obtained with the following results (The graph associated with each measure shows the performance measure over the last 19 fiscal years):

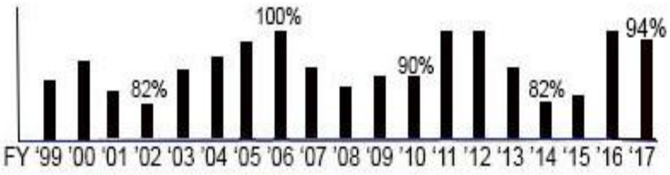
Measure 1. Overall, Work Met Expectations

Percent of FY2017 Program Sponsored Projects/Subprojects where sponsors reported that overall work met their expectations (agree or strongly agree) (29 out of 29) …….…………..... 100%



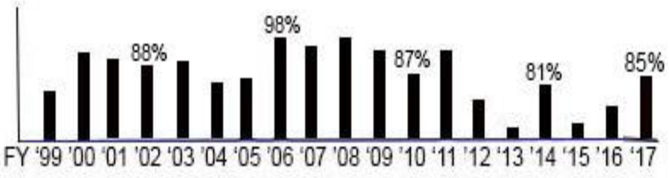
Measure 2. Established Major Deadlines Met

Percent of FY2017 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met (15 out of 16 responses) ……......……..…..…... 94%



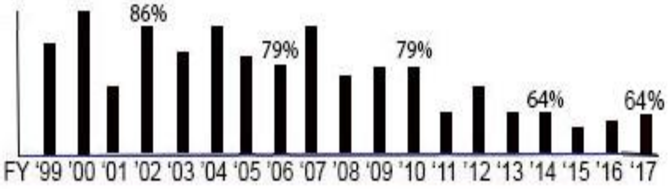
Measure 3a. At Least One Improved Method, Developed Technique , Solution, or New Insight

Percent of FY2017 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight (22 out of 26 responses) ………… 85%



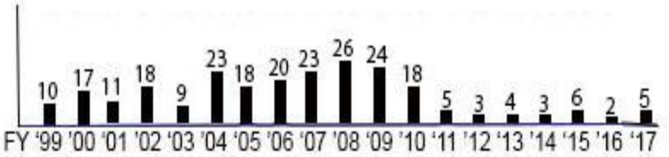
Measure 3b. Plans for Implementation

Of these FY2017 Program Sponsored Projects/Subprojects reporting at least one improved method, technique developed, solution, or new insight, the percent with plans for implementation (14 out of 22 responses) ………………….. 64%



Measure 4. Predict Cost Efficiencies

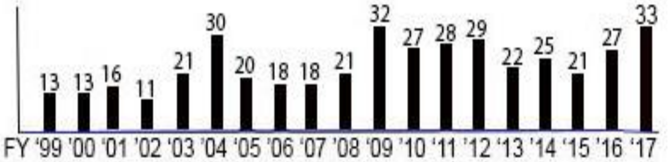
Number of FY2017 Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency" ………..…… 5



From Section 3 of this ANNUAL REPORT, we also have:

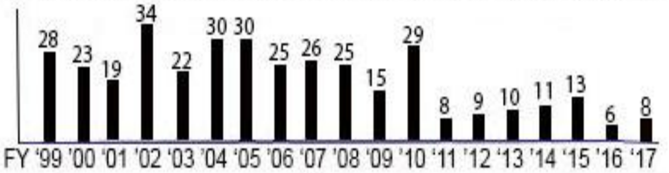
Measure 5. Journal Articles, Publications

Number of peer reviewed journal publications documenting research that appeared (18) or were accepted (15) in FY2017 ………………………………………………………………... 33



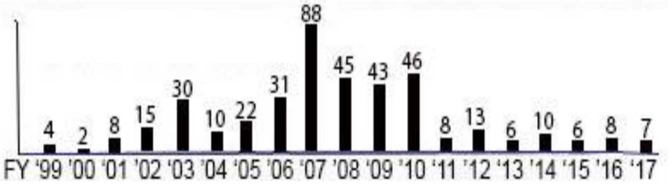
Measure 6. Proceedings, Publications

Number of proceedings publications documenting research that appeared in FY2017 ………………………………..………. 8



Measure 7. Center Research Reports/Studies, Publications

Number of center research reports/studies publications documenting research that appeared in FY2017 …………….. 7



Each completed questionnaire is shared with appropriate staff to help improve our future efforts.

[1] Reorganized from Statistical Research Division to Center for Statistical Research and Methodology, beginning in FY 2011.

TABLE OF CONTENTS

1. COLLABORATION

1.1 REDESIGNING FIELD OPERATIONS (Decennial Project 6650C23)

1.2 ADMINISTRATIVE RECORDS DATA (Decennial Project 6750C01)

1.3 DATA CODING, EDITING, AND IMPUTATION (Decennial Project 6550C01)

1.4 OPERATIONAL DESIGN (Decennial Project 6250C02)

A. Decennial Record Linkage

Description: Under this project, staff will provide advice, develop computer matching systems, and develop and perform analytic methods for adjusting statistical analyses for computer matching error with a decennial focus.

Highlights: During FY 2017, staff provided two write-ups related to a joint project between the Decennial Statistical Studies Division (DSSD) and the Center for Statistical Research and Methodology (CSRM). Staff provided a document describing the development of the Statistical Research Division (SRD) matching system used in the 1990, 2000, and 2010 Decennial Censuses. The document described some of the analytic/computer skills and the time needed by fifteen individuals (mostly DSSD) to write forty programs over four years. The document was background on how SRD/CSRM has participated in and developed new methods (particularly the Jaro-Winkler string comparator and methods for unsupervised learning in ~457 Census local offices) that were part of one of the main Decennial Census production systems.

Staff received 2010 Decennial data and began to combine the multiple files into a smaller number of files that is more appropriate to this research.

Staff: William Winkler (x34729), Emanuel Ben-David, Ned Porter

B. Coverage Measurement Research

Description: Staff members conduct research on model-based small area estimation of census coverage, and they consult and collaborate on modeling census coverage measurement (CCM).

Highlights: During FY 2017, staff attended weekly meetings for the Coverage Measurement full team and the Estimation sub-team. Additionally, staff has assisted in the creation of the research plans for the non-response weight adjustment research and the status of imputation research. Both of these research plans have been approved by the Decennial Research Objects and Methods (DROM) group and staff is assisting these two groups with implementing and interpreting the results.

Staff: Jerry Maples (x32873), Ryan Janicki, Eric Slud

C. Using 2010 Census Coverage Measurement Data to Compare Nonresponse Follow-up Proxy Responses with Administrative Records

Description: Research in preparation for the 2020 Census Nonresponse Follow-up (NRFU) investigates employing different contact strategies combined with the use of administrative records (AR) files in different ways in order to reduce the cost of the operation while maintaining data quality. Regardless of the contact strategy, the question arises as to whether the proxy responses are more accurate than ARs available for the NRFU housing units (HUs). The goal of this study is to use the results of the 2010 Census Coverage Measurement Program (CCM) to compare the accuracy of proxy responses for 2010 Census NRFU housing units in the CCM sample with the accuracy of the ARs available for the housing units.

Highlights: During FY 2017, staff completed revisions of a related article that subsequently appeared in the *Journal of Official Statistics* in June 2017. This project is now complete.

Staff: Mary Mulry (x31759)

D. Record Linkage Error-Rate Estimation Methods

Description: This project develops methods for estimating false-match and false-nonmatch rates without training data and with exceptionally small amounts of judiciously chosen training data. It also develops methods/software for adjusting statistical analyses of merged files when there is linkage error.

Highlights: Work is intended to extend Belin and Rubin (*JASA* 1995) and Larsen and Rubin (*JASA* 2001). Methods (with one exception) over the last thirty-plus years at other statistical agencies have failed because of the inability to develop expertise in name and address standardization, approximate string comparison, and very subtle methods of structuring data going into estimation software. The methods are very specific to pairs of files and often to subsets of files. For instance, for the ~500+ local offices in 2010 in which we have done Decennial Census matching, the parameters vary significantly even across adjacent regions.

During FY 2017, we received 2010 Decennial data which we are processing to put in simpler form so that our testing is more straightforward. We will be doing tests on the Center for Applied Technology (CAT) prior to moving to the main cloud machine for speed testing. Although *BigMatch* is known as by far the fastest software in the world (based on testing at the Isaac Newton Institute and other places), we hope doing work in the Bureau cloud will yield further improvements. (See subproject A.)

Staff: William E. Winkler (x34729), Emanuel Ben-David, Ned Porter, Tom Mule (DSSD)

E. Supplementing and Supporting Non-response with Administrative Records
Description: This project researches how to use administrative records in the planning, preparation, and implementation of nonresponse follow-up to significantly reduce decennial census cost while maintaining quality. The project is coordinated by one of the 2020 Census Integrated Project Teams.

Highlights: During FY 2017, staff continued to analyze the results of stepwise logistic regression models with topcoded Census Unedited File (CUF) household size as the dependent variable. The models used a national data file of 2010 Decennial Census nonresponse follow-up (NRFU) IDs. Staff fit models on a random 5% subsample and scored the models on the entire file. Staff began comparing results that include an additional four weeks of tax year 2009 IRS 1040 returns in the tax year 2009 IRS 1040 households to results without the additional weeks of returns. The two sets of results are similar except when the additional weeks of returns change the 2009 IRS 1040 household count from zero to at least one (for IDs where the undeliverable as addressed (UAA) flag is blank, indicating no UAA notice was received for that ID). Staff also examined the 2015 Indian Health Service (IHS) patient registration file to consider the usefulness of the IHS patient registration file as an additional file for modeling of household makeup or occupancy status. Presence on the IHS file with an assigned person identifier is already included in the current production modeling methodology. The number of persons in the IHS patient registration file in any individual cycle year is small compared to the Census and, apart from what is already in the modeling, the variables on the file do not appear likely to be useful for modeling purposes. Therefore, we concluded that the IHS patient registration file was not likely to be a useful addition to the general models. It is still possible that the file could be useful for a few localities, and there are variables that might be useful for characteristic imputation. Staff summarized the results of the analyses outlined above in draft documents that were sent to the Administrative Records Modeling subteam.

Staff: Michael Ikeda (x31756)

F. Identifying "Good" Administrative Records for 2020 Census NRFU Curtailment Targeting
Description: As part of the Census 2020 Administrative Records Modeling Team, staff are researching scenarios of nonresponse follow-up (NRFU) contact strategies and utilization of administrative records data. Staff want to identify scenarios that have reduction in NRFU workloads while still maintaining good census coverage. Staff are researching identification of "good" administrative records via models of the match between Census and administrative records person/address assignments for use in deciding which NRFU households to continue to contact and which to primary allocate. Staff are exploring various models, methods, and classification rules to determine a targeting strategy that obtains good Census coverage—and good characteristic enumeration—with the use of administrative records.

Highlights: During FY 2017, staff revised and received final acceptance of the paper "A Modeling Approach for Administrative Record Enumeration in the Decennial Census" to be published in a special issue of *Public Opinion Quarterly*. This paper compares classification methods for a person-place model for administrative records usage. Our staff and Decennial Statistical Studies Division (DSSD) staff revised and resubmitted the paper "A Distance Method for Administrative Record Modeling in the 2020 Census" to the *Journal of Official Statistics*. This paper documents a distance function approach based on predicted probabilities that were used to determine administrative record usage in the 2016 Census Test. Our staff and DSSD staff were awarded the 2017 Director's Award for Innovation for the work on these projects. Staff continued to attend meetings and provide input into research topics studied by the administrative records modeling team such as comparison of models using 2010 Census vs. ACS data and the analysis of USPS information in the 2016 Census test. Staff also began investigating real estate (MLS) data from CoreLogic—specifically assessing the relationship between housing unit vacancy and listing status.

Staff: Darcy Steeg Morris (x33989), Yves Thibaudeau

G. Evaluation of Response Error Using Administrative Records
Description: Censuses and their evaluations ask respondents to recall where they lived on Census Day, April 1. Some interviews for evaluations take place up to eleven months after this date. Respondents are asked when they moved to their current address, and the assumption has been that respondents who move around April 1 are able to give correct answers. Error in recalling a move or a move date may cause respondents to be enumerated at the wrong location in the census. This study investigates recall error in reports of moves and move dates in censuses and sample surveys using

data from survey files linked to administrative records.

Highlights: During FY 2017, staff continued to collaborate with colleagues in the Center for Survey Measurement (CSM) to complete revised analyses of recall error for reports of move dates in surveys using data from the National Longitudinal Survey of Youth linked to a third-party database. The data were prepared for the "Memory Recall of Migration Dates in the National Longitudinal Survey of Youth" developed under a contract with the National Opinion Research Center (NORC).

Staff: Mary Mulry (x31759)

H. 2020 Unduplication Research

Description: The goal of this project is to conduct research to guide the development and assessment of methods for conducting nationwide matching and unduplication in the 2020 Decennial Census, future Censuses and other matching projects. Our staff will also develop and test new methodologies for unduplication. The project is coordinated by one of the 2020 Census Integrated Project Teams.

Highlights: During FY 2017, staff obtained data files with Protected Identification Keys (person identifiers defined by the Center for Administrative Records and Applications (CARRA)) for use in the evaluation of matching systems. Staff continues to examine the files and extract data from them.

Staff: Michael Ikeda (x31756), Ned Porter, Bill Winkler, Emanuel Ben-David

I. 2020 Census Communications Campaign Statistical Analyses

Description: Both the 2000 and 2010 U.S. Censuses included a social marketing communications campaign that aided in maintaining the mail response rate in an environment when response to surveys was declining. As the 2020 U.S. Census approaches, the preparations include tests of new methodologies for enumeration that have the potential to reduce cost and improve quality. In parallel, the research includes formulating methods for the 2020 Census communications campaign that will aid the effectiveness of the enumeration operations. A team has been set up to conduct the research. For example, the 2015 Census Test in Savannah, GA included tests of Internet and mail response modes and of online delivery of social marketing communications focused on persuading the public to respond by Internet and mail. Analyses of the 2015 Census Test results and other data support the preparations for the 2020 Census communications campaign.

Highlights: During FY 2017, staff merged self-response results from three sources: (1) the 2015 Census Test in Savannah, GA, (2) the Low Response Score (LRS) found on the Census Bureau's Planning Database, and (3) Tapestry, a third-party population and geographic segmentation to create a dataset suitable for studying relationships between census response, the LRS and lifestyle segments. The use of the Tapestry is innovative in that it is designed for commercial marketing and not commonly used in sample survey research or census taking. The analyses found a great deal of variation in self-response and in the mean LRS between segments. The analyses provide insight regarding hard-to-survey populations, their response behavior, and interactions with social marketing communications. The analyses also provided some insight into variations among modes of self-response and the sources of internet responses across segments. Staff presented the results at the 2017 Joint Statistical Meetings.

Staff: Mary Mulry (x31759)

J. Undercount of Young Children

Description: The Census Bureau acknowledges the long-standing undercount of young children in decennial censuses and in Census Bureau sample surveys. Demographers have documented the high undercount of children under the age of 5 (e.g., West and Robinson 1999). Evaluations show that Census Bureau sample surveys like the American Community Survey (ACS), the Current Population Survey, and the Survey of Income and Program Participation also undercount young children, which can result in biased sample survey estimates (O'Hare and Jensen 2014). O'Hare (2015) found many other countries have a high net undercount of young children in their censuses. In 2014, the Census Bureau released a task force report summarizing this issue and recommending research to better understand the possible causes for this undercount (U.S. Census Bureau 2014). As a result, an interdivisional team is needed to work on several projects to review existing data sources that might provide insights into the high undercount of young children in the 2010 Census that would aid in lowering the undercount of young children in the 2020 Census.

Highlights: During FY 2017, staff used regression modeling to explore the relationship between the positive responses to child-specific probes on the 2010 Census questionnaire and 2010 Census Coverage Measurement weighted nonmatching children 0 to 4 years of age. The units of analysis are Tapestry segments which is a third-party segmentation of the population by geography and lifestyle segmentation. The modeling uses positive responses to child-specific probes on the census questionnaires to avoid possible confounding that might be caused by nonresponse in the Coverage Followup (CFU). After all, additions require a CFU interview. The CFU was a telephone interview and the operation was not able to contact all the households that gave positive responses to the coverage probes.

Because the CFU was able to add some young children, the next question that arises is whether most CCM nonmatches and CFU positive responses are in areas less likely to respond to the census or if some areas with average or even high response make errors when they complete their forms. This is the focus of the next phase of the research.

Staff: Mary Mulry (x31759)

K. Variables with Potential to Improve Small Area Estimation of Census Coverage Error

Description: This project investigates the potential for several variables to improve the small area estimation of components of census coverage error, namely omissions and erroneous enumerations. The approach investigates the relationship between variables available for all census responses that include 2010 Census Coverage Measurement (CCM) nonmatching young children. The approach uses ESRI's 67 Tapestry Segmentation, which is a geographic segmentation. Housing units are assigned to segments based on the tract of their address. Variables that appear able to predict the number or percentage of CCM nonmatching young children in a segment will be referred to other researchers focusing on small area estimation for further consideration.

Highlights: During FY 2017, staff presented preliminary results for a regression model of the 2010 CCM weighted nonmatching children 0 to 4 years of age on the number of positive responses to child-specific coverage probes on the 2010 Census questionnaire to the Coverage Measurement Design and Estimation Team. The presentation was given in response to a team member's request for suggestions for variables to use in models for small area estimation of components of census coverage error. Staff had suggested using positive responses to the coverage probes for all household members, not just those for young children, would be helpful in estimation of coverage error for small areas. These probes appear on all census forms, as opposed to being available on a sample basis. After seeing staff's preliminary results, the team decided to pursue research to determine whether positive responses to all the coverage probes, would be helpful variables in small area estimation models. Based on odds ratios, these variables proved to be influential in logistic regression models of the probability of census enumerations being correct, a duplicate, and other types of erroneous enumerations. A draft Decennial Statistical Studies Division (DSSD) memorandum with the results is undergoing review. These results also may be helpful in the census fieldwork.

Staff: Mary Mulry (x31759)

L. 2020 Census Privacy Research

Description: The Census Bureau is researching methods to make the published results of the 2020 Census differentially private. In support of that research, staff members are developing statistical models and visualization techniques to help facilitate the differentially private methods.

Highlights: During FY 2017, staff researched and developed differentially private methods to take a differentially private histogram and iteratively impute differentially private data at lower levels of geography given certain sets of constraints and additional noisy queries. They developed two main methods to do this geography imputation. The first was based on the idea of iterative proportional fitting and the second was based on a non-negative least squares optimization. Neither method produces integer counts, so staff further developed an integer rounding methodology that yields an integer solution that is close to the non-integer solution that still respects all constraints. Staff wrote code in Python to implement the methods and tested them on 2010 Census data to assess feasibility and accuracy. Additionally, staff developed interactive data visualizations using R Shiny that allowed team members to visualize high dimensional data and associated metrics as well as assess its accuracy compared with a second dataset.

Staff: Robert Ashmead (x31564), Brett Moran, Michael Ikeda, Philip Leclerc, John Abowd (ADRM), Daniel Kifer (ADRM)

M. Project to Study Priority of Tracts for Outreach and Advertising in Decennial Census 2020

Description: This project concerns research on how to augment for Census 2020 outreach purposes the Low Response Score (LRS) earlier developed for Census Tracts by C. Erdman and N. Bates to quantify nonresponse in terms of tract-level predictor variables available on the Planning Data Base (PDB). The idea is to re-weight the tracts according to which of their predictor variables producing extremely high LRS scores indicate membership in subpopulations for which outreach partnership efforts are likely to be successful.

Highlights: During FY 2017, regular meetings of a team were held and techniques for reweighting tracts using PDB data were explored using various codes in R. CSRM staff implemented several algorithms for reweighting hard-to-count tracts, upweighting for larger population and (separately) for high levels of LRS-score variables for which available interventions by "Partnership Specialists" seem likely to improve response. Lists of highest-priority tracts, both at national level and within target counties for the 2018 Census test, were circulated to field staff to obtain feedback on usefulness of the reweighting algorithm.

Staff: Eric Slud (x34991), Ryan Janicki, Nancy Bates (ADRM)

1.5 ADDRESS CANVASSING IN FIELD (Decennial Project 6350C02)

A. Statistical Evaluation of the In-office Image Review Process
Description: A large-scale address canvassing operation was carried out before the 2010 Census to ensure that the Master Address File was up to date on census day. For 2020, the Census Bureau is working to replace some of the fieldwork of the previous operation with steps that can be taken in-office, such as review of aerial imagery. This project will investigate whether the imagery can predict when in-office and in-field canvassing agrees. If successful, such a model could help to inform future field operations that are smaller in scale.

Highlights: During FY 2017, staff met with representatives from the Geography Division (GEO) and the Decennial Statistical Studies Division (DSSD) to establish scope for the problem. The GEO team constructed an initial dataset and staff began exploratory and predictive analysis using traditional (non-image) covariates.

Staff: Andrew Raim (x37894), Dan Weinberg, Scott Holan (ADRM)

B. Development of Block Tracking Database
Description: The Targeted Address Canvassing (TRMAC) project supports Reengineered Address Canvassing for the 2020 Census. The primary goal of the TRMAC project is to identify geographic areas to be managed in the office (i.e., in-office canvassing) and geographic areas to be canvassed in the field. The focus of the effort is on decreasing in-field and assuring the Master Address File (MAF) is current, complete, and accurate. The Block Assessment, Research, and Classification Application (BARCA) is an interactive review tool which will allow analysts to assess tabulation blocks—and later Basic Collection Units (BCUs)—by comparing housing units in 2010 imagery and current imagery, along with TIGER reference layers and MAF data.

Highlights: During FY 2017, staff continued to enhance the quality control feature and the system of automated reports for BARCA. Clerks completed the milestone of the first pass of In-Office Address Canvassing (IOAC) for all 11 million Census blocks. The next phase of IOAC has begun as new imaging and address lists become available. Clerks have also begun the 2nd phase of the In-Office Address Canvassing. As new address lists and new imaging become available, blocks will be triggered for another round of processing so the Master Address File can stay current.

Staff: Tom Petkunas (x33216)

1.6 AMERICAN COMMUNITY SURVEY (ACS) (Decennial Project 6385C70)

A. ACS Applications for Time Series Methods
Description: This project undertakes research and studies on applying time series methodology in support of the American Community Survey (ACS).

Highlights: During FY 2017, staff (a) completed work on methodology that provides kriging estimates in flow sample survey data, accounting for the contribution of sampling error to overall uncertainty. Code was updated and final results for all counties were generated; and (b) staff obtained monthly data files from the American Community Survey Office (ACSO), made extensions of the kriging framework to monthly data, involving continuous-time ARMA models with level shift regressors to handle changes that were made to the questionnaire. Issues of parametrization, numerical optimization, and modeling have been explored.

Staff: Tucker McElroy (x33227), Patrick Joyce

B. Data Analysis of ACS CATI-CAPI Contact History
Description: This project concluded a series of data analytic segments analyzing and reporting on contact history data from American Community Survey (ACS) Computer Assisted Telephone Interview (CATI) and Computer Assisted Personal Interview (CAPI) contact history data.

Highlights: While this collaborative project was largely completed during FY 2016, activity during FY 2017 consisted of writing, submitting, revising and publishing a paper in *Journal of Official Statistics* summarizing and abstracting the methodology of the previously completed series of American Community Survey (ACS) Research and Evaluation reports concerning the development of data-driven field intervention policies (for curtailment of field case followup) to minimize survey contact burden while maintaining maximal rate of interview completion. The title of the paper is "Adaptive Intervention Methodology for Reduction of Respondent Contact Burden in the American Community Survey".

Staff: Eric Slud (x34991), Robert Ashmead

C. Assessing Uncertainty in ACS Ranking Tables

Description: This project presents results from applying statistical methods which provide statements of how good the rankings are in the ACS Ranking Tables (see The Ranking Project: Methodology Development and Evaluation Research Section under Projects 0331000 and 0925000).

Highlights: [See General Research: Survey Sampling-Estimation and Modeling (C), The Ranking Project: Methodology Development and Evaluation]

Staff: Tommy Wright (x31702), Martin Klein, Jerzy Wieczarek (Carnegie Mellon University), Brett Moran, Nathan Yau, Michael Leibert

D. Confidence Intervals for Proportions in ACS Data

[See General Research: Small Area Estimation (B), Coverage Properties of Confidence Intervals for Proportions in Complex Surveys]

E. Voting Rights Section 203 Model Evaluation and Enhancements Towards Mid-Decadal Determinations

Description: Section 203 of the *Voting Rights Act* mandates determinations by the Census Bureau relating to rates of citizenship, limited English proficiency and limited education among specified small subpopulations (voting-age persons in various race and ethnicity groups called Language Minority Groups [LMGs] for small areas such as counties or minor civil divisions MCDs). The Section 203 determinations for 2016 will result in the legally enforceable requirement that certain geographic political subdivisions must provide voting materials in languages other than English in future elections. The research undertaken in this project consists of the development, assessment and estimation of regression-based small area models based on 5-year (2010-2014) American Community Survey (ACS) data. This modeling and estimation effort differs from the effort supporting Section 203 determinations in 2011. The 2016 models and estimates cannot make use of data from a nearly contemporary decennial census.

Highlights: During FY 2017, in consultation with the Decennial Statistical Studies Division (DSSD), the American Community Survey Office (ASCO), the Census Redistricting and Voting Rights Data Office (CRVRDO), and the Center for Statistical Research and Methodology (CSRM), staff continued investigations into the application of small area models relating to Section 203 of the Voting Rights Act. CSRM staff finished developing a parametric bootstrap methodology to estimate the standard errors of the modelled point estimates and derived the point estimates and standard errors from the 2010-2014 ACS 5-year data product. The point estimates were delivered to DSSD, and in early December, *The Federal Register Notice*, determinations, an executive summary of methods, and public use data files were published online.

Staff continued to work on technical documentation to support, document, and scrutinize the small area estimation methods and conducted additional research concerning model diagnostic methods for small area estimation and variance estimation methods using a hybrid BRR and parametric bootstrap method. Staff prepared research results on different aspects of the model selection, validation, and variance estimation and gave presentations on the results at the 2017 ISI Satellite Meeting on Small Area Estimation and the 2017 Joint Statistical Meetings. In addition to giving the presentation, staff discussed the small area models relating to Section 203 of the *Voting Rights Act* with other small area estimation experts.

Staff: Robert Ashmead (x31564), Eric Slud, Patrick Joyce, Mark Asiala (DSSD)

F. ACS Income Modeling

Description: A team has been assembled to investigate and model the extent to which American Community Survey (ACS) Income questions might be replaceable by modeled quantities using other ACS and Census Bureau survey data and administrative records.

Highlights: During FY 2017, staff attended several team meetings and assembled annotated bibliographies of literature based on previous Census Bureau reports on income estimation and associated administrative records. Meetings and formulation of research tasks are continuing.

Staff: Eric Slud (x34991), Tommy Wright, John Eltinge

1.7 DEMOGRAPHIC STATISTICAL METHODS DIVISION SPECIAL PROJECTS (Demographic Project TBA)

A. Research Balanced Repeated Replication and Other Variance Estimation Techniques for Use with Current Population Survey

Description: The current practice of variance estimation on the Current Population Survey (CPS) relies on Successive Difference Replication, which is a form of Balanced Repeated Replication (BRR) developed at the Census Bureau. Properties of this method, and comparison between it and alternatives, is the topic of this Demographic Statistical Methods (DSMD) research, on which CSRM staff consults. The scope of this project has now expanded to include model-based longitudinal analysis, design-based weighting and variance estimation concerning longitudinal gross flows in employment categories within the CPS.

Highlights: During FY 2017, staff met regularly to discuss definitions, properties, diagnostic tests, and possible variants of Balanced Repeated Replication (BRR) variance estimation methods in the expanded scope of longitudinal research on CPS. CSRM staff are investigating the use of loglinear models for gross flow analysis, with design-based analyses used to estimate large cell–counts in the gross flow cross-tabulations, and model-based analyses for the small cell-counts. Extended log-linear and generalized linear models with random effects are also under consideration for small area estimation of gross flows at state or lower levels of aggregation.

Staff: Eric Slud (x34991), Yves Thibaudeau, Yang Cheng (DSMD), Khoa Dong (DSMD), Tim Trudell (DSMD)

1.8 DEMOGRAPHIC SURVEYS DIVISION (DSD) SPECIAL PROJECTS (Demographic Project 0906/1444X00)

A. Data Integration

Description: The purpose of this research is to identify microdata records at risk of disclosure due to publicly available databases. Microdata from all Census Bureau sample surveys and censuses will be examined. Potentially linkable data files will be identified. Disclosure avoidance procedures will be developed and applied to protect any records at risk of disclosure.

Highlights: Staff integrated data files for the Privacy Preserving projects and Record Linkage. To accomplish this, staff wrote software and wrote documentation for these software and procedures. Staff extracted files for 11 different tables and documented the procedures on the 2020 DAS Wiki. Staff implemented version control on all software developed by the staff.

Staff: Ned Porter (x31798)

1.9 POPULATION DIVISION PROJECTS (Demographic Project TBA)

A. Introductory Sampling Workshop

Description: In support of Population Division's International Programs Area, staff will conduct (on request) introductory sampling workshops with focus on probability sampling for participants from various countries. These workshops are primarily funded by USAID or other sources.

Census Bureau Headquarters Workshop (Fall 2016)

Highlights: Over the two-week period (October 24-November 4, 2016), staff conducted a Workshop: Introduction to Survey Sampling (focus on Probability Sampling) at the Census Bureau Headquarters. As in the past, the workshop presented the main components of survey sampling with a focus on probability sampling and estimation techniques. The hands-on, interactive workshop included the production of estimates of population parameters from sample surveys as a function of sample design, weighting procedures, the computation of sampling errors of sample estimators, and the making of inferences from the sample to the population. The six workshop participants were staff from national statistical agencies: the National Institute of Health (Mozambique); Central Agency for Public Mobilization and Statistics (Egypt); and the Pakistan Bureau of Statistics. On the final day, the workshop featured a Panel on Sampling to give overviews of the American Community Survey, the Monthly/Annual Retail Trade Surveys, and the Current Population Survey. Plans are in place to offer the workshop in the Fall 2017 for international participants.

Staff: Tommy Wright (x31702), Michael Leibert

Rwanda Sampling Workshop (DEMO Project 8940100)

Staff traveled to Kigali, Rwanda to conduct a week-long (August 14-18, 2017) Workshop on Sampling at the National Institute of Statistics of Rwanda (NISR). NISR was formed in 2005. The sixteen workshop participants were mostly staff statisticians at NISR (includes 1 Intern) who conduct censuses and sample surveys, and who provide associated statistical methodology development. The workshop was sponsored by NISR and the U.S. Bureau of the Census.

This workshop presented the main components of survey sampling with a focus on probability sampling techniques such as simple random sampling, systematic sampling, stratification, cluster sampling, multistage sampling, and systematic sampling. When sample surveys are undertaken, NISR uses face-to-face and the response rates are near 99%. Many questions from participants on their current work were discussed. The workshop focused on understanding concepts and some theory relating to probability sampling.

Unfortunately, the five-day workshop included two holidays (the Assumption of Mary and the Inauguration of the President of Rwanda). However, most participants attended these two days. Also on both holidays, we ended a little early. T. Wright presented a special talk during the workshop closing session—"No Calculation When Observation Can Be Made" which highlighted: censuses, probability sampling, and big data.

Staff: Tommy Wright (x31702)

1.10 SOCIAL, ECONOMIC, AND HOUSING STATISTICS DIVISION SMALL AREA ESTIMATION PROJECTS (Demographic Project 7165017)

A. Research for Small Area Income and Poverty Estimates (SAIPE)
Description: The purpose of this research is to develop, in collaboration with the Small Area Estimates Branch in the Social, Economic, and Housing Statistics Division (SEHSD), methods to produce "reliable" income and poverty estimates for small geographic areas and/or small demographic domains (e.g., poor children age 5-17 for counties). The methods should also produce realistic measures of the accuracy of the estimates (standard errors). The investigation will include assessment of the value of various auxiliary data (from administrative records or surveys) in producing the desired estimates. Also included would be an evaluation of the techniques developed, along with documentation of the methodology.

Highlights: During FY 2017, staff completed two revisions of a paper related to estimating year-to-year changes for county poverty rates for school-aged children in poverty through measurement error models, and extended the results to estimating yearly county poverty rates. The estimates here are obtained using a bivariate model that jointly models two consecutive years of American Community Survey (ACS) yearly estimates and uses the ACS five-year estimates as a covariate with measurement error. The paper will appear in the *Journal of the Royal Statistical Society*. Staff delivered a presentation on this on a topic-contributed talk in the 2017 Joint Statistical Meetings.

Staff completed the debugging of a program that implements multivariate functional measurement error models through an efficient Markov Chain Monte Carlo (MCMC) algorithm. The program was used to obtain estimates for rates of children in poverty as well as yearly changes in these rates. Using this program, staff implemented bivariate and univariate measurement error models, as well as a naïve model that ignores errors in the covariates. Results were similar to those obtained with the generic software JAGS, which confirmed the accuracy of the program. This algorithm used a prior distribution for the parameters which we had previously proved to result in a proper posterior.

Staff studied the impact of ignoring measurement error (ME) when estimating year-to-year changes in poverty rates for school-aged children in poverty, such as might be done using SAIPE data. For year-to-year changes, it is appropriate to use a bivariate model. The bivariate model studied jointly models two consecutive years of ACS yearly estimates of county rates of 5-17 children in poverty and uses the ACS five-year estimates as a covariate with measurement error. Staff found that ignoring ME in this setting can yield incorrect and misleading estimates of prediction variances.

Staff studied the impact of assuming a univariate structural or functional measurement error model or a naïve model for estimating yearly SAIPE county poverty rates for school aged-children. SAIPE will likely eventually need to replace the Census long form estimates with five-year ACS estimates and this study sheds light on the most appropriate way to incorporate the sampling error in such estimates into the model. The results, which show the importance of properly accounting for measurement error for county poverty rates for school-aged children in poverty, were summarized and presented in an invited talk for a conference for J.N.K. Rao's 80th birthday. Staff also presented the results in the 2017 Small Area Conference in Paris, France.

Staff explored the impact of ignoring the correlation between the model errors when estimating year-to-year changes of county poverty rates for children through Fay-Herriot Models with measurement error. Staff found that for this application, ignoring such correlation by simply subtracting the estimates produced by univariate models and adding the corresponding variances can result both in incorrect estimates and grossly misstated standard errors.

Staff: Carolina Franco (x39959), Jerry Maples, William Bell (ADRM)

B. Small Area Health Insurance Estimates (SAHIE)
Description: At the request of staff from the Social, Economic, and Housing Statistics Division (SEHSD), our staff will review current methodology for making small area estimates for health insurance coverage by state and poverty level. Staff will work on selected topics of SAHIE estimation methodology, in conjunction with SEHSD.

<u>Small Area Estimation of Proportions</u>
Highlights: During FY 2017, staff continued to work on small area estimation of proportions using beta regression models and extended previous work to allow direct estimates to take values of 0 by using a zero-inflated beta density as a sampling model. The probability of observing a zero was modeled as a function of the true small area mean. A Markov Chain Monte Carlo (MCMC) algorithm was developed and used to fit the zero-inflated model to a data set consisting of administrative records, and direct estimates of the proportion of men without health insurance in the income to poverty ratio group 138 to 400 in counties. Simulation studies were performed to understand properties of the model-based predictions.

Staff: Ryan Janicki (x35725)

C. Sub-County Estimates of Poverty from Multi-year ACS Data
Description: This project is from the Development Case Proposal to improve the estimates of poverty related outcomes from the American Community Survey (ACS) at the tract level. Various modeling techniques, including model-based and model-assisted, will be used to improve on the design-based multi-year estimates currently produced by the ACS. The goal is to produce more accurate estimates of poverty and income at the tract level and develop a model framework that can be extended to outcomes beyond poverty and income.

Highlights: During FY 2017, staff developed and implemented models to estimate the number of poor school-aged children in poverty in census tracts using overdispersed Poisson distributions for the design-based estimator. Several variants of the model were compared to test how different adjustments made to the model to account for part of the survey design affected the mean squared error of the predictions. This research was presented at the Joint Statistical Meetings in Baltimore, Maryland in August 2017. Additional work on this project will continue into FY 2018.

Staff has also began additional work on the artificial population simulation framework to evaluate small area models. The new version will allow evaluation of both unit and area level models for all ACS collected characteristics (not just school-age child poverty) for both people and housing units. Development of this new version will continue into FY 2018.

Staff: Jerry Maples (x32873), Ryan Janicki, Carolina Franco, William Bell (ADRM)

1.11 ECONOMIC STATISTICAL COLLECTION (Economic Project 1183X01)

A. Research on Imputation Methodology for the Monthly Wholesale Trade Survey
Description: In the previous phase of this project, staff conducted a simulation study to investigate new imputation methodology for the Monthly Wholesale Trade Survey (MWTS). In this phase of the project, staff are creating a more realistic simulated wholesale trade population and investigating improved MWTS estimators. The MWTS is a longitudinal sample survey that provides month-to-month information on sales and inventories of U.S. merchant wholesalers. Key estimates produced from this sample survey include total sales, month-to-month relative change in sales, total inventories, and month-to-month relative change in inventories (overall and within industry subclasses). There are a number of challenges when developing estimators for the MWTS, including variables with highly skewed distributions, missing values in predictor variables from the Economic Census, and sample survey variables with trends that differ across industry classes. The longitudinal information in addition to a rich set of frame data available from the Economic Census can be used to build Bayesian models that address these challenges. It is expected that this model will be applicable to other business sample surveys.

Highlights: During FY 2017, staff developed and enhanced a version of a realistic, artificial population designed for drawing simulated Monthly Wholesale Trade Survey (MWTS) data to be used for conducting realistic simulation studies. In previous work, staff developed a version of this population representative of the two-year period from December 2008 to December 2010. This population provides a tool for evaluating the performance of statistical methodology applied to the MWTS or other similar surveys. Staff also expanded the artificial population to enable one to draw simulated MWTS data representative of the longer time period of December 2008 to March 2013. Staff expanded the population because simulated samples covering a multi-year period were needed to evaluate the performance of models that include a seasonal component. Staff assembled the available data and used sequential regression imputation with random forests as the conditional models to fill in data for missing values and population units that do not appear in the sample data. Staff evaluated the constructed population by comparing its features with corresponding features of the available sample data. Based on these evaluations, staff made further improvements to enhance the realism of the constructed population. Staff developed a procedure that fit a weighted nonparametric density estimate to the actual MWTS sample data, and transformed population values using the inverse cumulative distribution function of the fitted nonparametric density computed on the ranks of the population values. This procedure further improved the realism of marginal distributions of population variables. Staff prepared diagnostic plots and other summaries to compare distributions in the artificial population with those of the actual sample data. Staff prepared a deliverable version of this population along with accompanying documentation.

Staff: Martin Klein (x37856), Brett Moran, Joe Schafer (ADRM), Joanna Fane Lineback (CSM)

B. Use of Big Data for Retail Sales
Description: In this project, we are investigating the use of “Big Data” to fill gaps in retail sales estimates currently produced by the Census Bureau. Specifically, we are interested in how to use “Big Data” to

supplement existing monthly/annual retail surveys with a primary focus on exploring (1) how to use third party data to produce geographic level estimates more frequently than once every five years (i.e. a new product), and (2) the possibility of using third party data tabulations to improve/enhance Census Bureau estimates of monthly retail sales - for example, validation and calibration. Various types of data are being pursued such as credit card transaction data and scanner data.

Highlights: During FY 2017, staff participated in Big Data Knowledge Sharing meetings, attended meetings with Palantir staff regarding the continuation of the study of First Data credit card data pursuant to a recently awarded contract, and discussed further analysis of store-level data from the NPD Group. CSRM and Economic Directorate staff met with a *SUMMER AT CENSUS* scholar to exchange ideas and progress related to data science methods.

Staff: Darcy Steeg Morris (x33989), Osbert Pang, Tommy Wright, Rebecca Hutchinson (EID), Scott Scheleur (EID)

1.12 ECONOMIC CENSUS/SURVEY ENGINEERING: TIME SERIES RESEARCH; ECONOMIC MISSING DATA/PRODUCT LINE DATA; DEVELOPMENT/SAS (Economic Project 2270C10)

A. Seasonal Adjustment Support
Description: This is an amalgamation of projects whose composition varies from year to year but always includes maintenance of the seasonal adjustment software used by the Economic Directorate.

Highlights: During FY 2017, staff provided seasonal adjustment and software support for users within and outside the Census Bureau, including American Trucking Associations, CEIC Data (Bulgaria), First Guaranty Mortgage Corporation, Informed Portfolio Management, Mitsubishi UFJ Research and Consulting Co. (Japan), J,P. Morgan, Nikkei (Japan), SAS, Reserve Bank of Australia, BEA, Bureau of Labor Statistics (BLS), Federal Highway Administration, Washington Economic and Revenue Forecast Council, Department of Transportation, Australian Bureau of Statistics, Bangladesh Bureau of Statistics, Eurostat, Education Development Department (Canada), INEGI (Mexico), INSEE (France), Office of National Statistics (UK), Institute of Statistics Quebec, Statistics Canada, OECD, Statistics New Zealand, National Institute of Statistics and Informatics (INEI-Peru), National Institute of Statistics and Censuses (Argentina), National University of the Northeast (Argentina), University of Basel (Switzerland), University of Wisconsin School of Business. Staff participated in a time series workshop in London, UK on March 6 and 7, 2016, sponsored by the Office of National Statistics, presenting the current state of time series research and development at the Census Bureau, discussing issues related to preadjustments for seasonal adjustment, and participating in brainstorming sessions on future time series research. Staff, with staff from the Economic Statistical Methods Division (ESMD), organized a successful seasonal adjustment workshop held at the BLS Conference Center on November 4, 2016, and has begun to plan for a similar workshop in April of 2018. Several staff members participated in an inter-agency group searching for applied statistical techniques to solve problems related to residual seasonality in GNP, and developed comments on a summary of the seasonal adjustment system used by the Bureau of Economic Analysis when they generate seasonal adjustments for GDP and its components. Staff met with Center for Economic Studies (CES) staff to discuss the final seasonal adjustments of various quarterly series (both state and national level) and delivered files with final seasonal adjustments.

Staff: Brian Monsell (x31721), Tucker McElroy, James Livsey, Osbert Pang, Anindya Roy, William R. Bell (ADRM)

B. Seasonal Adjustment Software Development and Evaluation
Description: The goal of this project is a multi-platform computer program for seasonal adjustment, trend estimation, and calendar effect estimation that goes beyond the adjustment capabilities of the Census X-11 and Statistics Canada X-11-ARIMA programs, and provides more effective diagnostics. The goals for FY 2015 include: continuing to develop a version of the X-13ARIMA-SEATS program with accessible output and updated source code so that, when appropriate, the Economic Directorate can produce SEATS adjustments; and incorporating further improvements to the X-13ARIMA-SEATS user interface, output and documentation. In coordination and collaboration with the Time Series and Related Methods Staff of the Economic Statistical Methods Division (ESMD), staff will provide internal and/or external training in the use of X-13ARIMA-SEATS and the associated programs, such as X-13-Graph, when appropriate.

Highlights: During FY 2017, staff released an updated version of X-13ARIMA-SEATS (Version 1.1, Build 39), to the Economic Directorate for testing. They compared adjustments from this version of the software to the last released version of X-13ARIMA-SEATS (Version 1.1, Build 26) and found very small differences in the output due to a change in the precision of the value of pi used in the software. Some of the new features of this release are a new easter[0] regressor to available regressors; a new testalleaster argument to test all Easter regressors in the AICC testing procedure; a new test of seasonality for the quarterly version of the original series and seasonally adjusted series of monthly series; and standardized the use of pi in the software. After the Economic Directorate completed its testing, staff released this version of X-13ARIMA-SEATS (Version 1.1, Build 39), to the public. After the release of this version, staff fixed defects in the X-13ARIMA-SEATS software, including a module that saves input specifications related to regARIMA modeling, and fixed a problem with the program writing extremely small numbers. Staff organized and delivered a lecture series on the X-13ARIMA-SEATS source code. Staff also updated a summary of the variables used in the input routines of the X-13ARIMA-SEATS code to distribute to participants.

Staff continued the development of sigex, a suite of R routines for modeling multivariate time series. Staff revised the software to include canonical models, two new types of cyclical models, Hodrick-Prescott filtering, updated mean squared error calculations, and modifications to method-of-moments estimators to ensure invertibility. In addition, staff revised the software to include multi-step ahead forecasting and error covariance calculations, and the capacity to forecast extracted signals. The software was applied to several daily time series.

Staff: Brian Monsell (x31721), Tucker McElory, Osbert Pang

C. Research on Seasonal Time Series - Modeling and Adjustment Issues

Description: The main goal of this research is to discover new ways in which time series models can be used to improve seasonal and calendar effect adjustments. An important secondary goal is the development or improvement of modeling and adjustment diagnostics. This fiscal year's projects include: (1) continuing research on goodness of fit diagnostics (including signal extraction diagnostics and Ljung-Box statistics) to better assess time series models used in seasonal adjustment; (2) studying the effects of model based seasonal adjustment filters; (3) studying multiple testing problems arising from applying several statistics at once; (4) determining if information from the direct seasonally adjusted series of a composite seasonal adjustment can be used to modify the components of an indirect seasonal adjustment, and more generally investigating the topics of benchmarking and reconciliation for multiple time series; (5) studying alternative models of seasonality, such as Bayesian and/or long memory models and/or heteroskedastic models, to determine if improvement to seasonal adjustment methodology can be obtained; (6) studying the modeling of stock holiday and trading day on Census Bureau time series; (7) studying methods of seasonal adjustment when the data are no longer univariate or discrete (e.g., multiple frequencies or multiple series); (8) studying alternative seasonal adjustment methods that may reduce revisions or have alternative properties; and (9) studying nonparametric methods for estimating regression effects, and their behavior under long range dependence and/or extreme values.

Highlights: During FY 2017, staff (a) completed work on fitting vector moving averages by establishing stability properties and proving asymptotic properties; (b) developed modified method-of-moments estimation for long multivariate high-frequency time series, such that invertibility is guaranteed; (c) constructed weather regressors, and incorporated these into seasonal adjustment methodology, with application to regional construction series; (d) continued research on modeling of daily time series (New Zealand immigration data and credit card transaction data) with multiple periods of seasonality, with new work on canonical models and utilizing the Hodrick-Prescott filter to separate trend from annual seasonality; (e) compared mean squared errors of seasonal adjustments and determined the calibration of a seasonality detection diagnostic; (f) continued research into using weekly seasonal factors to better estimate monthly series from reported weekly data; and (g) continued investigation into seasonal vector form and the application to generating new seasonality diagnostics.

Staff: Tucker McElroy (x33227), James Livsey, Brian Monsell, Osbert Pang, Thomas Trimbur, William Bell (ADRM), David Findley (Private Collaborator)

D. Supporting Documentation and Software for X-13ARIMA-SEATS

Description: The purpose of this project is to develop supplementary documentation and utilities for X-13ARIMA-SEATS that enable both inexperienced seasonal adjustors and experts to use the program as effectively as their backgrounds permit. This fiscal year's goals include improving the X-13ARIMA-SEATS documentation, exploring the use of R packages that interface with X-13ARIMA-SEATS, and exploring the use of component and Java software developed at the National Bank of Belgium.

Highlights: During FY 2017, staff updated the *X-13ARIMA-SEATS REFERENCE MANUAL* to include information on new options. Staff updated HTML files to release new versions of X-13ARIMA-SEATS and Win X-13, and updated HTML documentation files for Win X-13. Staff also created new pages for the X-13-SAM program developed by Demetra Lytras from the Economic Statistical Methods Division (ESMD), and revised all the pages of the X-13ARIMA-SEATS website to include a tab for these pages on the X-13ARIMA-SEATS website. Several staff members are contributing to a paper that documents the various diagnostics used to evaluate seasonal adjustments as part of work with the Bureau of Economic Analysis (BEA) on the GDP residual seasonality problem. Staff also developed a draft paper with details of the diagnostics stored in the universal diagnostics file (.udg). Staff revised R routines used to generate diagnostic summaries for seasonal adjustment used in the study for the Center for Economic Studies (CES). Staff participated in meetings concerning the modernization of CSRM's web presence. Staff produced a listing of papers in the *CSRM Research Report Series* to aid in this effort. Staff worked with Christoph Sax to improve utilities related to the seasonal R package, and collaborated with Christoph Sax in creating a package named "x13story" to improve the Census Bureau's communication of seasonal adjustment and X-13ARIMA-SEATS.

Staff: Brian Monsell (x31721), Tucker McElroy, James Livsey, Osbert Pang, Anindya Roy, Thomas Trimbur, William R. Bell (ADRM)

1.13 INVESTIGATION OF ALTERNATIVE METHODS FOR RESOLVING BALANCE COMPLEX FAILURES IN StEPS (Economic Project TBA)

A. Investigation of Alternative Methods for Resolving Balance Complex Failures in StEPS

Description: The Standard Economic Processing System (StEPS) implements a raking algorithm for adjusting balance complexes in order to satisfy the requirement that the sum of items (details) in a balance complex balances to reported totals. In this project, we research alternative methods to raking when the data items are negative or when there is subtraction in the balance complex.

Highlights: During FY 2017, staff collaborated with the Economic Directorate on developing alternative methods to raking for the Quarterly Financial Report (QFR) as it migrates to StEPS II processing. The QFR data must satisfy several nested balance complexes in which the sum of detail items must balance to the reported totals. If a balance complex fails, then either the total or the set of details must be adjusted. StEPS implements a raking algorithm for adjusting balance complexes when the total and the summed details are within a specified tolerance. However, raking applies to positive data only and can lead to erroneous estimates when items can have negative values, as is the case for several QFR items. Staff developed four separate nonlinear programs that minimize loss functions under the specified additivity constraints. Staff designed a weighting scheme for the QFR items and incorporated the weights as "cost" in the objective functions, i.e. the weights represent the cost of adjusting the corresponding item when solving the nonlinear program. Staff applied these methods to failing balance complexes in the QFR, investigated additional survey-specific constraints, and evaluated the statistical properties of the adjusted items.

Staff: Maria Garcia (x31703), Yves Thibaudeau, Laura Bechtel (ESMD)

1.14 NATIONAL CANCER INSTITUTE (Census Bureau Project 7225010)

A. National Cancer Institute Tobacco Use Survey/Current Population Survey

Description: During the first and second quarters of FY 2017, staff started a new project using Current Population Survey (CPS) files from the Demographic Statistical Methods Division (DSMD) on a project for the National Cancer Institute (NCI), studying the relationship between smoking status and a range of geographic/demographic covariates. The Tobacco Use Supplement to the Current Population Survey (TUS-CPS) is a National Cancer Institute (NCI) sponsored survey of tobacco use that has been administered as part of the U.S. Census Bureau's Current Population Survey every two to four years since 1992. The TUS/CPS is designed to produce reliable estimates at the national and state levels. However, policy makers, cancer control planners, and researchers often need county level data for tobacco related measures to better evaluate tobacco control programs, monitor progress in the control of tobacco use, and conduct tobacco-related research. Staff is currently exploring the possibility of using model-based small area estimation (SAE) techniques due to insufficient samples in some small counties across the country. Using SAE techniques, staff is currently working to produce county level model based estimates for the following outcomes for 2014/2015 data cycle: current smoking prevalence among age 18+; ever smoking prevalence among age 18+; smoke-free work place policy prevalence among age 18+; smoke-free home rule prevalence among age 18+; percentage of at least attempted to quit for 24+ hours among those former smokers "at risk" of quitting during the past 12 months and current or everyday smokers (age 18+). The goal of this project is to produce model-based county level estimates for the above five outcomes using the 2014-2015 TUS/CPS data by applying the similar methodology developed in the previous two data cycles (2006/2007 and 2010/2011). Staff also plans to evaluate additional outcomes, e.g., physician/dental advice to quit attitudes to smoke-free cars and casinos, outdoor venues and see the possibility of producing small area estimates for the new outcomes.

Highlights: During FY 2017, staff continued to work on Tobacco Use Supplement to the Current Population Survey (TUS-CPS) project. As part of the project, staff has produced and prepared reliable direct smoking estimates for the five smoking outcomes of interest at the national, state and county levels for NCI. For this project, staff used model-based small area estimation (SAE) techniques due to insufficient samples in some of the small counties across the country to produce direct estimates at the county level for each of the five smoking outcomes of interest for 2014/2015 data cycle. Staff estimated the design effects for each of the five smoking outcomes of interests at the national, state and county levels using Gabler, Haeder and Lahiri (1999) small area estimation model-based method. Additionally, staff ran backward model selection procedure to choose initial covariates with the arcsine transformation-data method. Direct county smoking estimates and design effects for the five smoking outcomes of interests were validated by using the state and national estimates.

Staff: Isaac Dompreh (x36801), Benmei Liu (NCI)

1.15 PROGRAM DIVISION OVERHEAD (Census Bureau Project 0331000)

A. Center Leadership and Support

This staff provides ongoing leadership and support for the overall collaborative consulting, research, and administrative operation of the center.

Staff: Tommy Wright (x31702), Lauren Emanuel, Michael Hawkins, Michael Leibert, Erica Magruder, Eric Slud, Kelly Taylor, Bill Winkler

B. Research Computing

Description: This ongoing project is devoted to ensuring that Census Bureau researchers have the computers and software tools they need to develop new statistical methods and analyze Census Bureau data.

Highlights: During FY 2017, the Integrated Research Environment (IRE) team continued to develop the IRE, a shared Linux computing platform that will replace the current "compute clusters:" research1, research2, and the RDC cluster. The IRE will provide the logical separation of project data and activities that is provided in the RDC environment, but without using a separate login for each project. Staff developed a strategy for a shared applications "volume" so that certain applications could be shared across the nodes of the cluster and certain applications could be installed locally while still being presented as part of the same /apps tree. The scheme allows for software to be mounted (or not) at certain mount points based on a configuration file. Staff added additional compute nodes, bringing the total number of compute nodes to 20 and the total number of "cores" to 1080 (hyperthreading on). Staff experimented with a method to associate certain queues with sets of compute nodes, and explored ways of granting access to those queues to particular sets of users. Staff conducted pretesting and fixed various defects reported by the pretesters. A larger "real world" test is planned for the first quarter of FY 2018. Outstanding issues include: devising a scheme to load balance multiple login nodes, implementing a secure printing scheme that meets the requirements for handling Title 26 materials, and implementing effective procedures to allow the authorized transfer of files to and from the cluster while preventing unauthorized transfer.

Staff: Chad Russell (x33215)

2. RESEARCH

2.1 GENERAL RESEARCH AND SUPPORT (Census Bureau Project 0331000)

Missing Data, Edit, and Imputation

Motivation: Missing data problems are endemic to the conduct of statistical experiments and data collection projects. The instigators almost never observe all the outcomes they had set to record. When dealing with sample surveys or censuses that means individuals or entities in the survey omit to respond, or give only part of the information they are being asked to provide. In addition, the information provided may be logically inconsistent, which is tantamount to missing. To compute official statistics, agencies need to compensate for missing data. Available techniques for compensation include cell adjustments, imputation and editing. All these techniques involve mathematical modeling along with subject matter experience.

Research Problems: Compensating for missing data typically involves explicit or implicit modeling. Explicit methods include Bayesian multiple imputation and propensity score matching. Implicit methods revolve around donor-based techniques such as hot-deck imputation and predictive mean matching. All these techniques are subject to edit rules to ensure the logical consistency of remedial product. Research on integrating together statistical validity and logical requirements into the process of imputing continues to be challenging. Another important problem is that of correctly quantifying the reliability of predictors that have been produced in part through imputation, as their variance can be substantially greater than that computed nominally.

Potential Applications: Research on missing data leads to improved overall data quality and predictors accuracy for any census or sample survey with a substantial frequency of missing data. It also leads to methods to adjust the variance to reflect the additional uncertainty created by the missing data. Given the ever rising cost of conducting censuses and sample surveys, imputation and other missing-data compensation methods may come to replace actual data collection, in the future, in situations where collection is prohibitively expensive.

A. Editing
Description: This project covers development of methods for statistical data editing. Good methods allow us to produce efficient and accurate estimates and higher quality microdata for analyses.

Highlights: During FY 2017, staff collaborated with staff in the Economic Directorate on a project with the purpose of developing alternative methods to raking when data items are allowed to be negative. The Standard Economic Processing System (StEPS) implements a raking algorithm for adjusting balance edits in order to satisfy the requirement that the sum of detail items balances to final totals. The existing algorithm fails when the data items are negative or when there is subtraction in the balance complex. Staff proposed alternative methods based on solving nonlinear programming problems that minimize changes in the adjusted data under specified balance editing constraints. Staff applied the proposed methods to adjusting failing balance editing complexes in the Quarterly Financial Report (QFR), including developing a weighting scheme (costs) to control the frequency at which each particular item would be changed according to QFR analysts' specifications.

Staff: Maria Garcia (x31703), Yves Thibaudeau

B. Editing and Imputation
Description: Under this project, our staff provides advice, develops computer edit/imputation systems in support of demographic and economic projects, implements prototype production systems, and investigates edit/imputation methods.

Highlights: During FY 2017, staff completed the publication of a paper (Thibaudeau, Slud, and Gottschalck 2017) on eliciting small-area estimation models that are also useful for imputation. The models exploit log-linear constructs based on elicited conditional independence relationships identified in longitudinal Surveys, such as the Survey of Income and Program Participation. Staff extended the longitudinal models for general item imputation and missing data compensation. The later involves deriving estimators with “good statistical properties” to predict or estimate longitudinal outcomes, rather than substituting “snapshots” for missing items.

Staff also presented and revised a sequel paper at the World Statistics Congress in Marrakech, Morocco in July 2017. Staff submitted a proposal for a book chapter (*Methodology of Longitudinal Surveys*, Peter Lynn ed.) in a monograph dedicated to all issues in longitudinal surveys, including missing data and longitudinal imputation. Staff collaborated on the efforts to model and impute earnings on the American Community Survey. Staff helped review the state of the art in imputation and modeling and identified possible avenues. Staff continued to collaborate and support efforts to define in more details the imputation, count, and item operations for the 2020 Census.

Staff: Yves Thibaudeau (x31706), Maria Garcia, Martin Klein, Darcy Steeg Morris, Bill Winkler

C. Developing Economic Census Synthetic Microdata

Description: The purpose of this project is to develop synthetic industry-level Economic Census microdata that satisfies all edits and privacy restrictions, produce the same tabulations as the true data, are usable for other economic research purposes, and can be publicly released in place of suppressed estimates. Staff plans to implement existing edit/imputation/synthesis software developed by Hang Kim et al. (2015).

Highlights: During FY 2017, staff collaborated in research planning, participated in method/software training sessions, and used the software to edit, impute, and synthesize simulated test data. Staff used software developed by Professor Hang Kim that implements a nonparametric Bayesian method to produce multiple synthetic datasets given a set of “clean” data (edited/imputed.) Staff developed a two-phase research plan; Part I will be restricted to general statistics items and Part II will include all other economic census items, including sampled items. Staff worked on input/output data requirements, edit validation requirements, and industry selection requirements. Staff discussed approaches to formal privacy validation measures, preliminary evaluation measures, and planned how to ensure synthetic data products are consistent with the complete set of edit rules imposed by the Economic Census.

Staff: Maria Garcia (x31703), Yves Thibaudeau

Record Linkage

Motivation: Record linkage is intrinsic to efficient, modern survey operations. It is used for unduplicating and updating name and address lists. It is used for applications such as matching and inserting addresses for geocoding, coverage measurement, Primary Selection Algorithm during decennial processing, Business Register unduplication and updating, re-identification experiments verifying the confidentiality of public-use microdata files, and new applications with groups of administrative lists. Significant theoretical and algorithmic progress (Winkler 2004ab, 2006ab, 2008, 2009a; Yancey 2005, 2006, 2007, 2011) demonstrates the potential for this research. For cleaning up administrative records files that need to be linked, theoretical and extreme computational results (Winkler 2010, 2011b) yield methods for editing, missing data and even producing synthetic data with valid analytic properties and reduced/eliminated re-identification risk. Easy means of constructing synthetic make it straightforward to pass files among groups.

Research Problems: The research problems are in three major categories. First, we need to develop effective ways of further automating our major record linkage operations. The software needs improvements for matching large sets of files with hundreds of millions of records against other large sets of files. Second, a key open research question is how to effectively and automatically estimate matching error rates. Third, we need to investigate how to develop effective statistical analysis tools for analyzing data from groups of administrative records when unique identifiers are not available. These methods need to show how to do correct demographic, economic, and statistical analyses in the presence of matching error.

Potential Applications: Presently, the Census Bureau is contemplating or working on many projects involving record linkage. The projects encompass the Demographic, Economic, and Decennial areas.

A. Disclosure Avoidance for Microdata

Description: Our staff investigates methods of microdata masking that preserves analytic properties of public-use microdata and avoid disclosure.

Highlights: During FY 2017, staff reviewed papers from the privacy literature. Most of the papers were on differential privacy. Staff investigated how some of the current methods (Winkler 2008, 2010) might be applied to a Decennial Census short-form data set. The methods are sufficiently fast for national files (100+ times as fast as methods are other statistical institutes or commercial software). The methods need extension to differential privacy and drastic speeded-up in the new context.

Staff also reviewed papers on approximate- nearest-neighbor methods of computation. The methods have been used in a few situations for speeding up creation of synthetic data that satisfies differential privacy restraints (Park and Ghosh 2012). The methods (Shi et al. JMLR 2009, Weinberger et al. 2010) have also been used for analyzing very large Internet databases that are larger than sets of national files.

Staff provided 15 of 22 of the programs for modeling/edit/imputation that can be used for synthetic-data generation to Professor Jerry Reiter of Duke and Professor Cynthia Dwork of Harvard. Although three of the programs were not available (due to retirement), they could be replaced by other programs that are straightforward to write because they do not involve complicated, hierarchical age comparisons within households. The issue with any synthetic data is not providing contradictory data such as a child under 16 being married.

One staff member provided information on controlled rounding in 3 or more dimensions (Deville and Tille *Biometrika* 1999, 2004; Winkler 1986, 1987, 2001) that can be used (with extreme difficulty) for adjusting synthetic data to published tables. The work compares to Wei and Reiter (2017, under review) that presents methods for synthetic data and rounding.

Staff: William Winkler (x34729)

B. Record Linkage and Analytic Uses of Administrative Lists
Description: Under this project, staff will provide advice, develop computer matching systems, and develop and perform analytic methods for adjusting statistical analyses for computer matching error.

Highlights: During FY 2017, staff reviewed papers on adjusting statistical analyses for linkage error. Staff reviewed three Ph.D. dissertations from students of Professor Jerry Reiter related to adjusting statistical analyses for linkage error and for privacy/confidentiality. Some of the methods are related to Manrique-Vallier and Reiter (*JASA* 2017) on modeling/edit/imputation and to Winkler (1997, 2003, 2008, 2010).

One staff member reviewed papers on comparing data from two data sources when there is no linkage error. Most of the papers are from the European statistical institutes (De Waal et al. 2014-2017) plus papers from Raghunathan and Schenkar (2005) and Meng (2008-2017). The work by Meng and his students at Harvard is particularly impressive. These very new methods greatly extend multiple imputation. In particular, Meng (2014) provides new metrics when administrative data can be combined effectively with a well-done survey sample. The metrics show that the administrative data must be exceptionally clean in terms of missing and erroneous data and cover 95+% of its universe (conditions that currently are very difficult to verify). Any new extensions of the methods for adjusting statistical analysis for linkage error would need to account for linkage error, the edit/imputation needed for cleaning up individual files, and these new theoretical methods of Meng and his colleagues.

Staff provided advice, certain summaries, and papers to a large number of individuals within the Decennial Directorate and Research and Methodology Directorate. A challenge is using record linkage software originally written by Census Bureau staff who have retired. Much is written in SAS.

Staff: William Winkler (x34729), Ned Porter, Emanuel Ben-David

C. Modeling, Analysis, and Quality of Data
Description: Our staff investigates methods of the quality of microdata primarily via modeling methods and new software techniques that accurately describe one or two of the analytic properties of the microdata.

Highlights: During FY 2017, staff looked at a large number of papers on approximate nearest neighbor matching that provide new methods in the record linkage and edit/imputation work. A Schedule A appointee has new methods of locality-sensitive hashing to the blocking problem in record linkage. The new methods, among other things, allow easier estimation of the proportion of true matches within a set of pairs during record linkage operations. Staff looked at methods for parallel computation. One staff member gave a guest lecture on Gaussian Bayesian networks in the data science program at George Washington University.

One staff member circulated a document showing examples of how we developed new faster algorithms making our generalized software suitable for many large analyses and production systems. Some of these algorithms are still too slow for some of our largest situations with national files. Because of the manner in which the likelihood and optimization computation is performed, these methods are not presently suitable for computing in the cloud. To extend the concepts, we would need experts in those situations where the methods can be extended to the cloud.

Also during FY 2017, staff compiled a list of the seventeen papers over the last fifty-two years dealing with adjusting statistical analyses for linkage error. Staff is using approximately one hundred additional papers in this research. Staff reviewed heuristic methods of editing. Staff ran a seminar to arrive at a set of record linkage research problems to investigate. The first problem is estimating record linkage error (generally in the unsupervised learning situation) which is known as the "regression problem" in machine learning when there is a large amount of training data. Another problem is how to adjust statistical analyses for record linkage error.

Staff submitted two papers to *Bayesian Analysis* and *Bernoulli Journal*. Both papers are currently under revisions. Staff initiated a collaboration with a researcher from George Mason University working on regression analysis of linked files. The draft of a paper is expected soon. Staff also reviewed a paper for the Center for Economic Studies (CES).

Staff: William Winkler (x34729), Xiaoyun Lu, Ned Porter, Emanuel Ben-David, Maria Garcia

D. R Users Group

Description: The initial objective of the R Users Group is to identify the areas of the Census Bureau where R software is developed and those other areas that could benefit from such development. The scope of the topics is broad and it includes estimation, missing data methods, statistical modeling, Monte-Carlo and resampling methods. The ultimate goal is to move toward integrated R tools for statistical functionality at the Census Bureau.

Initially, the group will review basic skills in R and provide remedial instruction as needed. The first topic for deeper investigation is complex-survey infrastructure utilities, in particular an evaluation of the "Survey" package and its relevance at the Census Bureau in the context of weighing, replication, variance estimation and other structural issues.

Highlights: During FY 2017, staff attended presentations by Microsoft representatives on a commercial platform supporting development and production in the "R" environment. The platform has two parts. The first part is in the public domain and can be used free of charge. The second part includes proprietary software covering basic statistical functions. Staff may consider the software platform in the future.

Also during FY 2017, staff continued to meet with the diversity group and acting as Affinity Group Leader. Staff met with Robert Houser on the diversity staff to discuss possible directions for the "R Users" group.

Staff: Yves Thibaudeau (x31706), Chad Russell

Small Area Estimation

Motivation: Small area estimation is important in light of a continual demand by data users for finer geographic detail of published statistics. Traditional demographic surveys designed for national estimates do not provide large enough samples to produce reliable direct estimates for small areas such as counties and even most states. The use of valid statistical models can provide small area estimates with greater precision, however bias due to an incorrect model or failure to account for informative sampling can result. Methods will be investigated to provide estimates for geographic areas or subpopulations when sample sizes from these domains are inadequate.

Research Problems:

- Development/evaluation of multilevel random effects models for capture/recapture models.
- Development of small area models to assess bias in synthetic estimates.
- Development of expertise using nonparametric modeling methods as an adjunct to small area estimation models.
- Development/evaluation of Bayesian methods to combine multiple models.
- Development of models to improve design-based sampling variance estimates.
- Extension of current univariate small-area models to handle multivariate outcomes.

Potential Applications:

- Development/evaluation of binary, random effects models for small area estimation, in the presence of informative sampling, cuts across many small area issues at the Census Bureau.
- Using nonparametric techniques may help determine fixed effects and ascertain distributional form for random effects.
- Improving the estimated design-based sampling variance estimates leads to better small area models which assumes these sampling error variances are known.
- For practical reasons, separate models are often developed for counties, states, etc. There is a need to coordinate the resulting estimates so smaller levels sum up to larger ones in a way that correctly accounts for accuracy.
- Extension of small area models to estimators of design- base variance.

A. Coverage Properties of Confidence Intervals for Proportions in Complex Surveys

Description: This is primarily a simulation project to investigate the coverage behavior of confidence intervals for proportions estimated in complex surveys. The goal is ultimately to inform recommendations for interval estimates in the American Community Survey (ACS), so the issues of main interest are:
(i) whether the current Wald-type intervals (defined as a point-estimator plus or minus a margin-or-error (MOE) estimate) can be improved by empirical-Bayes modifications or by modified forms of intervals known to perform well in the setting of binomial proportion-estimators, (ii) whether failures of coverage in a simulated complex survey can be ascribed to poor estimation of effective sample size or to other aspects of inhomogeneity and clustering in proportions within realistically complex populations, and (iii) whether particular problems arising with coverage of intervals for small proportions can be overcome. Future research might address whether the confidence interval methods developed for single-domain design-based estimates can also be adapted to small area estimates that borrow strength across domains.

Highlights: During FY 2017, staff extensively studied a method of computing the sampling variance for complex sample surveys that staff derived in FY 2016 using the idea of anticipated variance (Isaki and Fuller, 1982). Staff refer to this method as the Kish2 method, as under some conditions this method reduces to the Kish method for estimating variances for estimation of design effects. Simulation studies revealed that this method yielded variance estimators with lower variances than the traditional design-based estimator in the scenarios we studied. Staff also studied the impact of using the resulting estimates to compute confidence intervals for proportions for complex surveys under several different methods. Staff found that the Kish2 method alleviated the undercoverage problem inherent in confidence intervals for proportions when there is clustering. Moreover, simulations suggest that using this method may be more efficient than using the ad-hoc modifications to the effective sample size in Dean and Pagano (2016). In particular, over the factorial design of the simulation study, there are many cases in which the Kish2 method has higher coverage and lower length than applying the Dean and Pagano modification to the design-based estimate of effective sample size, while there are hardly any cases in which the Dean and Pagano modification yields higher coverage and lower length than the Kish2 method.

Staff wrote a paper on the results and submitted it to a peer-reviewed journal. In the process, staff created several exhibits and data visualization tools to compare the different methods of improving the effective sample size estimation and the different methods of interval construction, including figures that simultaneously compare the length and coverage of different methods. These shed further light on the relationship between the Dean and Pagano modification to the effective sample size in comparison to the approach of improving the sampling variance using a superpopulation model.

Staff: Carolina Franco (x39959), Eric Slud, Thomas Louis (Johns Hopkins University), Rod Little (University of Michigan)

B. Small Area Estimates of Disability

Description: This project is from the Development Case proposal to create subnational estimates of specific disability characteristics (e.g., number of people with autism). This detailed data is collected in a supplement of the Survey of Income and Program Participation (SIPP). However, the SIPP is only designed for national level estimates. This project is to explore small area models to combine SIPP with the large sample size of the American Community Survey to produce state and county level estimates of reasonable quality.

Highlights: During FY 2017, staff revised a manuscript based on referee comments from the *Journal of the Royal Statistical Society – Series A*. The revised paper has been accepted for publication and will appear in late 2017 or early 2018. Staff also gave a presentation of this research at the 2017 Small Area Conference in Paris, France on July 10, 2017.

Staff: Jerry Maples (x32873), Amy Steinweg (SEHSD)

C. Using ACS Estimates to Improve Estimates from Smaller Surveys via Bivariate Small Area Estimation Models

Description: Staff will investigate the use of bivariate area-level models to improve small area estimates from one survey by borrowing strength from related estimates from a larger survey. In particular, staff will explore the potential of borrowing strength from estimates from the American Community Survey, the largest U.S. household survey, to improve estimates from smaller U.S. surveys, such as the National Health Interview Survey, the Survey of Income and Program Participation, and the Current Population Survey.

Highlights: No significant progress during FY 2017.

Staff: Carolina Franco (x39959), William R. Bell (ADRM)

D. Multivariate Fay-Harriot Hierarchical Bayesian Estimation of Small Area Means under Functional Measurement Error

Description: Area-level models have been extensively used in small area estimation to produce model-based estimates of a population characteristic for small areas (e.g., Fay and Herriot, 1979). Multivariate area level models have also been used to jointly model multiple characteristics of correlated responses (e.g., Huang and Bell, 2012, Franco and Bell, 2015). Such models may lead to more precise small area estimates than separate univariate modeling of each characteristic. Typically both univariate and multivariate small area estimation models use auxiliary information to borrow strength from other areas and covariates associated with a response variable or a response vector. However, auxiliary variables are sometimes measured or obtained from sample surveys and are subject to measurement or sampling error. Researchers recognized that ignoring measurement error in the covariates and using standard solutions developed for covariates measured without error may lead to suboptimal inference. It was demonstrated in the univariate small area estimation setup that this naïve approach can result in model-based small area estimators that are more variable than the direct estimators when some of the covariate values in a small area are measured with substantial error (cf. Ybarra and Lohr, 2008, *Biometrika*; Arima, Datta and Liseo, 2015, *Scandinavian Journal of Statistics*). We are investigating a multivariate Fay-Herriot model and develop Bayes small area estimates when one or more auxiliary variables are measured with error. We work out a hierarchical Bayesian analysis for the multivariate Fay-Herriot model with a functional measurement error treatment for the covariates measured with error.

Highlights: During FY 2017, staff completed two revisions of the paper submitted in FY 2016, and the paper was accepted for publication to *the Journal of the Royal Statistical Society-Series A*. When estimating year-to-year changes using small area measurement error models, staff showed the importance of jointly modeling both years rather than taking the difference of estimates derived from univariate models for each year.

For multivariate functional measurement error models, staff compared the "true" vs. "reported" mean squared errors of the naive model predictions using a first order approximation that assumes the measurement error model is correct and the model parameters are known. Staff derived the approximations and then applied them to data related to school-aged children in poverty, replacing the model parameters in these expressions with their posterior means from Markov Chain Monte Carlo (MCMC) simulations. This sheds some light on the impact of ignoring measurement error in small area estimation models, which can result in misleading estimates of the mean squared error.

Staff also performed analytical and simulation studies to understand the impact of using a structural versus a functional measurement error model. The functional measurement error model assumes that the true values of the covariate that is measured without error are fixed but unknown quantities. The structural measurement error model assumes that the true values follow a model, formulated as a multivariate model for the covariates and the original dependent variable(s). This distinction has important implications on the mean squared errors of the resulting estimates. Staff derived results regarding the asymptotic bias and the asymptotic mean squared errors of each of three predictors, the functional ME, the structural ME, and a naïve estimator, under the assumptions that either the functional measurement error model is true, or the structural measurement error is true. Staff devised a simulation with a factorial design to further understand the problems. Staff illustrated the theoretical results using SAIPE data. Staff delivered an invited talk on the subject in a conference in honor of J. N. K. Rao's 80th birthday. Staff also presented this work in the 2017 Small Area Estimation Conference in Paris.

Staff: Carolina Franco (x39959), Gauri Datta, William R. Bell (ADRM)

Survey Sampling-Estimation and Modeling

Motivation: The demographic sample surveys of the Census Bureau cover a wide range of topics but use similar statistical methods to calculate estimation weights. It is desirable to carry out a continuing program of research to improve the accuracy and efficiency of the estimates of characteristics of persons and households. Among the methods of interest are sample designs, adjustments for non-response, proper use of population estimates as weighting controls, small area estimation, and the effects of imputation on variances.

The Economic Directorate of the Census Bureau encounters a number of issues in sampling and estimation in which changes might increase the accuracy or efficiency of the survey estimates. These include, but are not restricted to, a) estimates of low-valued exports and imports not currently reported, b) influential values in retail trade survey, and c) surveys of government employment.

The Decennial Census is such a massive undertaking that careful planning requires testing proposed methodologies to achieve the best practical design possible. Also, the U.S. Census occurs only every ten years and is the optimal opportunity to conduct evaluations and experiments with methodologies that might improve the next census. Sampling and estimation are necessary components of the census testing, evaluations, and experiments. The scale and variety of census operations require an ongoing research program to achieve improvements in methodologies. Among the methods of interest are coverage measurement sampling and estimation, coverage measurement evaluation, evaluation of census operations, uses of administrative records in census operations, improvements in census processing, and analyses that aid in increasing census response.

Research Problems:

- How can methods making additional use of administrative records, such as model-assisted and balanced sampling, be used to increase the efficiency of household surveys?
- Can non-traditional design methods such as adaptive sampling be used to improve estimation for rare characteristics and populations?
- How can time series and spatial methods be used to improve ACS estimates or explain patterns in the data?
- Can generalized weighting methods be implemented via optimization procedures that allow better understanding of how the various steps relate to each other?
- Some unusual outlying responses in the surveys of retail trade and government employment are confirmed to be accurate, but can have an undesired large effect on the estimates - especially estimates of change. Procedures for detecting and addressing these influential values are being extended and examined through simulation to measure their effect on the estimates, and to determine how any such adjustment best conforms with the overall system of estimation (monthly and annual) and benchmarking.
- What models aid in assessing the combined effect of all the sources of estimable sampling and nonsampling error on the estimates of population size?
- How can administrative records improve census coverage measurement, and how can census coverage measurement data improve applications of

administrative records?

- What analyses will inform the development of census communications to encourage census response?
- How should a national computer matching system for the Decennial Census be designed in order to find the best balance between the conflicting goals of maximizing the detection of true duplicates and minimizing coincidental matches? How does the balance between these goals shift when modifying the system for use in other applications?
- What can we say about the additional information that could have been obtained if deleted census persons and housing units had been part of the Census Coverage Measurement (CCM) Survey?

Potential Applications:

- Improve estimates and reduce costs for household surveys via the introduction of additional design and estimation procedures.
- Produce improved ACS small area estimates through the use of time series and spatial methods.
- Apply the same weighting software to various surveys.
- New procedures for identifying and addressing influential values in the monthly trade surveys could provide statistical support for making changes to weights or reported values that produce more accurate estimates of month-to-month change and monthly level. The same is true for influential values in surveys of government employment.
- Provide a synthesis of the effect of nonsampling errors on estimates of net census coverage error, erroneous enumerations, and omissions and identify the types of nonsampling errors that have the greatest effects.
- Describe the uncertainty in estimates of foreign-born immigration based on American Community Survey (ACS) used by Demographic Analysis (DA) and the Postcensal Estimates Program (PEP) to form estimates of population size.
- Improve the estimates of census coverage error.
- Improve the mail response rate in censuses and thereby reduce the cost.
- Help reduce census errors by aiding in the detection and removal of census duplicates.
- Provide information useful for the evaluation of census quality.
- Provide a computer matching system that can be used with appropriate modifications for both the Decennial Census and several Decennial-related evaluations.

A. Household Survey Design and Estimation
[See Demographic Projects]

B. Sampling and Estimation Methodology: Economic Surveys
Description: The Economic Directorate of the Census Bureau encounters a number of issues in sampling and estimation in which changes might increase the accuracy or efficiency of the survey estimates. These include estimates of low-valued exports not currently reported, alternative estimation for the *Quarterly Financial Report*, and procedures to address nonresponse and reduce respondent burden in the surveys. Further, general simulation software might be created and structured to eliminate various individual research efforts. An observation is considered influential if the estimate of total monthly revenue is dominated by its weighted contribution. The goal of the research is to find methodology that uses the observation but in a manner that assures its contribution does not dominate the estimated total or the estimates of period-to-period change.

Highlights: During FY 2017, staff collaborated with staff in the Economic Directorate to revise a paper on an innovative solution to the basic and previously unanswered question of how to develop initial settings for the parameters required to implement M-estimation methodology for detecting and treating verified influential values in economic surveys. The economic populations of interest are highly skewed and are consequently highly stratified, making normal distribution theory inapplicable. The most challenging problem was to develop an “automatic” data-driven method for setting the initial value of the tuning constant φ, the parameter with the greatest influence on performance of the algorithm. Of all the methods considered, the team found that the methods defined in terms of the requirement for the accuracy of published estimates, namely coefficients of variation and standard errors, yielded the best performance when judged in terms of lack of convergence issues for the algorithm and appropriate detections. In addition, these methods can be implemented on a large scale for a wide variety of population distributions. Preparations are continuing for a side-by-side test of the methodology in the near future so that Monthly Wholesale Trade Survey (MWTS) staff can evaluate the effectiveness of the methodology when using it in a production setting.

Staff: Mary Mulry (x31759)

C. The Ranking Project: Methodology Development and Evaluation
Description: This project undertakes research into the development and evaluation of statistical procedures for using sample survey data to rank several populations with respect to a characteristic of interest. The research includes an investigation of methods for quantifying and presenting the uncertainty in an estimated ranking of populations. As an example, a series of ranking tables are released from the American Community Survey in which the fifty states and the District of Columbia are ordered based on estimates of certain characteristics of interest.

Highlights: During the first quarter of FY 2017, staff continued to develop and experiment with various visualizations of rankings while showing uncertainty in the rankings mostly using the bootstrap. Work continued of a related draft paper providing details of theory and methodology. During the second quarter of FY 2017, staff worked on the revision of a submitted paper on visualization of comparisons of pairs and groups of pairs. The aim is to provide conditions under which the visual is correctly interpreted by general users. For example, if confidence intervals for a population one parameter and another population two parameter do not overlap, the user wants to say that the populations are different. On the other hand, if the confidence intervals do overlap, the user does not want to say that the populations are different. We summarize the literature, provide some clarifying theoretical results with proofs, and provide visualizations.

Staff continued to draft a paper on uncertainty in rankings using the bootstrap. Efforts continue to focus on increasing the theoretical foundation of the method as well as understanding what the results mean. A basic visualization that expresses uncertainty has been determined.

Near the end of FY 2017, staff stumbled upon a novel and simple method of constructing a joint confidence region for a ranking of K population that is very promising.

Staff: Tommy Wright (x31702), Martin Klein, Jerzy Wieczorek (Carnegie Mellon University), Brett Moran, Nathan Yau, Michael Leibert

D. Sampling and Apportionment

Description: This short-term effort demonstrated the equivalence of two well-known problems–the optimal allocation of the fixed overall sample size among L strata under stratified random sampling and the optimal allocation of the H = 435 seats among the 50 states for the apportionment of the U.S. House of Representatives following each decennial census. This project continues development with new sample allocation algorithms.

Highlights: Staff continued detailed comparisons of the Economic Directorate's overall sample size determinations and optimal allocations approach with the exact optimal allocation approach using Algorithms III (mixed constraints) and IV (stated precision) of Wright (2012, 2014). Staff also worked on the revision of a submitted paper on exact optimal allocation. Staff also worked on theory for exact optimal allocation using cost constraints. Results on carlier algorithms were published (Wright, 2017).

Staff: Tommy Wright (x31702), Michael Leibert

E. Analysis and Estimation of Daily Response Propensities and Use of Contact History Instrument (CHI)

Description: Staff continue to use general research methodology to work on existing files to improve modeling accuracy and to provide suggestions based on information gathered from the National Crime Victimization Survey (NCVS) using the Contact History Instrument (CHI). Staff conducted discrete-time event history analysis to fit daily response propensities in the NCVS by: (1) specifying a suitable model for discrete-time hazard logistic regression model; (2) using NCVS CHI data to estimate the daily response propensity model parameters; (3) interpreting results in terms of daily response propensity research questions; (4) evaluating model fit, hypothesis test and constructed confidence intervals for model parameters; and (5) communicating our findings, modeling, and data limitations.

Highlights: During FY 2017, staff used existing files to improve on methods to fit and evaluate models that can predict daily response propensities in the National Crime Victimization Survey (NCVS). Staff updated existing methodology to describe how to fit daily response propensities along with actual survey indicators and survey outcomes to (1) evaluate model accuracy and determine whether the models need refinement; (2) investigate relationship between response propensity and key survey variables; and (3) determine how the daily response propensities may be used to manage fieldwork.

A formal report of the analysis and estimation documenting work on NCVS daily response propensity modeling and methodology was completed during this fiscal year. The report was developed with the goal to provide additional information to serve as a guideline for field representatives (FRs) to create intervention for NCVS fieldwork for cases with low response propensities. The completed report for the daily response propensities from propensity models can be used to strategically direct FR efforts to improve NCVS survey response rates and data quality and to reduce survey costs.

Staff: Isaac Dompreh (x36801), Joseph Schafer (ADRM)

Time Series and Seasonal Adjustment

Motivation: Seasonal adjustment is vital to the effective presentation of data collected from monthly and quarterly economic surveys by the Census Bureau and by other statistical agencies around the world. As the developer of the X-13ARIMA-SEATS Seasonal Adjustment Program, which has become a world standard, it is important for the Census Bureau to maintain an ongoing program of research related to seasonal adjustment methods and diagnostics, in order to keep X-13ARIMA-SEATS up-to-date and to improve how seasonal adjustment is done at the Census Bureau.

Research Problems:

- All contemporary seasonal adjustment programs of interest depend heavily on time series models for trading day and calendar effect estimation, for modeling abrupt changes in the trend, for providing required forecasts, and, in some cases, for the seasonal adjustment calculations. Better methods are needed for automatic model selection, for detection of inadequate models, and for assessing the uncertainty in modeling results due to model selection, outlier identification and non-normality. Also, new models are needed for complex holiday and calendar effects.
- Better diagnostics and measures of estimation and adjustment quality are needed, especially for model-based seasonal adjustment.
- For the seasonal, trading day and holiday adjustment of short time series, meaning series of length five years or less, more research into the properties of methods usually used for longer series, and perhaps into new methods, are needed.

Potential Applications:

- To the effective presentation of data collected from monthly and quarterly economic surveys by the Census Bureau and by other statistical agencies around the world.

A. Seasonal Adjustment

Description: This research is concerned with improvements to the general understanding of seasonal adjustment and signal extraction, with the goal of maintaining, expanding, and nurturing expertise in this topic at the Census Bureau.

Highlights: During FY 2017, staff (a) developed a reconciliation procedure for ensuring aggregates of time series have adequate adjustments, for use with enforcing adequacy of GDP, and completed a paper documenting the method; (b) developed signal extraction diagnostics methodology and software, incorporated into a report on seasonal adjustment diagnostics; (c) continued research into high-dimensional spatio-temporal seasonal adjustment; (d) developed theory and tools for understanding multivariate seasonal adjustment in the frequency domain; (e) developed code and written materials for a book on multivariate direct filter analysis; (f) developed algorithms for calculating multivariate seasonal adjustment and cycle estimation in the frequency domain; (g) developed algorithms and code for multivariate multi-step ahead forecasting error covariances, with applications to signal extraction uncertainty; and (h) participated in and led ISAT (Interagency Seasonal Adjustment Team) group A, with mission to guide other federal agencies as to best seasonal adjustment practices, and generate new research on diagnostics.

Staff: Tucker McElroy (x33227), James Livsey, Brian Monsell, Osbert Pang, Anindya Roy

B. Time Series Analysis

Description: This research is concerned with broad contributions to the theory and understanding of discrete and continuous time series, for univariate or multivariate time series. The goal is to maintain and expand expertise in this topic at the Census Bureau.

Highlights: During FY 2017, staff (a) continued work on stable parametrization of VARMA models fitted under parameter constraints and stability constraints; (b) completed research on co-integration tests, finalizing theory, simulations, and data analysis; (c) finalized work on computing residual entropy, and devised new method for calculating Gaussian orthant probabilities; (d) continued work on Frobenius norm tool, developing more application of method-of-moments estimators, such as a test for white noise; (e) continued implementation of vector band pass filters, utilizing canonical trends and cycles to get adequate modeling of series; (f) illustrated how multivariate bullwhip can be greater or less than the implied univariate effect; (g) extended methods and software for fitting high-dimensional VAR and tested empirical performance on QWI data; (h) developed simulation modeling of information rigidity behaviors, completing a model and paper of forecaster agent behavior according to attentive and inattentive personas; (i) obtained theoretical and empirical results on nonlinear prediction for time series forecasting, using Hermite polynomials; and (j) obtained initial results for quadratic prediction, providing a basis for nonlinear filtering.

Staff: Tucker McElroy (x33227), David Findley (Private Collaborator), Brian Monsell, James Livsey, Osbert Pang, Anindya Roy

C. Time Series Model Development

Description: This work develops a flexible integer-valued autoregressive (AR) model for count data motivated by the Conway-Maxwell-Poisson distribution thus allowing for data over- or under-dispersion (i.e. count data where the variance is larger or smaller than the mean, respectively). Such a model will contain Poisson and negative binomial AR models as special cases.

Highlights: During FY 2017, staff continued to develop theoretical results and computational codes in R to analyze relevant data. Staff are writing a manuscript on both the statistical methodology and results from this work to submit to a journal for publication.

Staff: Kimberly Sellers (x39808)

Experimentation and Statistical Modeling

Motivation: Experiments at the Census Bureau are used to answer many research questions, especially those related to testing, evaluating, and advancing survey sampling methods. A properly designed experiment provides a valid, cost-effective framework that ensures the right type of data is collected as well as sufficient sample sizes and power are attained to address the questions of interest. The use of valid statistical models is vital to both the analysis of results from designed experiments and in characterizing relationships between variables in the vast data sources available to the Census Bureau. Statistical modeling is an essential component for wisely integrating data from previous sources (e.g., censuses, sample surveys, and administrative records) in order to maximize the information that they can provide.

Research Problems:

- Investigate bootstrap methodology for sample surveys; implement the bootstrap under complex sample survey designs; investigate variance estimation for linear and non-linear statistics and confidence interval computation; incorporate survey weights in the bootstrap; investigate imputation and the bootstrap under various non-response mechanisms.
- Investigate methodology for experimental designs embedded in sample surveys; investigation of large-scale field experiments embedded in ongoing surveys; design based and model based analysis and variance estimation incorporating the sampling design and the experimental design; factorial designs embedded in sample surveys and the estimation of interactions; testing non-response using embedded experiments. Use simulation studies.
- Assess feasibility of established design methods (e.g., factorial designs) in Census Bureau experimental tests.
- Identify and develop statistical models (e.g., loglinear models, mixture models, and mixed-effects models) to characterize relationships between variables measured in censuses, sample surveys, and administrative records.
- Assess the applicability of post hoc methods (e.g., multiple comparisons and tolerance intervals) with future designed experiments and when reviewing previous data analyses.

Potential Applications:

- Modeling approaches with administrative records can help enhance the information obtained from various sample surveys.
- Experimental design can help guide and validate testing procedures proposed for the 2020 Census.
- Expanding the collection of experimental design procedures currently utilized with the American Community Survey.

A. Design and Analysis of Embedded Experiments

Description: This ongoing project will explore rigorous analysis of embedded experiments: from simple idealized designs to complex designs used in practice at the Census Bureau.

Highlights: During FY 2017, staff began exploration of design-based analysis for embedded experiments, which takes into account both the sampling design and the experimental design. Staff also explored recent literature on randomization-based confidence intervals and tests in the context of causal inference, which appears to be relevant to the proposed research. Staff began a research report on embedded experiments methodology under complex sample designs, such as those utilized in the American Community Survey and the Current Population Survey.

Staff: Thomas Mathew (x35337), Andrew Raim, Robert Ashmead

B. Multivariate Nonparametric Tolerance Regions

Description: A tolerance region for a multivariate population is a region computed using a random sample that will contain a specified proportion or more of the population, with a given confidence level. Typically, tolerance regions that have been computed for multivariate populations are elliptical in shape. A difficulty with an elliptical region is that it cannot provide information on the individual components of the measurement vector. However, such information can be obtained if we compute tolerance regions that are rectangular in shape. This project applies bootstrap ideas to compute multivariate tolerance regions in a nonparametric framework. Such an approach can be applied to multivariate economic data and aid in the editing process by identifying multivariate observations that are outlying in one or more attributes and subsequently should undergo further review.

Highlights: No significant progress during FY 2017.

Staff: Thomas Mathew (x35337)

C. Development of a Bivariate Distribution for Count Data where Data Dispersion is Present

Description: This project develops a bivariate form of the Conway-Maxwell-Poisson distribution to serve as a tool to describe variation and association for two count variables that express over- or under-dispersion (relationships where the variance of the data is larger or smaller than the mean, respectively).

Highlights: During FY 2017, staff developed an R package (multicmp) associated with this work and submitted it to The Comprehensive R Archive Network (CRAN). The multicmp package is now available for download in CRAN.

Staff: Kimberly Sellers (x39808), Darcy Steeg Morris

D. Developing a Flexible Stochastic Process for Significantly Dispersed Count Data
Description: The Bernoulli and Poisson are two popular count processes; however, both rely on strict assumptions that motivate their use. CSRM staff (with other collaborators) instead propose a generalized count process (hereafter named the Conway-Maxwell-Poisson process) that not only includes the Bernoulli and Poisson processes as special cases, but also serves as a flexible mechanism to describe count processes that approximate data with over- or under-dispersion. Staff introduce the process and its associated generalized waiting time distribution with several real-data applications to illustrate its flexibility for a variety of data structures. This new generalized process will enable analysts to better model count processes where data dispersion exists in a more accommodating and flexible manner.

Highlights: During FY 2017, staff created an R package (cmpprocess) associated with this work, now available on CRAN. Staff also responded to a letter to the editor regarding a recently published manuscript. Staff also helped develop a research poster that was presented at the 2017 Joint Statistical Meetings regarding the R package, cmpprocess.

Staff: Kimberly Sellers (x39808), Darcy Steeg Morris

E. Master Address File (MAF) Research—Developing a Generalized Regression Model for Count Data
Description: This project develops a zero-inflated version of a generalized regression model for count data based on the Conway-Maxwell-Poisson distribution to allow for data-dispersion and excess zeroes in the dataset. The objective of this project is to develop and consider an alternative regression model for use to describe associations with changes in the number of housing units (adds or deletes) on a block, and predict where housing growth or decline may occur in the MAF.

Highlights: During FY 2017, staff updated the COMPoissonReg package in R to accommodate zero-inflated regression modeling via the Conway-Maxwell-Poisson distribution and loaded the updated package onto the Comprehensive R Archive Network (CRAN). The updated package is now available on CRAN.

Staff: Kimberly Sellers (x39808), Andrew Raim

F. Analysis of Under-dispersed Count Data
Description: This research concerns contributions to the theory and understanding of under-dispersed count data, and models that accommodate such data. The goal is to expand understanding and expertise in this area at the Census Bureau.

Highlights: During FY 2017, the related manuscript was successfully accepted for publication in *Communications in Statistics—Theory and Methods*; this manuscript is available online but not yet published. Staff meanwhile presented this work in an invited session at the 2017 International Chinese Statistical Association Applied Symposium in Chicago, IL.

Staff: Kimberly Sellers (x39808), Darcy Morris

G. Spatio-Temporal Change of Support
Description: Spatio-temporal change of support methods are used for statistical inference and prediction on space-time domains which differ from the domains on which the data were observed. Bradley, Wikle, and Holan (2015; Stat) proposed a parsimonious class of Bayesian hierarchical spatio-temporal models for Gaussian outcomes through a motivating application involving the American Community Survey (ACS). The goal of this project is to develop an R package to make the methodology broadly accessible to public users of Census Bureau data and to the general R user community.

Highlights: During FY 2017, staff completed initial development of an R package and began large-scale testing with ACS data. A number of issues with model specification and computation were identified and addressed, including: data preparation, model fitting through Markov Chain Monte Carlo (MCMC), computation of spatio-temporal basis functions, use of sparse versus dense matrices, and choice of autocovariance structure for spatio-temporal random effects. Staff presented work at the 2017 Joint Statistical Meetings and submitted a paper to the proceedings which demonstrates the software on public ACS data and investigates some options in model specification.

Staff: Andrew Raim (x37894), Scott Holan (ADRM)

H. Conway-Maxwell-Poisson Model for Longitudinal Count Data
Description: Repeated measures count data have an inherent within-subject correlation that is commonly modeled with random effects in the standard Poisson regression. While this model allows for over-dispersion via the nature of the repeated measures, departures from equi-dispersion can exist due to the underlying count process mechanism. This work considers a cross-sectional Conway-Maxwell-Poisson (CMP) regression model incorporating random effects for analysis of longitudinal count data.

Highlights: During FY 2017, staff established preliminary results, developing a CMP longitudinal model in SAS via the NLMIXED procedure. Accordingly, staff presented some of this work at the 2017 SAS Global Forum in Orlando, FL and submitted an associated proceedings paper entitled "Fitting a Flexible Model for Longitudinal Count Data Using the NLMIXED Procedure". The paper is available online. This talk and paper detail how to fit the model via adaptations of existing SAS procedures. Staff also submitted a proceedings paper entitled "A COM-Poisson Mixed Model with Normal Random Effects for Clustered Count Data", which was presented at the ISI World Statistics Congress 2017 in Marrakech, Morocco. This talk and paper provide details of the model including analysis of simulated data to illustrate its flexibility.

Staff: Darcy Steeg Morris (x33989), Kimberly Sellers

Simulation and Statistical Modeling

Motivation: Simulation studies that are carefully designed under realistic survey conditions can be used to evaluate the quality of new statistical methodology for Census Bureau data. Furthermore, new computationally intensive statistical methodology is often beneficial because it can require less strict assumptions, offer more flexibility in sampling or modeling, accommodate complex features in the data, enable valid inference where other methods might fail, etc. Statistical modeling is at the core of the design of realistic simulation studies and the development of intensive computational statistical methods. Modeling also enables one to efficiently use all available information when producing estimates. Such studies can benefit from software such as *Tea* for data processing. Statistical disclosure avoidance methods are also developed and properties studied.

Research Problems:

- Systematically develop an environment for simulating complex surveys that can be used as a test-bed for new data analysis methods.
- Develop flexible model-based estimation methods for survey data.
- Develop new methods for statistical disclosure control that simultaneously protect confidential data from disclosure while enabling valid inferences to be drawn on relevant population parameters.
- Investigate the bootstrap for analyzing data from complex sample surveys.
- Continue to formalize the codebase and user interfacing for *Tea*, especially within the context of the current enterprise environment.
- Develop models for the analysis of measurement errors in Demographic sample surveys (e.g., Current Population Survey or the Survey of Income and Program Participation).
- Identify and develop statistical models (e.g., loglinear models, mixture models, and mixed-effects models) to characterize relationships between variables measured in censuses, sample surveys, and administrative records.
- Investigate noise multiplication for statistical disclosure control.

Potential Applications:

- Simulating data collection operations using Monte Carlo techniques can help the Census Bureau make more efficient changes.
- Use noise multiplication or synthetic data as an alternative to top coding for statistical disclosure control in publicly released data. Both noise multiplication and synthetic data have the potential to preserve more information in the released data over top coding.
- Rigorous statistical disclosure control methods allow for the release of new microdata products.
- *Tea* provides modeling and editing flexibility, especially with a focus on incorporating administrative data.
- Using an environment for simulating complex surveys, statistical properties of new methods for missing data imputation, model-based estimation, small area estimation, etc. can be evaluated.
- Model-based estimation procedures enable efficient use of auxiliary information (for example, Economic Census information in business surveys), and can be applied in situations where variables are highly skewed and sample sizes are not sufficiently large to justify normal approximations. These methods may also be applicable to analyze data arising from a mechanism other than random sampling.
- Variance estimates and confidence intervals in complex surveys can be obtained via the bootstrap.
- Modeling approaches with administrative records can help enhance the information obtained from various sample surveys.

A. Development and Evaluation of Methodology for Statistical Disclosure Control

Description: When survey organizations release data to the public, a major concern is the protection of individual records from disclosure while maintaining quality and utility of the released data. Procedures that deliberately alter data prior to their release fall under the general heading of statistical disclosure control. This project develops new methodology for statistical disclosure control, and evaluates properties of new and existing methods. We develop and study methods that yield valid statistical analyses, while simultaneously protecting individual records from disclosure.

Highlights: During FY 2017, staff developed a new method that uses the statistical principle of sufficiency to generate synthetic data under the normal multiple linear regression model. Under this approach, the values of certain sufficient statistics are fixed, and synthetic data are generated from the conditional distribution of the original data given the sufficient statistics. Staff showed that an advantage of this approach is that if the regression model assumed by the synthetic data producer is correctly specified, then the synthetic data have the same joint distribution as the original data. Therefore, one can use standard regression methodology and software to analyze the synthetic data. Staff also showed that if the same regression model used to generate the synthetic data is also used for data analysis, and the data are analyzed using standard regression methodology, then the synthetic data yield identical inference as the original data. Staff also studied the effects of overfitting or underfitting the linear regression model. Staff showed that even if the data producer overspecifies the regression model when generating synthetic data, the synthetic data will still have the same distribution as the original data, and hence valid inference can be obtained. Staff also showed that if the data producer underspecifies the regression model, then one cannot expect to obtain valid inference from the synthetic data. Staff studied the disclosure risk of the proposed method.

Staff continued work on likelihood based analysis of singly and multiply imputed synthetic data. Staff developed this methodology under linear regression and multivariate linear regression models, and studied properties of the methodology when certain assumptions (assumed when deriving the methods) do not hold. Staff derived a new result about the asymptotic distribution of the sample mean of a synthetic data set generated using plug-in sampling.

Also during FY 2017, staff formulated and examined a strong view that a privacy mechanism should ensure that no intruder would gain much new information about any respondent from his response. Staff showed that only certain functions can be used to formalize this notion. Staff then defined a canonical strict privacy protection criterion, and obtained a complete characterization of all randomized response procedures that satisfy this criterion. Interestingly, staff found that any privacy specification amounts to putting an upper bound on all Bayes factors. Thus, privacy needs may be assessed most appropriately in terms of Bayes factors. Staff compared all privacy preserving procedures by data utility and identified the class of all admissible procedures. Staff also obtained the optimal privacy preserving procedure under a particular information preservation criterion.

Based on this work, staff prepared a manuscript entitled "A Criterion for Privacy Protection in Data Collection and Its Attainment via Randomized Response Procedures," which will be submitted for a journal publication.

Staff: Martin Klein (x37856), Bimal Sinha, Thomas Mathew, Brett Moran, Tapan Nayak, Gauri Datta

B. Analysis and Estimation of Generalized Propensity Scores with Bootstrap, Simulation, Continuous Treatment and Causal Inference Methods

Description: Staff is currently using general research methodology to work on a simulation study to describe how to produce a generalized boosted regression modeling algorithm for estimating propensity scores with bootstrap and continuous treatment methods. For this simulation study, staff is looking into how to estimate the generalized propensity scores for causal inference, potential outcomes and treatment outcomes for this research. As part of this research, staff continues to conduct comprehensive model diagnostics to confirm that the simulation data generated using parametric distributions meet the normality assumptions. Three or more simulation scenarios are being considered for treatment and potential outcome variables by using both parametric and nonparametric statistical modeling for estimation of the generalized propensity scores.

Highlights: During FY 2017, staff continued to work on how to use bootstrap resampling methods to fit parametric and non-parametric bootstrap simulation files (a) to compute relative risk ratio estimates; (b) to compute standard errors; (b) to check the accuracy of the usual Gaussian based methods; (c) to compute an accurate confidence intervals for a variety of statistics and (d) for a variety of complex sampling methods, and how to perform significance tests in some of these settings.

Staff continue to work on bootstrap and simulation methods to fit parametric and non-parametric statistical analysis to check sampling variability, standard error, bias, central limit theorem, significance tests and p-values. An internal draft documenting work on a simulation study and modeling for estimation of the generalized propensity scores for continuous treatment and causal inference is currently in progress.

Staff: Isaac Dompreh (x36801)

Summer at Census

Description: For each summer since 2009, recognized scholars in the following and related fields applicable to censuses and large-scale sample surveys are invited for short-term visits (one to five days) primarily between May and September: statistics, survey methodology, demography, economics, geography, social and behavioral sciences, and computer science. Scholars present a seminar based on their research and engage in collaborative research with Census Bureau researchers and staff.

Scholars are identified through an annual Census Bureau-wide solicitation by the Center for Statistical Research and Methodology.

Highlights: Staff facilitated all the details and background with staff from around the Census Bureau to host *2017 SUMMER AT CENSUS* with thirty-one scholars.

Staff: Tommy Wright (x31702), Michael Leibert

Research Support and Assistance

This staff provides substantive support in the conduct of research, research assistance, technical assistance, and secretarial support for the various research efforts.

Staff: Erica Magruder, Brett Moran, Kelly Taylor

3. PUBLICATIONS

3.1 JOURNAL ARTICLES, PUBLICATIONS

Abramowitz, J., O'hara, B., and Morris, D.S. (2017). "Risking Life and Limb: Estimating A Measure of Medical Care Economic Risk and Considering Its Implications," *Health Economics*, *26*, 469-485.

Arima, S., Bell, W. R., Datta, G. S., Franco, C., and Liseo, B. (In Press). "Multivariate Fay-Herriot Bayesian estimation of Small Area Means Under Functional Measurement Error," *Journal of the Royal Statistical Society--Series A*.

Ashmead, R., Slud, E., and Hughes, T. (In Press), "Adaptive Intervention Methodology for Reduction of Respondent Contact Burden in the American Community Survey," *Journal of Official Statistics.*

Blakely, C. and McElroy, T. (2017). "Signal Extraction Goodness-of-fit Diagnostic Tests Under Model Parameter Uncertainty: Formulations and Empirical Evaluation." *Econometric Reviews, 36(4):* 447-467.

Brown, D.A., Datta, G.S., and Lazar, N. (In Press). "A Bayesian Generalized CAR Model for Correlated Signal Detection," *Statistica Sinica*, DOI:10.5705/ss.202015.0382.

Datta, G.S., Delaigle, A., Hall, P., and Wang, L. (In Press). "Semi-parametric Prediction Intervals in Small Areas when Auxiliary Data are Measured with Error," *Statistica Sinica*. DOI:10.5705/ss.202016.0416.

Ghosal, S. and Roy, A. (2017). "Discussion of 'Should We Sample A Time Series More Frequently? Decision Support via Multirate Spectrum Estimation,'" by Nason, G., Powell, B., Elliott, D. and Smith, P. A. *Journal of Royal Statistical Society, Ser A*, 180: 393-394.

Holan, S., McElroy, T., and Wu, G. (2017). "The Cepstral Model for Multivariate Time Series: The Vector Exponential Model," *Statistica Sinica, 27:* 23-42.

Janicki, R. and Vesper, A. (2017). "Benchmarking Techniques for Reconciling Small Area Models at Distinct Geographic Levels," *Statistical Methods and Applications*. DOI: https://doi.org/10.1007/s10260-017-0379-x

Klein, M. and Datta, G. (In Press). "Statistical Disclosure Control Via Sufficiency Under the Multiple Linear Regression Model," *Journal of Statistical Theory and Practice*.

Livsey, J., Lund, R., Kechagias, S., and Pipiras, V. (In Press). "Multivariate Integer-valued Time Series with Flexible Autocovariances and Their Application to Major Hurricane Counts," *Annals of Applied Statistics*.

Lu, X. (In Press). "On Min-max Pair in Tournaments," *Journal of Graphs and Combinatorics*.

Lu, B. and Ashmead, R. (In Press). "Propensity Score Matching Analysis for Causal Effect with MNAR Covariates," *Statistica Sinica.*

Maples, J. (In Press). "Improving Small Area Estimates of Disability: Combining the American Community Survey with the Survey of Income and Program Participation," *Journal of the Royal Statistical Society—Series A*.

McElroy, T. (In Press). "Recursive Computation for Block Nested Covariance Matrices," *Journal of Time Series Analysis.*

McElroy, T. (2017). "Computation of Vector ARMA Autocovariances*," Statistics and Probability Letters*, *124*: 92-96.

McElroy, T. (2017). "Multivariate Seasonal Adjustment, Economic Identities, and Seasonal Taxonomy," *Journal of Business and Economics Statistics, 35(4):* 511-525.

McElroy, T. and McCracken, M. (2017). "Multi-step Ahead Forecasting of Vector Time Series," *Econometric Reviews, 36(5):* 495-513.

McElroy, T. and Roy A., (In Press). "The Inverse Kullback Leibler Method for Fitting Vector Moving Averages. *Journal of Time Series Analysis*.

Morris, D.S. (2017). “A Modeling Approach for Administrative Record Enumeration in the Decennial Census,” *Public Opinion Quarterly: Special Issue on Survey Research, Today and Tomorrow, 81(S1):* 357-384.

Moura, R., Klein, M., Coelho, C.A., and Sinha, B. (2017). “Inference for Multivariate Regression Model Based On Synthetic Data Generated Under Fixed-Posterior Predictive Sampling: Comparison With Plug-in Sampling,” *REVSTAT – Statistical Journal, 15(2)*: 155-186.

Mulry, M. H. and Keller, A. (2017). “Comparison of 2010 Census Nonresponse Follow-Up Proxy Responses with Administrative Records Using Census Coverage Measurement Results.” *Journal of Official Statistics, 33(2):* 455–475. DOI: https://doi.org/10.1515/jos-2017-0022.

Mulry, M.H., Oliver, B., Kaputa, S, and Thompson, K. J. (2016). “Cautionary Note on Clark Winsorization,” *Survey Methodology, 42 (2):* 297-305. http://www.statcan.gc.ca/pub/12-001-x/2016002/article/14676-eng.pdf.

Raim, A., Neerchal, N., and Morel, J. (In Press). “An Extension of Generalized Linear Models to Finite Mixture Outcome Distributions,” *Journal of Computational and Graphical Statistics*.

Roy, A., McElroy, T., and Linton, P. (In Press). “Estimation of Causal Invertible VARMA Models,” *Statistica Sinica.*

Sellers, K., Benn, E.K.T., Garcia, M., and Kellam, M. (2017). “Addressing Implicit Bias among Women Statisticians and Data Scientists,” *Chance, Vol. 30, No. 2.*

Sellers, K.F. and Morris, D.S. (In Press). “Under-dispersion Models: Models That Are ‘Under The Radar’,” *Communications in Statistics—Theory and Methods*. http://dx.doi.org/10.1080/03610926.2017.1291976.

Sellers, K.F., Morris, D.S., Shmueli, G., and Zhu, L. (2017). “Reply: Models for Count Data (A Response to A Letter to The Editor),” *The American Statistician, 71(2)*: 190

Thibaudeau, Y., Slud, E., and Gottschalck, A. (2017). “Modeling Log-Linear Conditional Probabilities for Estimation in Surveys,*” Annals of Applied Statistics, 11(2):* 680-697.

Trimbur, T. and McElroy, T. (2017). “Signal Extraction for Nonstationary Time Series With Diverse Sampling Rules,” *Journal of Time Series Econometrics, 9(1).*

Wright, T. (2017). “Exact Optimal Sample Allocation: More Efficient than Neyman,” *Statistics and Probability Letters, 129,* 50-57.

Wright, T., Klein, M., and Wieczorek, J. (In Press). “A Primer on Visualizations for Comparing Populations, Including the Issue of Overlapping Confidence Intervals,” *The American Statistician.*

Zhu, L., Sellers, K.F., Morris, D.S., and Shmueli, G. (2017). “Bridging the Gap: A Generalized Stochastic Process For Count Data,” *The American Statistician, 71 (1):* 71-80.

3.2 BOOKS/BOOK CHAPTERS

Christen, P. and Winkler, W. (2017). “Record Linkage,” In Sammut, C., and Webb, G. (Eds). *Encyclopedia of Machine Learning and Data Mining,* Springer.

Mulry, M.H, Nichols, E.M., and Hunter Childs, J. (2017). “Using Administrative Records Data at the U.S. Census Bureau: Lessons Learned from Two Research Projects Evaluating Survey Data, Part 2 of Establishing Infrastructure for the Use of Big Data to Understand Total Survey Error: Examples from Four Survey Research Organizations” In Biemer, P.P, Eckman, S., Edwards, B., Lyberg, L., Tucker, C., de Leeuw, E., Kreuter, F., and West, B.T. (Eds.), *Total Survey Error in Practice*. Wiley. New York. 467-473. DOI: 10.1002/9781119041702.ch21

Winkler, W. (In Press). “Cleaning and Using Administrative Lists: Enhanced Practices and Computational Algorithms for Record Linkage and Modeling/Editing/Imputation,” In Chun, A.Y. and Larson, M. (Eds). *Administrative Records for Survey Methodology,* New York, NY: Wiley.

3.3 PROCEEDINGS PAPERS

Statistics Canada Methodology Symposium, Gatineau, Quebec, Canada, March 22 – 24, 2016

- Mary Mulry, Elizabeth M. Nichols, and Jennifer Hunter Childs, "Using Administrative Records to Evaluate Survey Data." http://www.statcan.gc.ca/eng/conferences/symposium2016/program/14711-eng.pdf.

Fifth International Conference on Establishment Surveys (ICES-V). Geneva, Switzerland, June 21 – 24, 2016.

- Mary Mulry, Stephen Kaputa, and Katherine J. Thompson, "Setting Parameters for M-estimation." http://ww2.amstat.org/meetings/ices/2016/proceedings/024_ices15Final00260.pdf

Joint Statistical Meetings, American Statistical Association, Chicago, Illinois, July 31-August 4, 2016
2016 Proceedings of the American Statistical Association

- Robert Ashmead and Eric Slud, "Inference from Complex Survey-Embedded Field Experiments", 786-793.
- Krista Heim and Andrew Raim, "Predicting Coverage Error on the Master Address File Using Spatial Modeling Methods at the Block Level", 1541-1555.
- Yves Thibaudeau and Darcy Morris, "Bayesian Decision Theory to Optimize the Use of Administrative Records in Census NRFU", 472-478.
- Mary Mulry, Tom Mule, and Brian Clark, "Using the 2015 Census Test Evaluation Follow-up to Compare the Nonresponse Follow-up with Administrative Record", 503-516.
- Andrew Raim, "Informing Maintenance to the U.S. Census Bureau's Master Address File Using Statistical Decision Theory", 648-659.
- Dan Weinberg, Tucker McElroy, and Soumendra Lahiri, "Estimation of Locally Stationary Spatial Processes with Application to the American Community Survey".

SAS Global Forum Proceedings, SAS Institute, Cary, North Carolina, April 2-5, 2017

- Darcy Morris, Kimberly Sellers, and Austin Menger, "Fitting a Flexible Model for Longitudinal Count Data Using the NLMIXED Procedure." http://support.sas.com/resources/papers/proceedings17/0202-2017.pdf.

3.4 CENTER FOR STATISTICAL RESEARCH & METHODOLOGY RESEARCH REPORTS

<http://www.census.gov/srd/csrmreports/byyear.html>

SS (Statistics #2016-01): Claire McKay Bowen. "NSF GRIP Report: American Community Survey Simulation Study," October 25, 2016.

RR (Statistics #2017-01): Tucker S. McElroy and Anindya Roy. "Detection of Seasonality in the Frequency Domain," January 11, 2017.

RR (Statistics #2017-02): Tucker S. McElroy and Richard Penny. "Maximum Entropy-Value Seasonal Adjustment," January 11, 2017.

RR (Statistics #2017-03): David Findley, Demetra P. Lytras, and Tucker S. McElroy. "Detecting Seasonality in Seasonally Adjusted Monthly Time Series," February 13, 2017.

RR (Statistics #2017-04): Anirban Sanyal, Pratik Mitra, Tucker S. McElroy, Anindya Roy. "Holiday Effects in Indian Manufacturing Series," August 10, 2017.

RR (Statistics #2017-05): William R. Bell and Carolina Franco. "Small Area Estimation – State Poverty Rate Model Research Data Files," September 20, 2017.

RR (Statistics #2017-06): Tucker S. McElroy, Osbert Pang, and George Sheldon. "Custom Epoch Estimation for Surveys," September 29, 2017.

3.5 OTHER REPORTS

Adragni, K.P., Martin, S.R., Raim, A.M., and Huang, W. (2017). "ManifoldOptim: An R Interface to the 'ROPTLIB' Library for Riemannian Manifold Optimization," https://cran.r-project.org/package=ManifoldOptim.

Livsey, J. and Sax, C. (2017). “Interactive Stories on Seasonal Adjustment with X-13ARIMA-SEATS,” http://github.com/christophsax/x13story.

Raim, A.M. (2017). “mixlink: Mixture Link Regression,” https://cran.r-project.org/package=mixlink.

Sellers K.F., Morris D.S., Balakrishnan N., and Davenport, D. (2017) “multicmp: Flexible Modeling of Multivariate Count Data via the multivariate Conway-Maxwell-Poisson Distribution,” https://cran.r-project.org/web/packages/multicmp/index.html

Sellers, K.F., Lotze, T., and Raim, A. (2017). “COMPoissonReg: Conway-Maxwell-Poisson regression,” Version 0.4.0, 0.4.1, https://cran.r-project.org/web/packages/COMPoissonReg/index.html

Wieczorek, J. (2017). “Ranking Project: The Ranking Project: Visualizations for Comparing Populations,” R package version 0.1.1, https://cran.r-project.org/package=RankingProject [Complements Wright, Klein, and Wieczorek (In Press)].

Zhu, L., Sellers, K., Morris D., Shmueli, G., and Davenport, D. (2017). “cmpprocess: Flexible Modeling of Count Processes,” https://cran.r-project.org/web/packages/cmpprocess/index.html

4. TALKS AND PRESENTATIONS

International Conference on Advances in Interdisciplinary Statistics and Combinatorics, Greensboro, North Carolina, September 30–October 2, 2016.

- Tapan Nayak, "A Novel Approach to Setting a Strict Identity Disclosure Control Goal and Achieving It Efficiently."

George Washington University, Washington, D.C., October 7, 2016

- Tucker McElroy, "Testing Collinearity of Vector Time Series."

National Center for Health Statistics Seminar, Hyattsville, Maryland, October 2016.

- Eric Slud, "Design of Sample Surveys which Complement Observational Data to Achieve Population Coverage."

International Conference on Statistical Distributions and Applications (ICOSDA), Crowne Plaza, Niagara Falls, Canada, October 15, 2016

- Darcy Steeg Morris, "The Bivariate Conway-Maxwell-Poisson Distribution."
- Andrew Raim, "A Flexible Zero-inflated Model to Address Data Dispersion."
- Kimberly Sellers, "Don't Count on Poisson! Introducing the Conway-Maxwell-Poisson Distribution for Statistical Methodology regarding Count Data."

George Mason ASA Group, George Mason University, Fairfax, Virginia, October 27, 2016.

- Brian Monsell, "A (Mostly) Painless Introduction to Seasonal Adjustment."

Seasonal Adjustment Practitioners Workshop, Bureau of Labor Statistics, Washington, D.C., November 4, 2016

- James Livsey, "Learning and Discussing Seasonal Adjustment with R."
- Tucker McElroy, "Modeling and Seasonal Adjustment of Daily Retail Series."
- Brian Monsell, "So You Just Got 300 Files You Need To Seasonally Adjust."
- Thomas Trimbur, "The Effects of Seasonal Heteroskedasticity in Time Series on Trend Estimation and Seasonal Adjustment."

University of San Francisco, San Francisco, California, November 30, 2016

- James Livsey, "Integer-valued Time Series: Superposition Methods."

Data Science Program, Columbian College of Arts and Sciences, George Washington University, Washington, D.C., December 9, 2016.

- Emanuel Ben-David, "Gaussian Bayesian Nets."

Federal Economic Statistics Advisory Committee (FESAC) Meeting, U.S. Census Bureau, Washington, D.C., December 9, 2016.

- Tucker McElroy, "Challenges with Seasonal Adjustment."

Platinum Jubilee International Conference on Applications of Statistics , Department of Statistics, University of Calcutta, Kolkata, India, December 21-23, 2016.

- Tapan Nayak, "Privacy and Confidentiality Protection via Randomization."
- Tommy Wright, "No Calculations When Observations Can Be Made."

Department of Statistics, University of Missouri, Columbia, Missouri, February 22, 2017.

- Gauri Datta, "Robust Methods in Small Area Estimation."

2nd Annual International Time Series/Seasonal Adjustment Workshop hosted by the Office of National Statistics, Royal Statistical Society, London, England, March 7, 2017

- James Livsey, "Challenges with Seasonal Adjustment of GDP."

48th Southeastern International Conference on Combinatorics, Graph Theory & Computing, Florida Atlantic University, Boca Raton, Florida, March 7, 2017.

- Xiaoyun Lu, "On Min-max Pair in Tournaments."

4th African International Conference (AIC), University of Limpopo, Polokwane, South Africa, March 20-23, 2017.

- Tapan Nayak, “Confidentiality Protection in Mocrodata Release Using Unbiased Post-Randomization.”
- Anindya Roy, “Detection of Seasonality in the Frequency Domain.”
- Bimal Sinha, “Some Aspects of Data Analysis under Confidentiality Protection.”

Economic Area Methodology Seminar, U.S. Census Bureau, Washington, D.C., March 29, 2017

- Brian Monsell, “A Small Piece of X-13 Pi.”

SAS Global Forum 2017, Orlando, Florida, April 2-5, 2017.

- Darcy Steeg Morris, “Fitting a Flexible Model for Longitudinal Count Data Using the NLMIXED Procedure.”

Department of Mathematics, Burapha University, Saen Suk, Thailand, April 26-28, 2017.

- Bimal Sinha, “Some Aspects of Data Analysis under Confidentiality Protection,” and “Likelihood-Based Finite Sample Inference for Singly Imputed Synthetic Data under Multivariate Normal and Multiple Linear Regression Models.”

Office of Planning, Research and Evaluation, U.S. Department of Health and Human Services, Washington, D.C., May 24, 2017.

- Tommy Wright, “No Calculation When Observation Can Be Made.”

Contemporary Theory and Practice of Survey Sampling: A Celebration of Research Contributions of J.N.K. Rao, Kunming, China, May 24-27, 2017.

- Gauri Datta, “Robust HB Methods for Small Area Estimation.”
- Carolina Franco, “Measurement Error in Small Area Estimation: Functional vs. Structural Models.”

University of California, San Diego, San Diego, California, June 7, 2017.

- Tucker McElroy, “Testing Collinearity of Vector Time Series.”

Statistical Society of Canada Annual Meeting, Winnipeg, Manitoba, Canada, June 11-14, 2017.

- Gauri Datta, “Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model.”

International Chinese Statistical Association Applied Symposium 2017, Chicago, Illinois, June 25-28, 2017.

- Kimberly Sellers, “Under-dispersion Models: Models that are ‘Under the Radar’”; “Statistical Analysis for Non-normal Data” (Invited Session).

International Symposium on Forecasting, Cairns, Australia, June 26-28, 2017.

- Tucker McElroy, “Multi-Step Ahead Forecasting and Signal Extraction of High Frequency Vector Time Series.”

Indian Statistical Institute, Kolkata, India, June 28, 2017.

- Gauri Datta, “Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model.”

Department of Statistics, University of Calcutta, Kolkata, India, June 30, 2017.

- Gauri Datta, “Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model.”

Small Area Estimation 61st International Statistical Institute’s World Statistics Congress Satellite Meeting (2017), Paris, France, July 10-12, 2017.

- Robert Ashmead and Eric Slud, “Small Area Model Diagnostics and Validation with Applications to the U.S. Voting Rights Act Section 203.”
- Ryan Janicki, “Beta Regression Models for Small Area Estimation of Proportions.”
- Jerry Maples, “Improving Small Area Estimates of Disability: Combining the American Community Survey with the Survey of Income and Program Participation.”

Maastricht University School of Business and Economics, Maastricht, Netherlands, July 13, 2017.

- Gauri Datta, “Multivariate Fay-Herriot Bayesian Estimation of Small Area Means under Functional Measurement Error.”

61st International Statistical Institute's World Statistics Congress (2017), Marrakech, Morocco, July 16-21, 2017.

- Darcy Morris, "A COM-Poisson Mixed Model with Normal Random Effects for Clusters Count Data."
- Tapan Nayak, "A Novel Unbiased Post-randomization Method for Guaranteed Control of Identification Risk in Microdata Release."
- Kimberly Sellers, "Bivariate Conway-Maxwell-Poisson Distributions: Formulation, Properties, and Inference."
- Yves Thibaudeau, "Maximum Likelihood Parametrization for Conditional Probabilities."
- Tommy Wright, "Simple Exact Optimal Sample Allocation Algorithms."

2017 Consumer Expenditure Survey Methods Symposium, Washington, D.C., July 18, 2017.

- Robert Ashmead, "Reducing Respondent Contact Burden in the ACS Using a Cumulative Burden Score."

Veterans Administration, State of Illinois, Illinois, July 21, 2017.

- Bimal Sinha, "Some Aspects of Data Analysis under Confidentiality Protection."

Joint Statistical Meetings, American Statistical Association, Baltimore, Maryland, July 30-August 3, 2017.

- Robert Ashmead and Eric Slud, "Small Area Model Diagnostics and Validation with Applications to the Voting Rights Act, Section 203."
- Laura Bechtel, Nicole Czaplicki, Maria Garcia, and Jeremy Knutson, "Resolving Balance Complex Discrepancies in the Presences of Negative Data."
- William Bell, Gauri Datta, Carolina Franco, and Hee Chung, "Measurement Error in Small Area Estimation: Functional vs. Structure vs. Naïve Models."
- Gauri Datta, "Robust Hierarchical Bayes Small Area Estimation."
- Diag Davenport, Kimberly Sellers, Darcy Morris, and Li Zhu, "Conway-Maxwell-Poisson Process Implementation in R."
- Carolina Franco, William Bell, Serena Arima, Bruneo Liseo, and Gauri Datta, "Multivariate Fay-Herriot Bayesian Estimation of Small Area Means under Functional Measurement Error."
- Martin Klein and Gauri Datta, "Statistical Disclosure Control via Sufficiency under the Multiple Linear Regression Model."
- Joanna Lineback, Martin Klein, and Joseph Schafer, "Exploring New Estimation Techniques for the Monthly Wholesale Trade Survey."
- James Livsey and James Wilson, "Autoregressive Regime Switching Models for Dynamic Networks."
- Jerry Maples and Adam Maidman, "Small Area Models for Over-Dispersed Poisson Counts."
- Tucker McElroy, "Seasonal Adjustment Subject to Accounting Constraints."
- Brian Monsell and Tucker McElroy, "Issues Related to the Modeling and Adjustment of High Frequency Time Series."
- Mary Mulry, Nancy Bates, and Matt Virgile, "Lifestyle Segments, Social Marketing and Hard-to-Survey Populations: Understanding Participation in the 2015 Census Test."
- Osbert Pang, Brian Monsell, and William Bell, "Comparing Mean Squared Errors in X-13ARIMA-SEATS Using Components Models."
- Brandon Park and Tucker McElroy, "Implicit Multi-Layer Network for Time Series Data."
- Andrew Raim, Scott Holan, Jonathan Bradley, and Christopher Wikle, "An R Package for Spatio-Temporal Change of Support."
- Anindya Roy and Tucker McElroy, "Bayesian Estimation of Optimal Differencing Operator in Cointegrated Systems."
- Kimberly Sellers and Andrew Raim, "Introducing a Flexible Zero-Inflated Count Model to Address Data Dispersion."
- Eric Slud and Robert Ashmead, "Hybrid BRR and Parametric-Bootstrap Variance Estimates for Small Domains in Large Surveys."
- Daniel Weinberg, Tucker McElroy, and Soumendra Lahiri, "Estimation of Locally Stationary Spatial Processes with Applications to the American Community Survey."
- Chuanhua Xing and Andrew Raim, "A Bayesian Integrative Model for Deciphering High-Dimensional Genotype-Phenotype Map."

Department of Statistics and Acturial Science, University of Hong Kong, Hong Kong, August 21-23, 2017.

- Bimal Sinha, "Likelihood-Based Finite Sample Inference for Singly Imputed Synthetic Data under Multivariate Normal and Multiple Linear Regression Models."

Department of Mathematics, Mahidol University, Salaya, Thailand, August 24-25, 2017

- Bimal Sinha, “Likelihood-Based Finite Sample Inference for Singly Imputed Synthetic Data under Multivariate Normal and Multiple Linear Regression Models.”

Mathematics and Statistics Department Colloquium, University of Maryland, Baltimore County, Baltimore County, Maryland, September 8, 2017.

- Tommy Wright, “No Calculation When Observation Can Be Made.”

Advanced Statistical Programming with Rcpp Workshop, University of Maryland, Baltimore County, Baltimore County, Maryland. September 22, 2017.

- Andrew Raim and Iris Gauran, “Advanced Statistical Programming with Rcpp.”

5. CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY SEMINAR SERIES

Gauri Datta, CSRM, U.S. Census Bureau & University of Georgia, “Robust Methods in Small Area Estimation,” January 18, 2017.

Joshua Day, North Carolina State University, “Julia for High Performance Technical Computing,” March 21, 2017.

Matthew Simpson, University of Missouri, “Introduction to Stan for Markov Chain Monte Carlo,” April 25, 2017.

Dongchu Sun, University of Missouri-Columbia, *SUMMER AT CENSUS*, “Estimation and Prediction in the Presence of Spatial Confounding for Spatial Linear Models,” May 9, 2017.

Tommy Wright, CSRM, U.S. Census Bureau, “No Calculation When Observation Can Be Made,” May 23, 2017.

Martin Slawski, George Mason University, “On the Use of Random Projections for Dimension Reduction in Linear Regression,” May 31, 2017.

Robert Schumacher, GfK, *SUMMER AT CENSUS*, “The Role of Human-Centered Design in Survey Instruments and Beyond,” May 31, 2017.

Victoria Stodden, University of Illinois at Urbana-Champaign, *SUMMER AT CENSUS*, “Enhancing Reproducibility for Computational Methods,” June 5, 2017.

Toshihiko Mukoyama, University of Virginia, *SUMMER AT CENSUS*, “Barriers to Reallocation and Economic Growth: The Effects of Firing Costs,” June 5, 2017.

Andrew Penner, University of California, Irvine, *SUMMER AT CENSUS*, “The Causal Effects of Course Failure,” June 6, 2017.

Jonathan Azose, Pacific Northwest National Laboratory, *SUMMER AT CENSUS*, “Probabilistic Population Projections With Migration Uncertainty,” June 6, 2017.

Thurston Domina, University of North Carolina at Chapel Hill, *SUMMER AT CENSUS*, “Civil Society Goes to School: A Contextual Analysis of Parent Teacher Associations,” June 8, 2017.

Daniel Feenberg, National Bureau of Economic Research (NBER), *SUMMER AT CENSUS*, “An Introduction to TAXSIM,” June 19, 2017.

Caroline Weber, University of Oregon, *SUMMER AT CENSUS*, “Estimating the Elasticity of Broad Income for High-Income Taxpayers,” June 19, 2017.

Jon Bakija, Williams College, *SUMMER AT CENSUS*, “A Comprehensive Historical U.S. Federal and State Income Tax Calculator Program,” June 20, 2017.

Richard D. De Veaux, Williams College, *SUMMER AT CENSUS*, “The Seven Deadly Sins of Big Data,” June 27, 2017.

Guoyi Zhang, University of New Mexico, *SUMMER AT CENSUS*, “Neyman Smooth-Type Goodness of Fit Tests in Survey Data,” June 27, 2017.

Richard D. De Veaux, Williams College, *SUMMER AT CENSUS*, “Aging and Human Performance – What Are the Limits?,” June 28, 2017.

Hang Kim, University of Cincinnati, *SUMMER AT CENSUS*, “Statistically Integrated Data-Processing for 2017 Economic Census Microdata,” June 28, 2017.

Richard D. De Veaux, Williams College, *SUMMER AT CENSUS*, “Presentation of the Data Science Guidelines for Undergraduate Programs,” June 29, 2017.

Aloysius Siow, University of Toronto, *SUMMER AT CENSUS*, “A Quality View of Earnings Inequality,” July 10, 2017.

Brad N. Greenwood, University of Minnesota, *SUMMER AT CENSUS*, “Uber Might Buy Me a Mercedes Benz: An Empirical Investigation of the Sharing Economy and Durable Goods Purchase,” July 10, 2017.

Bikas K. Sinha, Retired Professor of Statistics, Indian Statistical Institute, Kolkata, *SUMMER AT CENSUS*, “Unbiased Estimation of a Finite Population Proportion in the Light of Randomized Reporting,” July 11, 2017.

Bikas K. Sinha, Retired Professor of Statistics, Indian Statistical Institute, Kolkata, *SUMMER AT CENSUS*, “Under-Estimation Problem – Further Thoughts,” July 12, 2017.

Donald P. Green, Columbia University, *SUMMER AT CENSUS*, “Field Experimental Designs for the Study of Media Effects,” July 12, 2017.

Xiaofeng Shao, University of Illinois at Urbana-Champaign, *SUMMER AT CENSUS*, “Martingale Difference Divergence and Its Applications to Contemporary Statistics,” July 18, 2017.

Hilal Atasoy, Temple University, *SUMMER AT CENSUS*, “The Impact of Health Information Exchange Use on Chronic Disease Management: An Empirical Investigation,” July 19, 2017.

Vishesh Karwa, Harvard University (Postdoc), The Ohio State University, “Differential Privacy and Statistical Inference: A Statistician’s Perspective,” July 25, 2017.

Mehdi Bozorgmehr, City University of New York, *SUMMER AT CENSUS*, “Middle Eastern American Panethnicity,” July 25, 2017.

Marie Haldorson, Statistics Sweden, *SUMMER AT CENSUS*, “GEOSTAT 2 – A Point-Based Foundation for Statistics: A Model for a Point-Based Geocoding Infrastructure for Statistics Based on a Geocoded Address, Building, and Dwelling Register,” July 27, 2017.

Dan Hammer, University of California, Berkeley, *SUMMER AT CENSUS*, “Change Detection and Validation Metrics for Relevant Use of Commercial Satellite Imagery,” August 1, 2017.

Rahul Mazumder, Massachusetts Institute of Technology, “Sparse Statistical Learning with Discrete Optimization,” August 3, 2017.

Duncan Elliott, Office of National Statistics (UK), *SUMMER AT CENSUS*, “A Comparison of New and Established Benchmarking Methods,” August 7, 2017.

John Aston, University of Cambridge, *SUMMER AT CENSUS*, “Wavelet Benchmarking and Seasonal Adjustment – Some Theory and Some Results,” August 7, 2017.

Sezgin Ayabakan, Temple University, *SUMMER AT CENSUS*, “Reevaluating Readmission Reduction Policies: The Role of Telehealth and Latent Patient Health Status,” August 8, 2017.

Christopher Clapp, Florida State University, *SUMMER AT CENSUS*, “Interactions of Public Paratransit and Vocational Rehabilitation,” August 15, 2017.

Michael Mueller-Smith, University of Michigan, *SUMMER AT CENSUS*, “Avoiding Convictions: Regression Discontinuity Evidence on Court Deferrals for First-Time Drug Offenders,” August 15, 2017.

Thuan Nguyen, Oregon Health & Science University, *SUMMER AT CENSUS*, “Fence Methods for Small Area Estimation: Recent Development and Implementation,” August 16, 2017.

George Alter, University of Michigan, *SUMMER AT CENSUS*, “Continuous Capture of Metadata for Statistical Data,” August 29, 2017.

Kirsten Early, (U.S. Census Bureau Dissertation Fellow), Carnegie Mellon University, “Dynamic Question Ordering: Obtaining Useful Information While Reducing User Burden,” August 30, 2017.

Pingfang Zhu, Shanghai Academy of Social Sciences, *SUMMER AT CENSUS*, “R&D Spill-Over Effects: Evidence from Chinese Industrial Sectors,” September 18, 2017.

6. PERSONNEL ITEMS

6.1 HONORS/AWARDS/SPECIAL RECOGNITION

Director's Award for Innovation

- **Darcy Steeg Morris** (Team Award)—"For investigating, developing, and implementing methods to use administrative records and third-party data to reduce the number of in-person field visits in the 2020 Census during its Nonresponse Follow-Up operation. This application is projected to lower costs in the census field operations by around $1.4 billion, while maintaining the Census Bureau's high standard of quality in its data."

Fellow, American Statistical Association

- **Anindya Roy**—"For major contributions to biostatistics, time series analysis, and statistical methodology for federal agencies, especially in Bayesian analysis, environmental risk assessment, and biomarker studies; for outstanding mentoring of graduate and undergraduate students; and for sustained service to the profession.

6.2 SIGNIFICANT SERVICE TO PROFESSION

Robert Ashmead

- Organizer, 2017 JSM Session: "New Developments in Small Area Estimation Research at the U.S. Census Bureau."
- Refereed a paper for the *Journal of Official Statistics*

Emanuel Ben-David

- Refereed papers for *Statistica Sinica, Mathematical Reviews,* and *The American Statistician*

Gauri Datta

- Associate Editor, *Sankhya*
- Associate Editor, *Statistical Methods and Applications*
- Assistant Editor, *Calcutta Statistical Association Bulletin*
- Associate Editor, *Environmental and Ecological Statistics*
- Refereed papers for *Journal of the Royal Statistical Society-A, TEST, Canadian Journal of Statistics* and *Scandinavian Journal of Statistics*

Carolina Franco

- Refereed papers for the *Journal for the Royal Statistical Society-Series A*, *Survey Methodology,* and *Computational Statistics and Data Analysis*
- Organizer, invited session on small area estimation at the 2017 International Indian Statistical Association Conference in Hyderabad, India

Maria Garcia

- Member, Program Committee, 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference
- Panelist, Women in Statistics and Data Science Conference, Charlotte, North Carolina, October 2016

Ryan Janicki

- Refereed papers for *Journal of the American Statistical Association*, *The American Statistician*, and *Biometrika*

Patrick Joyce

- Refereed a paper for *Biometrika*

Martin Klein

- Member, Co-advisor, and Dissertation Reader, Ph.D. Dissertation in Statistics Committee, University of Maryland, Baltimore County
- Refereed papers for *The American Statistician* and *Journal of Official Statistics*
- Chair, Special Alumni Session at the 11th Annual Probability & Statistics Day, University of Maryland, Baltimore County

James Livsey

- Organizer, 2017 JSM Session: "Time Series, Reconciliation, and National Statistics."
- Organizer and Chair, 2017 JSM Session: "Advances in Time Series Methodology."
- Refereed papers for *Communications in Statistics-Theory and Methods* and *Sankhya B, The Indian Journal of Statistics*

Xiaoyun Lu

- Refereed a paper for *Discussions Mathematicae Graph Theory*

Thomas Mathew

- Associate Editor, *Journal of the American Statistical Association*
- Associate Editor, *Sankhya*
- Editorial Board member, *Journal of Occupational and Environmental Hygiene*
- Co-editor for a special issue of the *Journal of Statistical Theory and Practice*
- Refereed papers for *The American Statistician, Statistica Neerlandica, Biometrics,* and *Statistics in Medicine*
- Member, American Statistical Association's Committee on W.J. Youden Award in Inter-laboratory Testing

Tucker McElroy

- Refereed papers for *Annals of Statistics, Computational Statistics and Data Analysis, Electronic Journal of Statistics, Journal of Time Series Analysis, Statistica Sinica, Econometric Reviews, Journal of Applied Econometrics, Journal of Official Statistics, Extremes, Sankhya, Statistics Sinica,* and *Journal of the American Statistical Association*
- Chair, 2017 JSM Session, "Recent Developments in Seasonal Adjustment"
- Reviewer, grants for Chilean National Science and Technology Commission and National Science Foundation

Darcy Steeg Morris

- Refereed papers for *Journal of Official Statistics, Statistica Neerlandica*, and *Hacettepe Journal of Mathematics and Statistics*
- Judge, Science Fair for STEM Students, TC Williams High School, Alexandria, Virginia

Brian Monsell

- Organizer, First Seasonal Adjustment Practitioner's Workshop, November 4, 2016

Mary H. Mulry

- Associate Editor, *Journal of Official Statistics*
- Methodology co-Editor, *Statistical Journal of the International Association of Official Statistics*

Tapan Nayak

- Associate Editor, *Communications in Statistics - Theory and Methods*
- Associate Editor, *Communications in Statistics - Simulation and Computation*
- Associate Editor, *Journal of Statistical Theory and Practice*
- Refereed papers for *The American Statistician, Journal of Official Statistics,* and *International Statistical Review*

Andrew Raim

- Reviewed papers for *Biometrical Journal, Hacettepe Journal of Mathematics and Statistics,* and *Mathematical Population Studies*
- Member, Statistics Ph.D. Committee, University of Maryland, Baltimore County

Anindya Roy

- Associate Editor, *Sankhya A*
- Associate Editor, *Sankhya B*
- Member, American Statistical Association's committee on Retention and Recruitment
- Reviewer, grant proposals for Research Foundation Flanders, Belgium

Kimberly Sellers

- Chairperson, American Statistical Association Committee on Women in Statistics
- Associate Editor, *The American Statistician*

- Associate Editor, *Journal of Computational and Graphical Statistics*
- International Black Doctoral Network Association, Incorporated -- Advisory Board member; Director, BDN STEMers
- Member, Adele's Circle of Women, University of Maryland, College Park, MD
- Scientific Program Committee member, International Conference on Statistical Distributions and Applications (ICOSDA) 2016
- Refereed papers for *Communications in Statistics – Theory and Methods,* and *Statistics and Probability Letters*

Bimal Sinha

- Chair, Doctoral Dissertation Committee, University of Maryland, Baltimore County

Eric Slud

- Associate Editor, *Biometrika*
- Associate Editor, *Journal of Survey Statistics and Methodology*
- Associate Editor, *Lifetime Data Analysis*
- Member, Hansen Lecture Committee, Washington Statistical Society (completed term of service)
- Refereed papers for *Journal of Official Statistics* and *Journal of Survey Statistics and Methodology*
- Reviewer, Report for National Academy of Sciences

Yves Thibaudeau

- Refereed papers for *The Journal of Official Statistics* and *Communication in Probability –Simulation*.

Thomas Trimbur

- Refereed papers for *The American Statistician*, *Journal of Applied Econometrics, Journal of Business and Economics Statistics, Journal of Forecasting,* and *Journal of Official Statistics*.

William Winkler

- Associate Editor, *Journal of Privacy and Confidentiality* and *Transactions on Data Privacy*
- Member, Statistics Ph.D. Committee, University of Maryland, College Park
- Member, Program Committee of the IEEE Workshop on Data Integration and Applications 2017 and of Statistical Data Protection 2017
- Refereed papers for *Statistics in Medicine* and *Journal of Official Statistics*

Tommy Wright

- Associate Editor, *The American Statistician*
- Member (Past Chair), Waksberg Award Committee, *Survey Methodology*
- Member, Board of Trustees, National Institute of Statistical Sciences

6.3 PERSONNEL NOTES

Bret Hanlon left the Census Bureau.

Bimal Sinha (Statistics Faculty at University of Maryland, Baltimore County) was reassigned to our Center on a Schedule A appointment.

Tapan Nayak (Statistics Faculty at The George Washington University) was reassigned to our Center on a Schedule A appointment.

Soumendra Lahiri (Statistics Faculty at North Carolina State University) joined our Center on a Schedule A appointment.

Thomas Trimbur joined our Time Series Research Group in a permanent position.

Michael Leibert joined the Postal Regulatory Commission.

Erica Magruder joined Tower Radiology (LaPlata, MD).

APPENDIX A **Center for Statistical Research and Methodology FY 2017**

Program Sponsored Projects/Subprojects With Substantial Activity and Progress and Sponsor Feedback

(Basis for PERFORMANCE MEASURES)

Project #	Project/Subproject Sponsor(s)	CSRM Contact	Sponsor Contact
	DECENNIAL		
6650C23	Redesigning Field Operations		
6750C01	Administrative Records Data		
6550C01	Data Coding, Editing, and Imputation		
6250C02	Operational Design		
	1. Decennial Record Linkage	William Winkler	Tom Mule
	2. Coverage Measurement Research	Jerry Maples	Tim Kennel
	3. Record Linkage Error-Rate Estimation Methods	William Winkler	Tom Mule
	4. Supplementing and Supporting Non-Response with Administrative Records	Michael Ikeda	Tom Mule
	5. Identifying "Good" Administrative Records for 2020 Census NRFU Curtailment Targeting	Darcy Steeg Morris	Tom Mule
	6. 2020 Census Communications Campaign Statistical Analyses	Mary Mulry	Gina Waljeko
	7. Undercount of Young Children	Mary Mulry	Scott Konicki
	8. 2020 Census Privacy Research	Robert Ashmead	John Abowd
	9. Project to Study Priority of Tracts for Outreach and Advertising in Decennial Census 2020	Eric Slud	Nancy Bates
	Address Canvassing In Field		
	10. Statistical Evaluation of the In-office Image Review Process	Andrew Raim	April Avnayim
6350C02	*11. Development of Block Data Tracking Database*	Tom Petkunas	Graham Baggett
	American Community Survey (ACS)		
	12. ACS Applications for Time Series Methods	Tucker McElroy	Mark Asiala
6385C70	*13. Confidence Intervals for Proportions in ACS Data*	Carolina Franco	Mark Asiala
	14. Voting Rights Section in 203 Model Evaluation and Enhancements Towards Mid-Decadal Determinations	Robert Ashmead	James Whitehorne
	15. ACS Income Modeling	Eric Slud	John Eltinge
	DEMOGRAPHIC		
TBA	Demographic Statistical Methods Division Special Projects		
	16. Research Balanced Repeated Replication and Other Variance Estimation Techniques for Use with Current Population Survey	Eric Slud	Yang Cheng
0906/1444X00	Demographic Surveys Division (DSD) Special Projects		
	17. Data Integration	Ned Porter	Christopher Boniface
TBA	Population Division Special Projects		
	18. Introductory Sampling Workshop	Tommy Wright	Oliver Fischer
7165017	Social, Economic, and Housing Statistics Division Small Area Estimation Projects		
	19. Research for Small Area Income and Poverty Estimates (SAIPE)	Carolina Franco	Wes Basel
	20. Small Area Health Insurance Estimates (SAHIE)	Ryan Janicki	Wes Basel
	21. Sub-County Estimates of Poverty from Multi-year ACS Data	Jerry Maples	Wes Basel
	ECONOMIC		
1183X01	Economic Statistical Collection		
	22. Research on Imputation Methodology for the Monthly Wholesale Trade Survey	Martin Klein	Joe Schafer
	23. Use of Big Data for Retail Sales	Darcy Steeg Morris	Rebecca Hutchinson
2270C10	Economic Census/Survey Engineering: Time Series Research; Economic Missing Data/Product Line Data; Development/SAS		
	24. Seasonal Adjustment Support	Brian Monsell	Kathleen McDonald-Johnson
	25. Seasonal Adjustment Software Development and Evaluation	Brian Monsell	Kathleen McDonald-Johnson
	26. Research on Seasonal Time Series: Modeling & Adjustment Issues	Tucker McElroy	Kathleen McDonald-Johnson
	27. Supporting Documentation & Software: X-13ARIMA-SEATS	Brian Monsell	Kathleen McDonald-Johnson
TBA	Investigation of Alternative Methods for Resolving Balance Complex Failures in StEPS		
	28. Investigation of Alternative Methods for Resolving Balance Complex Failures in StEPS	Maria Garcia	Laura Bechtel
	CENSUS BUREAU		
7225010	*29. National Cancer Institute Tobacco Use Survey/CPS*	Isaac Dompreh	Benmei Liu

APPENDIX B

FY 2017 PROJECT PERFORMANCE MEASUREMENT QUESTIONNAIRE

CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY

Dear

In a continuing effort to obtain and document feedback from program area sponsors of our projects or subprojects, the Center for Statistical Research and Methodology will attempt for the nineteenth year to provide *seven measures of performance,* this time for the fiscal year 2017. For FY 2017, the *measures of performance* for our center are:

Measure 1. Overall, Work Met Expectations: Percent of FY 2017 Program Sponsored Projects/Subprojects where sponsors reported that work met their expectations.

Measure 2. Established Major Deadlines Met: Percent of FY 2017 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met.

Measure 3a. At Least One Improved Method, Developed Technique, Solution, or New Insight: Percent of FY 2017 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight.

Measure 3b. Plans for Implementation: Of the FY 2017 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight, the percent with plans for implementation.

Measure 4. Predict Cost Efficiencies: Number of FY 2017 Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency."

Measure 5. Journal Articles, Publications: Number of journal articles (peer review) and publications documenting research that appeared or were accepted in FY 2017.

Measure 6. Proceedings Publications: Number of proceedings publications documenting research that appeared in FY 2017.

These measures will be based on response to the five questions on this form from our sponsors as well as from members of our center and will be used to help improve our efforts.

To construct these seven measures for our center, we will combine the information for all of our program area sponsored projects or subprojects obtained during November 14 thru November 21, 2017 using this questionnaire. Your feedback is requested for:

Project Number and Name: ______________
Sponsoring Division(s): ________________

After all information has been provided, the CSRM Contact ____________ will ensure that the signatures are obtained in the order indicated on the last page of this questionnaire. We very much appreciate your assistance in this undertaking.

____

Tommy Wright Date
Chief, Center for Statistical Research and Methodology

Brief Project Description ***(CSRM Contact will provide from Division's Quarterly Report):***

Brief Description of Results/Products from FY 2017 ***(CSRM Contact will provide)****:*

(over)

TIMELINESS:
Established Major Deadlines/Schedules Met

1(a). Were all established major deadlines associated with this project or subproject met? **(Sponsor Contact)**

□ Yes □ No □ No Established Major Deadlines

1(b). If the response to 1(a) is No, please suggest how future schedules can be better maintained for this project or subproject. **(Sponsor Contact)**

QUALITY & PRODUCTIVITY/RELEVANCY:
Improved Methods / Developed Techniques / Solutions / New Insights

2. Listed below are at most 2 of the top improved methods, developed techniques, solutions, or new insights offered or applied on this project or subproject in FY 2017 where an CSRM staff member was a significant contributor. Review "a" and "b" below **(provided by CSRM Contact)** and make any additions or deletions as necessary. For each, please indicate whether or not there are plans for implementation. If there are no plans for implementation, please comment.

□ No improved methods/techniques/solutions/new insights developed or applied.

□ Yes as listed below. (See a and b.)

	Plans for Implementation?
a. ____________________________	Yes □ No □
____________________________	
____________________________	
____________________________	
____________________________	
b. ____________________________	Yes □ No □
____________________________	
____________________________	
____________________________	
____________________________	

Comments (Sponsor Contact):

COST:
Predict Cost Efficiencies

3. Listed **(provided by CSRM Contact)** below are at most two research results or products produced for this project or subproject in FY 2017 that predict cost efficiencies. Review the list, and make any additions or deletions as necessary. Add any comments.

□ No cost efficiencies predicted.
□ Yes as listed below. (See a and b.)

a.

b.

Comments (Sponsor Contact):

OVERALL:
Expectations Met/Improving Future Communications

4. Overall, work on this project or subproject by CSRM staff during FY 2017 met expectations. **(Sponsor Contact)**

□ Strongly Agree
□ Agree
□ Disagree
□ Strongly Disagree

5. Please provide suggestions for future improved communications or any area needing attention on this project or subproject. **(Sponsor Contact)**

(CSRM Contact will coordinate the signatures as noted and pass to CSRM Chief.)

First____
Sponsor Contact Signature Date

Second____
CSRM Contact Signature Date

Center for Statistical Research and Methodology

Research & Methodology Directorate

STATISTICAL COMPUTING AREA

Bill Winkler (Acting)
- VACANT

Machine Learning & Computational Statistics Research

Bill Winkler
- Emanuel Ben-David
- Xiaoyun Lu

Missing Data Methods Research

Yves Thibaudeau
- Maria Garcia
- Darcy Morris
- Jun Shao (U. of WI)

Research Computing Systems & Applications

Chad Russell
- Tom Petkunas
- Ned Porter

Simulation, Modeling, & Data Visualization Research

Martin Klein
- Isaac Dompreh
- Brett Moran
- Bimal Sinha (UMBC)
- Nathan Yau (FLOWINGDATA.COM)

MATHEMATICAL STATISTICS AREA

Eric Slud
- VACANT

Sampling & Estimation Research

Eric Slud (Acting)
- Robert Ashmead
- Mike Ikeda
- Patrick Joyce
- Mary Mulry
- Tapan Nayak (GWU)

Small Area Estimation Research

Jerry Maples
- Gauri Datta (U. of GA)
- Carolina Franco
- Ryan Janicki

Time Series Research

Brian Monsell
- Osbert Pang

Tucker McElroy
- Soumendra Lahiri (NCSU)
- James Livsey
- Aninyda Roy (UMBC)
- Thomas Trimbur
- Dan Weinberg

Experimentation & Modeling Research

Tommy Wright (Acting)
- Thomas Mathew (UMBC)
- Andrew Raim
- Kimberly Sellers (Georgetown U.)

Tommy Wright, Chief
- Kelly Taylor
- Lauren Emanuel
- Jae-Kwang Kim (F)
- Michael Hawkins
- VACANT

(F) ASA/NSF/Census Research Fellow

September 30, 2017