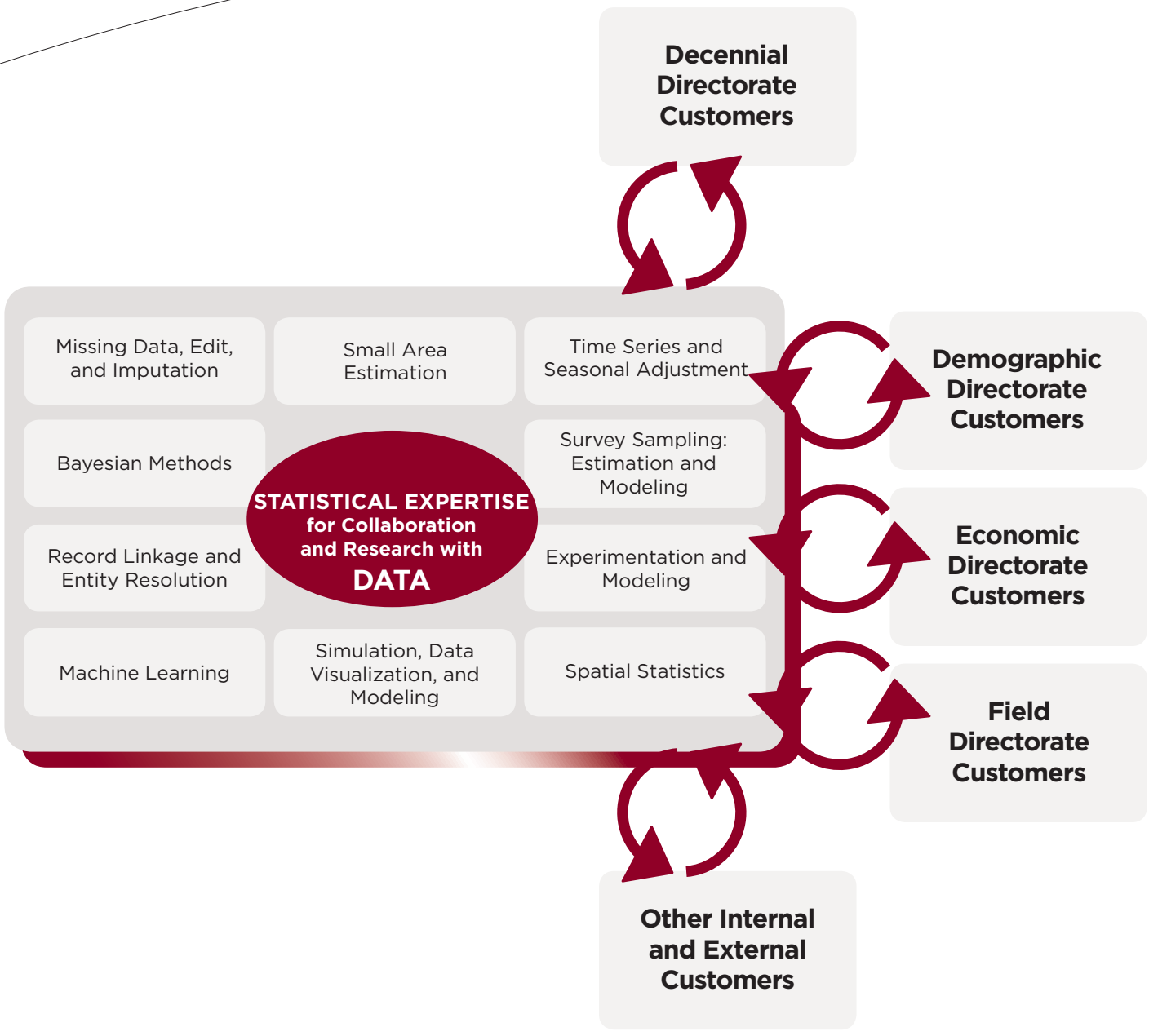
Annual Report of the Center for Statistical Research and Methodology

Research and Methodology Directorate

Fiscal Year 2018



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

We help the Census Bureau improve its processes and products. For fiscal year 2018, this report is an accounting of our work and our results.

Center for Statistical Research & Methodology
https://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 2018 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 coupled with complementary logistic regression to improve the efficiency (reducing the sampling error) of estimates of gross flows and estimate of gross flows proportions from month to month, classified by demographic variables; (2) Investigated the feasibility of using third party ("big") data from First Data – a large payment processor to supplement or enhance retail sales estimates in the Monthly/Annual Retail Trade Surveys (MRTS and ARTS).

- *Record Linkage:* (1) Began record linkage methodology study group to increase knowledge among staff; (2) Conducted research to combine regression with record linkage, such that the error from the linkage process propagates exactly into the resulting regression.

- *Small Area Estimation*: (1) Developed a Multinomial-Dirichlet to model school district pieces population counts from county population totals and compared results against a Generalized Poisson small area model; (2) Obtained several new results, both theoretical and empirical, related to the understanding of functional and structural measurement error Fay-Herriot models, as well as of naïve models that ignore measurement errors.

- *Survey Sampling-Estimation and Modeling*: (1) Demonstrated the potential for a market segmentation from an external source to improve self-response propensity models using data from the 2010 Census and the American Community Survey; (2) Contributed to team development of methods for producing differentially private decennial census tabulations conforming to legally mandated error-free disclosure of block-level population totals under *Public Law 94* as well as *Title 13* requirements for nondisclosure of individual-level data.

- *Time Series and Seasonal Adjustment*: (1) Developed new algorithms for midcasting (imputing missing values, with uncertainty) and signal extraction of multivariate time series; (2) Developed new autoregressive diagnostics for seasonality.

- *Experimentation and Statistical Modeling:* (1) Studied finite-sample bias in the fiducial distribution of some functions of multinomial probabilities; (2) Started work on a paper for the spatio-temporal change of support modeling package (stcos), featuring a small case study using ACS data.

- *Simulation and Statistical Modeling:* (1) Developed new methodology using principles of multiple imputation to analyze data under a differentially private Laplace mechanism; (2) Developed new methodology for using sample survey estimates to construct a joint confidence region for a ranking of populations.

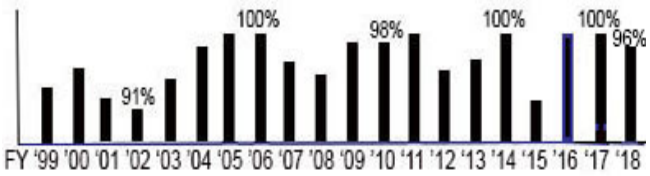
- *SUMMER AT CENSUS:* Sponsored, with divisions around the Census Bureau, scholarly, short-term visits by 42 researchers/leaders who collaborated extensively with us and presented seminars on their research. For a list of the 2018 *SUMMER AT CENSUS* scholars, see *http://www.census.gov/research/summer_at_census/*.

How Did We[1] Do...

For the 20th year, we received feedback from our sponsors. Near the end of fiscal year 2018, our efforts on 27 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 27 questionnaires were obtained with the following results (The graph associated with each measure shows the performance measure over the last 20 fiscal years):

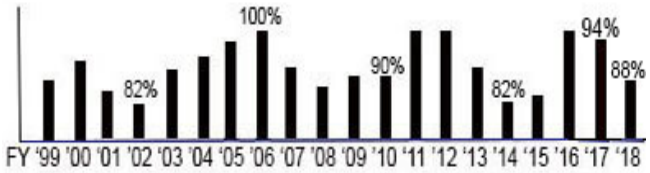
Measure 1. *Overall, Work Met Expectations*

Percent of FY2018 Program Sponsored Projects/Subprojects where sponsors reported that overall work met their expectations (agree or strongly agree) (26 out of 27) ……..…….... 96%



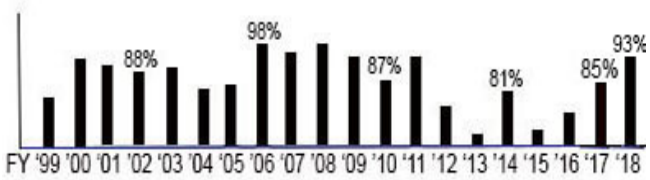
Measure 2. *Established Major Deadlines Met*

Percent of FY2018 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met (14 out of 16 responses) …….....……..…..….. 88%



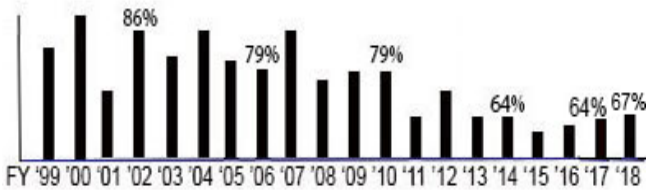
Measure 3a. *At Least One Improved Method, Developed Technique , Solution, or New Insight*

Percent of FY2018 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight (25 out of 27 responses) …… 93%



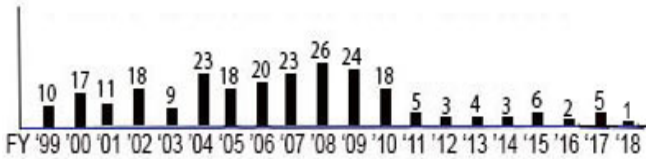
Measure 3b. *Plans for Implementation*

Of these FY2018 Program Sponsored Projects/Subprojects reporting at least one improved method, technique developed, solution, or new insight, the percent with plans for implementation (16 out of 24 responses) …….. 67%



Measure 4. *Predict Cost Efficiencies*

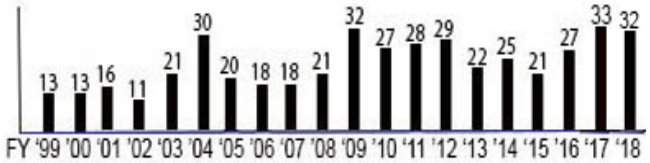
Number of FY2018 Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency" ……..…… 1



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

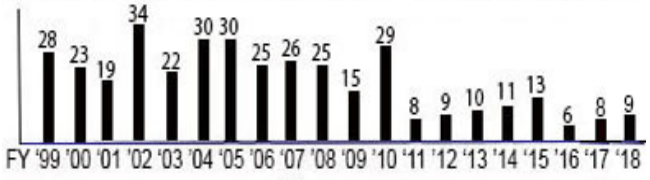
Measure 5. *Journal Articles, Publications*

Number of peer reviewed journal publications documenting research that appeared (19) or were accepted (13) in FY2018 ……... 32



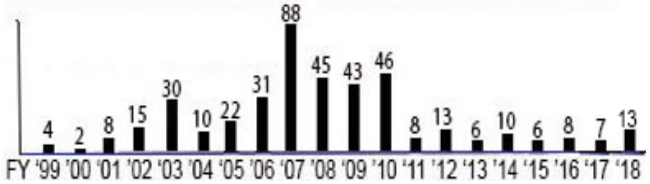
Measure 6. *Proceedings, Publications*

Number of proceedings publications documenting research that appeared in FY2018 ……..……. 9



Measure 7. *Center Research Reports/Studies, Publications*

Number of center research reports/studies publications documenting research that appeared in FY2018 …… 13



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 OPERATIONAL DESIGN (DECENNIAL Project 6250D02)

1.2 DATA CODING/EDITING/IMPUTATION (Decennial Project 6550D01)

1.3 CCM PLANNING AND PROJECT MANAGEMENT (Decennial Project 6650D01)

1.4 2020 EVALUATIONS-PLANNING AND PROJECT MANAGEMENT (Decennial Project 6650D20)

1.5 ADMINISTRATIVE RECORDS DATA (Decennial Project 6750D01)

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: This is on hold due to changed requirements. During FY 2018, staff began a group study of Census Bureau record linkage software (BIGMATCH, SRD Matcher, PVS System).

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

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 2018, staff attended weekly meetings for the Coverage Measurement team and Estimation subteam. Staff assisted team members from the Decennial Statistical Studies Division in implementation and interpretation of the nonresponse interview adjustment study plan and the status imputation study plan. Staff also provided consulting to DSSD staff to assist them on their research tasks in nonresponse modeling and status imputation.

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

C. 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: During FY 2018, we began some research by reviewing additional literature. When there are massive training sets, this is known as the Regression Problem (Vapnik 1995, Hastie, Thibshirani, Friedman 2009) that is considered the most difficult machine learning problem. We will need to develop methods for the unsupervised situation.

We continue to anticipate the delivery of empirical data.

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

D. 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 2018, staff continued to compare results using four additional weeks of IRS 1040 returns (tax year 2009 received in weeks 4-21 of 2010) to results without the four additional weeks of returns for directly modeling the Census household count. The two sets of results are generally similar. Staff extended the comparison to look at the effect of using cutoffs for the maximum estimated probability in determining when to use the household count based on administrative records (AR). The results using weeks 4-17 and the results using weeks 4-21 generally produce similar predicted values for corresponding IDs and similar population estimates for corresponding cutoffs. Staff summarized the results of the cutoff analysis in an overview that was sent to the Administrative Records Modeling subteam. Staff also compared the results of modeling using the planned 2020 production AR modeling methodology with and without state data from specific programs targeted towards low-income households. The state data included Supplemental Nutrition Assistance Program data from eleven states, Temporary Assistance for Needy Families data from three of these eleven states, and Low Income Energy Assistance Program data from a twelfth state. The models with and without state data are the same except that the models with state data include variables for presence of state data and presence in state data elsewhere and the AR household for the models with state data can include persons only in state data. There are three sets of models: AR occupied, AR vacant, and AR delete. In all three sets, the model with state data generally produced similar results to the model without

state data. There are some differences between the units added as AR occupied units by the state data models and the units dropped as AR occupied units by the state data models and there were two states with an unusually high proportion of units dropped by the state data models (compared to the proportion of units added). Staff summarized the results of the comparison of modeling with and without state data in a draft document (co-authored with Darcy Morris). The document has been distributed to the Administrative Records Modeling subteam. Results from the draft document will be an input into the decision on whether to use state data in modeling for AR removals from nonresponse follow-up.

Staff: Michael Ikeda (x31756)

E. 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 2018, Center for Statistical Research and Methodology (CSRM) and Decennial Statistical Studies Division (DSSD) staff completed revisions and published the paper "A Distance Method for Administrative Record Modeling in the 2020 Census" in 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 and 2018 Census Tests. Relatedly, staff presented "A Modeling Approach for Administrative Record Enumeration in the Decennial Census" at the DC-AAPOR POQ Special Issue Conference. CSRM staff concluded preliminary analysis using real estate (MLS) data from CoreLogic to assess the relationship between housing unit vacancy and listing status and presented results to the Census Scientific Advisory Committee working group on administrative records. CSRM staff also attended meetings of 2018 end-to-end test updates and assisted in the analysis of state program administrative data (SNAP data) for potential use in identifying and enumerating occupied housing units.

Staff: Darcy Steeg Morris (x33989), Yves Thibaudeau

F. 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 2018, 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. Currently staff is investigating the potential for using robust regression in the analyses. The data was 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 (682-305-8809)

G. 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 2018, staff acquired Protected Identification Key (PIK) codes for 2010 Census and Census Coverage Measurement (CCM) and cleaned data to prepare for matching.

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

H. 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 2018, staff continued to collaborate with staff in the Research and Methodology Directorate to explore the use of a lifestyle segmentation of the U.S. population to gain insight about variation in self-response in censuses and surveys and determine whether the segments could be useful in the 2020 Census Communications Campaign. The use of the Tapestry segmentation is innovative in that it is designed for commercial marketing and not commonly used in survey research or census taking. The initial analyses used merged data from three sources: (1) self-response results from 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. Recent analyses have included fitting models to predict propensity for self-response in the 2010 Census and in the American Community Survey with the Tapestry segments and the LRS as independent variables. The new models show that adding lifestyle segments to the LRS improves the self-response propensity models, Staff produced a revised manuscript that includes the new results and submitted it to a journal. Staff also prepared a presentation entitled "Using Lifestyle Segments to Enhance Response Propensity and Design Nonresponse Interventions," that incorporated recent work, and presented it at the International Total Survey Conference in Durham, NC in June.

Staff: Mary Mulry (682-305-8809)

I. 2020 Census Communications Campaign Statistical Modeling

Description: As part of the broader effort under the Integrated Partnership and Communications Research and Analytics Team, a team of CSRM staff was assembled to create Tract-level predictive models of self-response to the ACS and Decennial Census, both in Mail and Internet modes, in terms of data on the Planning Data Base (PDB), ACS and FCC data on internet usage, and ESRI (commercial) "Tapestry" data files concerning marketing and lifestyle segmentation of the tract-level population. The goal is to improve on previous predictive scores (the Low Response Score developed in research of Erdman and Bates, 2013) and preliminary research undertaken by a contractor, with particular emphasis on trying Statistical Machine Learning techniques.

Highlights: During FY 2018, a Team was organized and conducted a series of exploratory data analyses on the combined datasets from the PDB, from ACS tract-level aggregated variables on internet usage, and from ESRI 'Tapestry' data, augmented also by a new Urban/Rural tract-level variable aggregated from Geography Division block-level datasets. The methods explored included: variable selection techniques for linear models using the dataset variables and interactions; decision-tree and random forest techniques resulting in new thresholded variables that we also then included in linear model variable-selection analyses; support vector machine analyses; and LASSO variable selection techniques. Members of the Team then met with the Y&R contractors and explained the best predictor variables we had found for modeling tract-level overall and internet ACS self-response rates. Our activities on this project concluded with that meeting; with transmission of the newly created tract-level Urban/Rural data-files to the ongoing Communications Modeling Team (including Y&R); and with critical review a few weeks later of the Contractor's proposed modeling and analysis plan. This project was terminated during the second quarter of FY 2018.

Staff: Eric Slud (x34991), Emanuel Ben-David, Kimberly Sellers, Darcy Morris, Gina Walejko (DSSD)

J. Census Coverage Error of Young Children

Description: As part of the preparations for the 2020 U.S. Census, the U.S. Census Bureau has conducted research to gain insight about the causes of the long-standing undercount of young children in decennial censuses and in Census Bureau sample surveys. This project explores the relationship between the positive responses to child-specific probes on the 2010 Census questionnaire and 2010 Census Coverage Measurement (CCM) weighted nonmatching children 0 to 4 years of age, which are an indicator of enumeration problems, including being missed by the census or in the census but not with enough information to be uniquely identified. The coverage probes for young children and other household members appeared for the first time on the 2010 Census questionnaire and therefore, comparable data are not available from previous censuses. The 2010 Coverage Follow-up (CFU) was a telephone interview and was unable to contact all the households that gave positive responses to the coverage probes. The results of these analyses also may provide insight for census and survey fieldwork and in the census coverage measurement program.

Highlights: During FY 2018, staff used regression modeling to explore whether a relationship exists between hard-to-count status as indicated by the Low Response Score (LRS) of an area and the rate of positive

responses to child-specific probes on the 2010 Census questionnaire in the area. This relationship is of interest because the mean LRS is available nationwide prior to the 2020 Census, and therefore, could be helpful in planning the field operations and the 2020 Census Communications Campaign and future censuses and surveys. The coverage probes for young children and other household members appeared for the first time on the 2010 Census questionnaire so analyses of comparable data from previous censuses are not available.

The results of the regression modeling indicate that for segments in 12 of the 14 LifeModes, the mean LRS in a segment was predictive of the number of positive responses to child specific probes per thousand households for segments. This result may be interpreted as indicating that the segments that tend to have low response rates in censuses and surveys also tend to be the areas where a high percentage of respondents were uncertain about whether to include one or more young children in their household census population count. However, there appears to be a subpopulation composed of the segments in two LifeModes where the rates of positive responses to the child-specific coverage probes are uncorrelated with the mean LRS, and the observed rates tend to be lower than would be expected for the rest of the segments with comparable values of their mean LRS. These two LifeModes are two of the youngest and have the fewest number of children of all the LifeModes.

Staff: Mary Mulry (682-305-8809)

K. 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, creating visualization techniques, and engineering algorithms to accomplish this task.

Highlights: During FY 2018, staff made major strides in implementing and refining the Topdown algorithm to release results from the 2020 Decennial Census. The algorithm code is stored in a git repository and summarized in an unpublished paper "TopDown94: Large-scale Differentially Private Publication of Query Answers and Microdata". Staff also finished the first draft of a paper titled "Effective Privacy after Adjusting for Constraints, with applications to the 2020 Census" which is intended to discuss the implications of including invariants in algorithm based on differential privacy. Staff met regularly with stakeholders and internal subject matter experts to discuss experimental results and the progress of the project.

Staff: Robert Ashmead (x31564), Brett Moran, Michael Ikeda, Ned Porter, Philip Leclerc (CDAR), Simson Garkinkel (CDAR), William Sexton (CDAR), John Abowd (ADRM), Daniel Kifer (CDAR)

L. 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 2018, lists and maps of reweighted high-priority tracts were run and submitted to Regional Office Field staff for feedback. As a result, Field decided that the reweighting did not provide enough useful new information to be developed generally as a tool for outreach efforts in the 2020 Census, and this project concluded early in the second quarter of FY 2018.

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

M. Creation of Record-Level Public-Use File from the Census Barriers, Attitudes, and Motivators Survey Confidential Micro-data

Description: Between February 20 and April 17, 2018, the Census Bureau conducted a national stratified random sample survey of households with the title Census Barriers, Attitudes, and Motivators Survey (CBAMS) to assess the nation's reactions to various questions relating to participation in the 2020 Census. The approximately17,500 responses to CBAMS reside in a confidential data file which will help the Census Bureau plan strategies for maximizing participation in the 2020 Census.

It was desired to grant access to a protected record-level file created from the confidential data file to a contractor who is a member of the 2020 Communications Campaign Team so that it could help identify effective communications strategies for the 2020 Census.

Highlights: During FY 2018, staff members were provided specific and detailed instructions for producing ten disclosure protected files for potential public use corresponding to the values $\varepsilon = 1, 2, \ldots, 9, 25$ of a privacy loss budget under a local definition of ε-differential privacy. For each of these disclosure protected files, a measure of accuracy based on mean square error was computed. For each variable (most were discrete) on each record, noise was added using "multinomial randomized response" for the discrete variables and Laplace additive randomized noise for the continuous variables. The desire was to add noise while preserving distributions of the variables. Results and details are documented in an

unpublished report, "Implementation of Multinomial Randomized Response and Laplace Additive Noise to Create a Record-Level Public-Use File from the Census Barriers, Attitudes, and Motivators Survey Confidential Micro-data," DPST Assignment 1: July 20, 2018.

Staff: Tommy Wright (x31702), Rolando Rodriguez (CDAR), Martin Klein

1.6 ADDRESS CANVASSING IN FIELD (Decennial Project 6350D02)

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 2018, staff carried out regression analysis with non-image covariates, with the response taken to be whether in-office assignment of a block to a "passive" status was consistent with fieldwork's assessment of the block. An internal write-up of the analysis was given to staff from the Geography Division.

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 2018, the In-Office Address Canvassing Interactive Review reached the goal of completing 100% of all Census tabulation blocks in the U.S. and continues to update housing counts in the MAF. As updated lists and images become available, the next stage of Address Canvassing includes triggering tabulation blocks for further review.

Staff: Tom Petkunas (x33216)

C. Statistical Inference on Call Volumes to Helplines
Description: The Census Bureau will provide telephone helplines to assist respondents for the 2020 Census. Experiments are being conducted to study how the schedule of census mailings affects both response rates and volumes of calls to the helplines. An ideal strategy would lead to a constant volume of calls throughout the week without a negative impact to response rates. A flat distribution of calls is easier to staff than one where the majority of calls tend to occur on one or two days.

Highlights: During FY 2018, staff proposed hypothesis tests and confidence intervals to compare call volumes in a statistical framework. The procedures consider entropy between discrete probability distributions based on observed multinomial data. Staff completed a technical report discussing the methodology and detailed results. Results were presented to the Decennial Research Objectives and Methods (DROM) working group, and were used to complete a 2020 Research and Testing Report on mailing strategy.

Staff: Andrew Raim (x37894), Thomas Mathew, Elizabeth Nichols (CSM)

1.7 AMERICAN COMMUNITY SURVEY (ACS) (Decennial Project 6385D70)

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 2018, staff made extensions to a kriging framework based on continuous-time ARMA models, for application to American Community Survey Office monthly data, in order to generate customized pooled estimates. Issues of parametrization and numerical optimization were explored, and simulation studies supporting the modeling efforts were completed. Some preliminary data analysis was also done.

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: The *Journal of Official Statistics* paper titled "Adaptive Intervention Methodology for Reduction of Respondent Contact Burden in the American Community Survey" written by staff was published in FY 2018. This project is now completed.

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 Project 0331000).

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

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 2021 Determinations
Description: Section 203 of the *Voting Rights Act* mandates the Census Bureau to make estimates every five years relating to totals and proportions 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 result in the legally enforceable requirement that certain geographic political subdivisions must provide language assistance during elections for groups of citizens who are unable to speak or understand English adequately enough to participate in the electoral process. The research undertaken in this project consists of the development, assessment and estimation of regression-based small area models based on 5-year American Community Survey (ACS) data and the Decennial Census.

Highlights: During FY 2018, staff continued to document the small area methods, published estimates, and model assessments as part of the 2016 models. In addition, staff discussed possible modelling approaches for the 2021 determinations. The 2021 models can make use of the 2020 Decennial Census tabulations, in contrast with the 2016 models, where the 2010 Decennial Census tabulations where judged to be too far out of date. Staff considered how proposed changes to the 2020 Decennial Census would affect the modeling approach to the 2021 determinations.

Staff finished the initial draft of the report documenting the 2016 models and sent it to internal reviewers. Staff discussed technical issues such as sampling error of covariates and replicate weight variance estimation for small domains.

Staff [Eric Slud] initiated research into the validation by simulation of the hybrid variance-estimation methodology used in the 2016 *Voting Rights Act* project, and that this research was incorporated into the preparation of a presentation by Staff at the SAE 2018 Small Area Estimation conference in Shanghai, China.

Staff completed and revised the technical report in response to reviewer comments, and continued design and coding of a simulation study to validate the hybrid variance-estimation methodology.

Staff: Robert Ashmead (x31564), Carolina Franco, Eric Slud, Mark Asiala (DSSD)

F. Model-based Estimates to Improve Data Confidentiality for ACS Special Tabulations
Description: ACS special tabulations are custom data releases requested by external customers. The released tables, which are often based on small sample sizes, raise concerns with data privacy and confidentiality. This project is to create model based estimates of the special tabs.

Highlights: During FY 2018, staff started working on modeling a set of ACS special tabulations on young children for the Minnesota Department of Education. The purpose is to provide model-based estimates to help lessen the potential for disclosure of data. The staff is working on spatially dependent models using the framework developed by Bradley, Holan and Wikle (2018). Preliminary model fits have been obtained and staff is in the process of refining and improving the model to better fit the data.

Staff: Jerry Maples (x32873), Andrew Raim, Ryan Janicki, Scott Holan (ADRM), Tommy Wright, John Eltinge (ADRM), John Abowd (ADRM)

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

A. Research on 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 2018, new CPS datasets were produced and analyzed in relation to research of CSRM staff on methods of weighted and unweighted analysis, using models with and without random effects, on longitudinal survey data.

On the longitudinal project, research continued in which gross-flow cell estimates with and without weighting and random effects on CPS data for two successive months' data have been analyzed and prepared for a conference presentation and a book chapter by CSRM staff. Novel elements of the research on this topic concerned the extension of (unweighted) mixed-effect model estimation methodology applicable to CPS employment-related gross-flows, and of BRR variance estimation methodology to unweighted mixed-model predictions of small cell probabilities.

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

1.9 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: During FY 2018, staff cleaned data and provided nine tables for the 2020 Decennial Disclosure review project. Staff archived software on the Census GitHub site (github.ti.census.gov) and documented on the project wiki. Staff also provide data cleaning and support for related work. Additional staff provided record linkage, data cleaning and documentation for the 2010 Data reconstruction and reidentificaiton project for the 2020 Decennial Disclosure. The documentation included summary metadata and a collaborative report series.

Staff: Ned Porter (x31798)

1.10 SOCIAL, ECONOMIC, AND HOUSING STATISTICS DIVISION SMALL AREA ESTIMATION PROJECTS (Demographic Project 7165018)

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 2018, staff continued the work begun in FY2017 comparing different types of measurement error in the univariate case with application to school-aged children in poverty. Staff obtained several new empirical results that further elucidate the impact of ignoring or incorrectly specifying measurement error for SAIPE. Staff wrote a paper about this work, submitted it for publication and completed a revision. The paper was accepted to *Survey Methodology*. Staff gave several related professional talks. Staff also developed and implemented a method to infuse noise on the figures from the above paper that involved SAIPE data, in order to protect the confidentiality of the data, and had the resulting plots evaluated and approved by the Disclosure Review Board.

Staff also began developing multiple model frameworks to estimate the number of children and number of children in poverty for school districts. The first method is a small area generalized Poisson model which takes into account the overdispersion of the sampling distribution of the sample survey estimates. The second framework uses a square root transformation of the counts which is motivated by the variance stabilizing transformation. The final method is a small area share model which models the relative proportion of the counts in the school district to the county. Staff also compared models for county-level school-aged children in poverty rates by testing three different models on the artificial population of counties which was previously created for the purpose of SAIPE model comparison and evaluation. The models considered were a univariate version of the Binomial Logit Normal (BLN) Model proposed in Franco and Bell (2013, 2015), a Fay-Herriot (FH) model on log transformed poverty counts similar to the SAIPE production model, and a FH model on poverty rates. The true sampling variances and

design effects were used in order to separate the problem of estimation of sampling variances from that of selecting the best model form. The models will also be fit with estimated sampling variances in future work. Preliminary results suggest the BLN achieves lower MSEs for the majority of counties.

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.

Small Area Estimation of Proportions
Highlights: In FY 2018, staff continued work on small area estimation of proportions, with a focus on analysis of data sets with estimates clustered near the boundary and which include a point mass at zero. A comparison of different methods was performed, to investigate differences in point estimates and uncertainty estimates by analyzing 2014 SAHIE data. The models considered were the Fay-Herriot, transformed Fay-Herriot, zero inflated beta regression, and a censored model. Analysis of two different SAHIE data sets, one with a small and one with a moderate number of exact zeros, using different model assumptions was completed. A comparison of predictions and interval estimates using each model for the small areas with direct estimates of zero was made.

Review and comparison of unit-level small area modeling strategies
The goal of this project is to perform a comprehensive literature review of unit-level small area modeling strategies, and to evaluate these strategies by fitting different unit-level models on simulated data and on ACS data, with a focus on estimation of proportions with health insurance. Work in Q2 of FY2018 includes beginning a literature review, constructing a data set for model fitting, and writing an outline of a paper for comparing unit level small area models.

Highlights: In FY 2018, staff conducted a comprehensive literature review of unit-level small area modeling strategies when the sampling design is informative, with a goal of investigating which methods can be applied to estimation of health insurance coverage at the county level. Four model-based methods, which make use of the sample survey design and sampling weights, were chosen for this application, and an evaluation of these strategies was started through fitting these models on simulated data and on ACS data. The initial data analysis was restricted to a single state, and the modeling strategies included regression of response variables on survey weights, regression of survey weights on responses, and pseudo-likelihood methods. A comparison of predictions and uncertainty estimates using each method with each other and with unweighted methods was made.

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: In FY 2018, staff created two frameworks to modeling the number of children in school districts. The first method is a small area model based on the generalized Poisson (Consul and Jain 1973) with a gamma distributed random effect. Staff had to solve the problem of specifying the overdispersion parameter from estimates of the sampling variance from the source survey data. Staff developed a generalized variance function (GVF) to give more stable estimates of the overdispersion. This modeling framework was also extended to estimating the number of school-age children in poverty in school districts. Portions of this research was presented at the Small Area Estimation Conference in Shanghai, China in June 2018.

The second framework jointly modeled the school district piece to county population share. The model was based on the Dirichlet-Multinomial (DM) distribution. This model form had some unique differences than other applications of the DM model in that the number of areas could be different in each county, i.e. each county may contain a different number of school district pieces, and the ordering of the school district pieces was not relevant. This model was applied to school district data from the state of Nebraska and the results were presented at the Joint Statistical Meetings in Vancouver, Canada in August 2018.

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

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: In FY 2018, staff continued development of a realistic, artificial population that can be used to repeatedly draw simulated Monthly Wholesale Trade Survey (MWTS) data. The population is created as a tool for evaluating the performance of statistical methodology applied to the MWTS or other similar surveys. Staff developed a procedure to adjust the proportion of observed zero values for certain variables and to make the proportion more realistic. This procedure resulted in an updated version of the population. Based on evaluations of this version of the population, staff made further refinements. Using the latest version of the constructed population, staff conducted simulation studies to evaluate current estimation and imputation procedures for the MWTS.

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

B. Use of Big Data for Retail Sales Estimates

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: In FY 2018, a new round of data analysis began with updated First Data purchase transaction data from Palantir. Staff worked on exploratory data analysis of this new monthly retail sales data based on raw aggregates, specifically investigating suppression rates and measures of coverage. CSRM staff worked with statisticians and subject matter experts in the Economic Directorate to understand trends in data availability by industry, state and region in order to determine promising industries and geographies for producing experimental estimates. CSRM and ESMD staff finished and disseminated a report that developed industry-level "quality profiles" and described the methodology for producing geographically granular retail sales experimental estimates for select industries.

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

1.12 ECONOMIC CENSUS/SURVEY ENGINEERING: TIME SERIES RESEARCH; ECONOMIC MISSING DATA/PRODUCT LINE DATA; DEVELOPMENT/SAS (Economic Project 2270D10)

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 2018, staff provided seasonal adjustment and software support for users within and outside the Census Bureau, including Capital Economics, Barclay's Bank, Lloyds Banking Group, U. S. Bank, CEPREMAP (France), SAS, Citadel, Freddie Mac, Warner Brothers, Macrobond (Sweden), International Monetary Fund, Banco Central do Brazil, National Bank of Kazakhstan, Bundesbank, Federal Reserve Board, Bureau of Labor Statistics, Bureau of Economic Analysis, State Of Washington Economic & Revenue Forecast Council, Colorado Department of Labor, Australian Government Department of Jobs and Small Business, Ethiopian Development Research Institute, Statistics Center (UAE), Central Statistics Office (Ireland), INEI-Peru, Statistics Norway, Statistics Austria, Australian Government Department of Jobs and

Small Business, Cass Business School (UK), Concordia University (Canada).

Brian Monsell, working with Kathleen McDonald-Johnson and Demetra Lytras (ESMD), organized a seasonal adjustment workshop held at the Bureau of Labor Statistics Conference Center on April 28, 2018. Several staff members have participated in an inter-agency group to study and make recommendation on mitigating residual seasonality in estimates of GDP, which has evolved into a group to discuss research in seasonality diagnostics and residual seasonality.

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 for several quarters of publication. Staff also met with employees from the Economic Directorate to transfer the responsibilities of generating the seasonal adjustments of the Business Formation statistics from the Research Directorate to the Economic Directorate, with an overview presentation led by Mr. Monsell and subsequent shadowing sessions where employees from the Economic Directorate would work on BFS series.

Staff: Brian Monsell (x31721), Tucker McElroy (ADRM), 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 2018 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 2018, staff released an updated version of X-13ARIMA-SEATS (Version 1.1, Build 48), to the Economic Directorate for testing. Staff and compared adjustments from this version of the software to the last released version of X-13ARIMA-SEATS (Version 1.1, Build 39) and no difference in the seasonally adjusted series was found; this was confirmed by testing by groups in the Economic Directorate. Since this release, staff continued to changes in the X-13ARIMA-SEATS source code related to HTML table output, correcting defects in the SEATS code of X-13ARIMA-SEATS, flagging errors in the automatic model identification routine, adding an option to specify what series are used to generate the Hodrick-Prescott cycle, correcting defects in the code that saves the user defined regression model of X-13ARIMA-SEATS, and adding an optional test for seasonal overdifferencing to the automatic model identification procedure.

Staff continued the development of sigex, a suite of R routines for modeling multivariate time series, by improving the documentation and establishing a version on github. Improvements to the speed of multi-step forecasts and the simulation module were made.

Staff: Brian Monsell (x31721), Tucker McElroy (ADRM), Osbert Pang, James Livsey

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 2018, staff (a) continued study of the impact of weather regressors on seasonal adjustment; (b) continued research on modeling of daily time series (New Zealand immigration data and credit card

transaction data) with multiple periods of seasonality, utilizing ideal low-pass filters to separate trend from annual seasonality; (c) compared various diagnostics of seasonality, and continued research into issues of residual seasonality detection; (d) continued research into the use of weekly seasonal factors to improve monthly seasonal adjustment; and (e) investigated the application of seasonal vector form to understanding seasonality and using it to seasonally adjust time series; (f) continued simulations and empirical work for consultation project, involving daily time series data used to improve the calendarization of 4-4-5 data; (g) developed routines in R for a reconciliation project to solve an issue in temporal aggregation of monthly series that results in residual seasonality for quarterly series; (h) examine the performance of seasonal adjustment diagnostics on seasonally adjusted series.

Staff: Tucker McElroy (x33227; ADRM), James Livsey, Brian Monsell, Osbert Pang, Thomas Trimbur, William Bell (ADRM)

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 2018, 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 is developing materials documenting aspects of the source code of X-13ARIMA-SEATS and the documentation for the software and presenting X-13 code walks for interested employees. Staff participated in meetings concerning the modernization of CSRM's web presence and developed comments on converted web pages generated by IT Directorate staff; these pagers are intended to migrate CSRM to the Census Bureau's content management system. Staff shared utilities related to generating seasonal adjustment and model diagnostic summaries to colleagues in the Economic Directorate and at Statistics Canada.

Staff: Brian Monsell (x31721), Tucker McElroy (ADRM), 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 2018, we collaborated with Economic Statistical Methods staff on developing alternative methods to raking when items can have negative values. We developed, tested, evaluated, and delivered four separate nonlinear programs that minimize a loss function under specified balance linear constraints. We designed a weighting scheme in which the weight of an item represents the cost of adjusting the item when solving the nonlinear program. We applied the methodology to the QFR and provided assistance on one of the deliverable programs as the QFR migrated into StEPS II processing. This project is completed.

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

1.14 NATIONAL CANCER CENTER (Census Bureau Project 7225010)

A. National Cancer Center 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 2018, staff ran backward model selection procedure to choose covariates with the arcsine transformation method that were used to produce Hierarchical Bayesian (HB) estimates. Staff produced and prepared matched records for 3,134 counties considered for this project with TUS, CPS and ACS files.

Staff conducted the following (a) ran small area estimation Bayesian model to produce HB estimates for each of the five smoking outcomes; (b) ran model diagnostics to validate the HB estimates; (c) staff is currently working on benchmarking to validate the final county level HB estimates that will be benchmarked to the corresponding state level direct estimates. Staff is currently working on steps to fit random effects models to test state effects of the five smoking outcomes of interest.

Additionally, staff identified data problems and proposed solutions to address these important data problems on the TUS-CPS small area estimation project. Some of the data difficulties and important data issues that were identified include the following: (a) conducted rigorous data steps to remove duplicate records in order to produce accurate direct estimates and design effects for state and county levels; (b) recalculated design effects for the five outcomes (current smokers, ever smokers, work ban policies, home ban smoking policies, and attempt to quit smoking) and (c) conducted a comprehensive data steps to identify unique cases for national, state and county level design effects.

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), Joe Engmark, Michael Hawkins, 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: In FY 2018, the Integrated Research Environment (IRE) went into production. After addressing issues identified during user testing, all Research Data Center (RDC) projects were migrated to it. Planning for the migration of research1 and research2 projects to the IRE is underway and is scheduled to be complete in March 2019.

Also in FY 2018, the R&M Directorate launched a pilot project called the “Cloud Research Environment” (CRE). The goal of the project is to take four existing projects that do not use Title 26 data and migrate the work to systems in Amazon GovCloud and to evaluate the customer experience and cost of that versus working in IRE or other “on premise” systems. Requirements gathering has taken place and the project is now in the design phase with CSVD drafting the initial design documents.

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 2018, we completed development, testing, and evaluation of alternatives methods to raking that are feasible for negative data and solve nonlinear programs that minimize loss functions under specified constraints while ensuring balance edits are satisfied.

We continue collaboration in researching nonparametric Bayesian methods developed by Kim (2017) to edit, multiply impute, and produce synthetic industry-level Economic Census microdata. We resolved issues when applying this methodology to industries in the manufacturing and mining sectors. The editing problem is complex for these two sectors: there is a larger number of ratio edits, a larger number of variables, and several balance editing constraints. In these cases, the Bayesian editing and imputation method fails to converge. We used a connected graph representation of the ratio edits to analyze the variables and edits. We added and deleted variables to the model based on the connectivity represented in the ratio edits graph to find an initial set of variables for which the model converges. Given the model convergence for a set of core variables, we were able to test and resolve convergence issues for several industries in these two sectors when using ratio edits and balance edits for a larger number of variables.

We implemented an updated version of the software to edit and multiply impute test data for sample industries in all sectors. We used the software graphics utility to produce convergence plots for guidance in determining the total number of burn-in iterations for the MCMC step and the number of iterations between posterior draws.

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 2018, staff extended log-linear models to estimate the gross flows cells where there is no unit available from the data of interest. This is a common situation when dealing with fine-grain cross-classification with zero cells and estimatable conditional probabilities. Staff applied the methods to estimate gross-flows based on data fielded through the Current Population Survey and will present results in a forthcoming book chapter.

Staff continued to expand the model-based and model-assisted methodologies in longitudinal surveys to estimate variance through balanced repeated replication when some of the observation maybe missing (Thibaudeau 2013). Staff initiated efforts to identify best record-linkage solutions in the context of a virtual “administrative record census”. Staff has initiated a comparison between “BigMatch” and

"Unimatch" (the current record linkage process for PVS matching).

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 2018, we continue collaboration with an interdivisional team researching the implementation of nonparametric Bayesian methods developed by Kim et al. (2017) to edit, multiply impute, and synthetize Economic Census data. We proposed and calculated several measures to classify study industries at the industry and imputation cell level. We implemented the editing and imputation modules of Kim's software for seven separate economic census sectors datasets and tested with multiple industries. We resolved difficulties when applying the methodology to industries in the manufacturing and mining sectors. Data in these sectors must satisfy balance edit constraints in addition to ratio edit restrictions. In these cases, there are convergence issues during the editing and imputation steps. Additional research identified problems in the R, C++ programs compatibility that we resolved with simple programming fixes and convergence is no longer an issue. We developed and implemented quality metrics measures for selecting what type of data (raw data with limited corrections or data with analyst's corrections) to use in the next project steps. We developed quality metric measures based on correlations and regressions of core items and calculation of sample moments (mean, variance, skewness, and kurtosis). Sample moments' quality metrics are used, if needed, as a possible tiebreaker when formal hypothesis tests are not informative. We wrote a draft of a User Guide for the software.

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 2018, one staff person reviewed two long technical reports on privacy and confidentiality about the Decennial Short-Form. Staff refereed one paper for a monograph on privacy and confidentiality.

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 2018, staff reviewed extensive literature and refereed one paper for the *Annals of Applied Statistics* and one paper for *American Political Science Review*. Staff co-authored a paper with George Mason University faculty on using signal recovery techniques in an application of adjusting statistical analyses for linkage

error.

One staff person did additional review and research with two statistics professors at George Mason University. The three individuals are looking at improving both the theoretical ideas and the computational methods. Staff provided several large papers on methods and three large write-ups on record linkage for administrative lists for R&M leadership.

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 2018, staff provided a set of comments to two individuals at Statistics Canada about some new record linkage methods that they are investigating. Two staff persons provided a background document on record linkage.

Staff worked on some projects in collaboration with two statistics faculty members at George Mason University. These projects concern the problem of regression analysis with linked data files. In these projects, staff with university collaborators proposed to use robust regression with group LASSSO and pseudo-likelihood estimation with mixture models for estimating the regression parameters. Simulation studies demonstrate a significant improvement in estimation and computation speed-ups.

Staff worked on ongoing project on graphical models. In this project staff and university collaborators propose to use non-local priors for estimation and graph selection in the context of Gaussian directed acyclic graphical models.

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 2018, the R Group matured and expanded in several independent directions, including integration to an interagency task force. Staff continues to maintain an infrastructure for many Census Bureau users.

Staff: Yves Thibaudeau (x31706), Chad Russell

E. Entity Resolution and Merging Noisy Databases
Description: Work is underway on the problem of merging noisy databases to remove duplicate entities (individuals, households, etc.), where typically a unique identifier is not known. This problem in the literature is known as entity resolution or record linkage. Work is undertaken on improved methodology, and scalability, and testing such methods on both synthetic and real data.

Highlights: Two main results can be reported from this work so far. The first is an accepted paper with Brunero Liseo and Andrea Tancredi on combining the approach of regression with record linkage, such that the error from the linkage process propagates exactly into the resulting regression. This approach was tested on two datasets, which were both sample survey data, illustrating that we are able to correct for biases in the record linkage in the regression step.

The second piece of work is scaling prior work from Steorts (2015), which is available on CRAN and is known as blink. In a recently submitted paper with others, we propose the first scalable Bayesian method for entity resolution, which scales to 1/2 million records using a Bayesian latent variable model and using distributed computing. Specifically, we propose dblink a distributed extension to blink---a state-of-the-art model from statistics due to (Steorts, 2015)---by contributing several key ideas, including: (i)~a reparameterization of the model that induces conditional independencies, allowing groups of entities to be updated in parallel; (ii)~application of partially-collapsed Gibbs sampling; and (iii)~a novel perturbation sampling algorithm (leveraging the Vose-Alias method) that enables fast updates of the entity attributes. We demonstrate empirically that dblink can achieve efficiency gains of over 100 times when compared to a non-distributed Gibbs sampler baseline.

We are continuing to work on scaling dblink.

Staff: Rebecca C. Steorts (919-485-9415)

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 2018, staff completed two revisions of the paper submitted in FY 2017, which was subsequently accepted for publication. Staff added the Logit interval for computing confidence intervals for proportions to the simulation study, and analyzed the results through several tables and figures to determine whether the Logit interval is a viable alternative to the other intervals studied in the paper. Staff identified and corrected a coding error that affected limited simulation results. Staff did new analysis on the biases, variances, and mean squared errors of the design-based vs proposed model-based estimates of effective sample sizes. Staff created and tested user friendly code designed to be supplied to users who want to apply the proposed method of sampling variance estimation as well as the alternative methods of computing confidence intervals for proportions from complex sample surveys. Staff prepared and delivered an invited talk on this research in the 2018 Small Area Estimation Meeting in Shanghai, China.

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: In FY 2018, the peer reviewed paper for this project has been published in the in *Journal of the Royal Statistical Society – Series A*. Staff also gave an invited talk at the 50th Anniversary Celebration Conference (of the statistics department) at The Pennsylvania State University.

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: In FY 2018, staff illustrated the potential from borrowing strength from ACS to improve estimates from other surveys via bivariate models, specifically through using ACS estimates to improve estimates from the NHIS of health insurance coverage for U.S. States. Staff implemented and compared six different models, three bivariate and three univariate models, and showed that drastic improvements in terms of MSE can be achieved by using such bivariate models instead of the direct estimates. Decomposed the results into components involving pure shrinkage and actual gain from using the ACS estimates. Staff began writing the results in a paper to be submitted to a peer-reviewed journal, which will also include three additional applications.

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 2018, staff obtained several new results, both theoretical and empirical, related to the understanding of functional and structural measurement error Fay-Herriot models, as well as of naïve models that ignore measurement error. Staff created several contour plots that elucidate how incorrectly specifying measurement error when the structural measurement error Fay-Herriot model is true affects the resulting mean squared errors for different values of the relevant parameters. Staff also did empirical work that shows how the different predictors compare in their mean squared errors for the SAIPE application of school-aged children in poverty assuming the functional measurement error model is true. Staff developed and implemented a method to infuse plots using SAIPE data with noise, in order to protect the confidentiality of the data. Staff wrote a paper about this work and completed a revision, and the paper was accepted to *Survey Methodology*. Staff delivered several invited presentations at conferences about this work, as well as a seminar at Charles III University of Spain.

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

E. Robust Hierarchical Bayes Small Area Estimation for Nested Error Linear Regression Model

Description: Standard model-based small area estimates perform poorly in presence of outliers. Sinha and Rao (2009) developed robust frequentist predictors of small area means. In this article, we present a robust Bayesian method to handle outliers in unit-level data by extending the nested error regression model. We consider a finite mixture of normal distributions for the unit-level error to model outliers and produce noninformative Bayes predictors of small area means. Our solution generalizes the solution by Datta and Ghosh (1991, *Annals of Statistics*) under the normality assumption. Application of our method to a data set, which is suspected to contain an outlier, confirms this suspicion and correctly identifies the suspected outlier, and produces robust predictors and posterior standard deviations of the small area means. Evaluation of several procedures via simulations shows that our proposed procedure is as good as the other procedures in terms of bias, variability, and coverage

probability of confidence or credible intervals, when there are no outliers. In presence of outliers, while our method and Sinha-Rao method perform similarly, they improve over the other methods. This superior performance of our procedure shows its dual (Bayes and frequentist) dominance, and is attractive to all practitioners, Bayesians and frequentists, of small area estimation.

Highlights: During FY 2018, staff prepared a manuscript, submitted to a journal to review for publication, and submitted a revised manuscript. Staff created and delivered an invited presentation in the 2017 International Indian Statistical conference in Hyderabad, India.

Staff: Gauri Datta (x33426), Adrijo Chakraborty (NORC), Abhyuday Mandal (University of Georgia)

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: The *Journal of Official Statistics* published "Setting M-Estimation Parameters for Detection and Treatment of Influential Values," which describes methodology developed by staff in collaboration with staff in the Economic Directorate. The methodology provides 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. A side-by-side test of the methodology began in the Monthly Wholesale Trade Survey (MWTS) began in the summer of 2018. The test results will enable the evaluation of the effectiveness of the methodology when using it in a production setting.

Staff: Mary Mulry (682-305-8809)

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: Staff developed a simple and novel measure of uncertainty for an estimated ranking with theory, using American Community Survey travel time to work data, and with a visualization. Specifically, we construct a joint confidence region for the true unknown ranking of K populations based on an assumption that a real-valued parameter for each population is known to be between two real numbers while the value of each population's parameter is unknown. Two simulations confirm the theory. Staff worked on various visualizations of the joint confidence region of a true unknown ranking. Theory was published in a *CSRM Research Report* and submitted for publication.

Staff: Tommy Wright (x31702), Martin Klein, Jerzy Wieczorek (Colby College), Nathan Yau

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.

<u>Sample Allocation</u>
Highlights: During FY 2018, staff obtained access to data from the 2007 Economic Census and began to develop and apply algorithms to replicate (Panel) sample size determinations for various Economic sample surveys based on Neyman allocation and to compute sample sizes using the exact probability sampling methods in Wright (2017).

Staff: Tommy Wright (x31702), Andrew Raim

Apportionment

Highlights: Staff extended the current equal proportions methodology by appealing to probability sampling results in Wright (2017). Staff further extended the mathematical theory underlying the method used to apportion the seats in the U.S. House of Representatives.

Staff: Tommy Wright (x31702)

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: A formal report of the analysis and estimation documenting work on NCVS daily response propensity modeling and methodology was completed in FY2017.

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 2018, staff worked on the following projects: (a) Staff continued work on reconciliation for aggregates of time series, with extensions to the conversion of monthly to quarterly flow time series, such that seasonal adjustment adequacy is preserved. The methodology was updated, allowing for both the case of cross-series aggregation as well as aggregation across frequencies. (b) Staff continued work on signal extraction diagnostics methodology, including new diagnostics for seasonal adjustment based on autoregressive roots. Simulation studies and data analysis were completed. (c) Staff examined the Expectation-Maximization algorithm to assist with the fitting of multivariate time series models as well as the calculation of signal extraction estimates. The methods have been coded and tested on low-dimensional time series. (d) Staff continued the development of tools for understanding multivariate seasonal adjustment in the frequency domain. Spectral density estimation and diagnostics have been developed and implemented. (e) Staff continued writing code and text for a book on multivariate real-time seasonal adjustment and forecasting. These methods have been tested on different case studies. (f) Staff completed new algorithmic work on multivariate seasonal adjustment and missing value imputation. A new software product (Ecce Signum) has been developed to implement the missing value methods, allowing for a broad range of applications in forecasting and extreme-value adjustment. (g) Staff 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; ADRM), 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 2018, staff worked on the following projects:(a) Staff continued algorithmic work on computing Gaussian orthant probabilities, with applications to multivariate analysis. A new algorithm was tested for sorting through the extensive combinatorial calculations. (b) Staff completed simulation work for Frobenius norm methodology for fitting and comparing multivariate time series models in frequency domain. The finished project gives a new distribution theory for a broad range of statistics for vector time series. (c) Staff completed implementation of vector band pass filters, utilizing canonical trends and cycles to get adequate modeling of series. One facet is a paper on vector low pass filters, and a separate facet is improving the class of cycle component models. (d) Staff completed methodology and software for fitting high-dimensional VAR, and made some further extensions to allow for fitting of low-dimensional VAR with missing values. (e) Staff completed simulation work and project on information rigidities, developing a new State Space algorithm to handle cases where the dimension of data observed is time-varying. This framework allows for the modeling of the behavior of forecasters in the presence of crises. Also, new results were obtained on multi-step ahead forecast uncertainties. (f) Staff continued theoretical and applied work on nonlinear prediction for time series forecasting, using Hermite polynomials. A facet of this work involves using autocumulants to construct optimal quadratic predictors. (g) Staff obtained new results for computing and estimating multivariate inverse autocovariances, which is useful for fitting vector moving averages. (h) Staff continued research into methods for count time series.

Staff: Tucker McElroy (x33227; ADRM), 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 that contain data over- or under-dispersion (i.e. count data where the variance is larger or smaller than the mean, respectively). This model contains Poisson and negative binomial AR models as special cases.

Highlights: Staff continue to edit and revise a manuscript on the statistical methodology and results from this work to submit for publication. The manuscript was submitted for review with a journal, and is now being revised per the reviewer comments.

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: Experiments embedded within an ongoing survey are performed in order to test the effect of alternative treatments or survey methodologies on estimates of finite population parameters. This ongoing project will explore rigorous analysis of embedded experiments: from simple idealized designs to complex designs used in practice at the Census Bureau. Earlier work on this topic include a series of articles by Fienberg, Tanur, van den Brakel, and co-authors. They have developed the analysis of embedded experiments, mostly using a model-based approach, especially in the context of higher-order factorial experiments. A design-based analysis is yet to be taken up, especially for ratio and regression estimation (analysis of covariance), post-stratification, categorical data analysis, etc., in the context of multi-factor experiments embedded within complex surveys.

Highlights: No significant progress during FY 2018.

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

B. 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: The related manuscript was published in *Communications in Statistics – Theory and Methods*.

Staff: Kimberly Sellers (x39808), Darcy Steeg Morris

C. Developing Flexible Distributions and Statistical Modeling for Count Data Containing Dispersion
Description: Projects address myriad issues surrounding count data that do not conform to data equi-dispersion (i.e. where the (conditional) variance and mean equal). These projects utilize the Conway-Maxwell-Poisson (CMP) distribution and related distributions, and are applicable to numerous Census interests that involve count variables.

Highlights: During FY 2018, (1) staff drafted a manuscript regarding developing a broader zero-inflated version of a CMP regression model that includes zero-inflated negative binomial regression and zero-inflated binomial regression as special cases. Staff submitted the manuscript for review with a journal. (2) Staff are deriving the form of a multivariate CMP distribution to serve as a tool to describe variation and association for several count variables that express over- or under-dispersion (relationships where the variance of the data is larger or smaller than the mean, respectively). This work will generalize the previous results achieved developing a bivariate CMP model. (3) Staff derived moments associated with the bivariate CMP model in order to establish a sampling distribution for the components that comprise the bivariate Fisher index for dispersion (denoted FI_2). (4) Staff are deriving a generalization of the negative binomial distribution extended via the CMP distribution, and have identified three special cases of this new Conway-Maxwell-negative-binomial (COMNB) distribution.

Staff: Kimberly Sellers (x39808), Darcy Steeg Morris

D. 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 data and to the general R user community.

Highlights: During FY 2018, staff carried out simulations to exercise the R package and further investigate some of the results observed in their 2017 JSM Proceedings paper, which were based on the package. Staff began writing an article demonstrating the software and spatio-temporal change of support analysis, to be submitted for peer-reviewed publication.

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

E. 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 Conway-Maxwell-Poisson (CMP) regression model *incorporating* random effects for analysis of longitudinal count data.

Highlights: In FY 2018, staff developed R functions to incorporate normal- and conjugate-distributed random effects in a clustered/longitudinal COM-Poisson model. Staff investigated the behavior of the COM-Poisson conjugate distribution and implemented a simulation study for understanding the behavior and sensitivity of assumptions of the COM-Poisson random effect model. Staff continued working on a manuscript and presented this work at the "Conference on Multivariate Count Analysis" at the University of Franche-Comte in France. Staff also assisted external academic researchers implement previously

developed SAS code to applied problems in fields such as urban planning.

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.

Subproject A.1. (Nayak): Protecting the confidentiality of data collected from individuals, households and other survey units is a solemn obligation of data agencies, including the U.S. Census Bureau. One major concern is the potential for identifying the records of targeted survey units in the released data by matching the values of some variables, which can be obtained easily from other sources. Staff worked on developing rigorous methods for controlling identification risk. Staff published a paper titled "*Measuring Identification Risk in Microdata Release and Its Control by Post-randomisation*" in *International Statistical Review.* This paper presents a novel approach to measuring the risk of identifying a survey unit in released data by matching the values of some key variables. The measure is then used to set a strict disclosure control goals, which asks to ensure that the probability of correctly identifying any respondent or surveyed unit would not exceed a pre-specified value. This paper also develops an unbiased

post-randomisation procedure, which achieves this goal without losing much data utility. Staff continued research on this topic and issued a technical report titled "*Post-randomization for Identification Risk Limited Microdata Release from General Surveys*," which presents another method for controlling identification risk in general surveys, where survey weights are used to make statistical inferences.

Staff worked on developing statistical theory and methods for protecting respondent's privacy when collecting data by surveys. A general idea for privacy protection is to record a randomized version of the true response. However, randomized response theory lacks rigorous privacy measures and comparison of various randomization procedures. Staff developed a rigorous privacy criterion in terms of gain of information about a respondent from his/her randomized response. Staff identified all privacy satisfying randomization procedures and compared them broadly and characterized all admissible procedures. These and some related results are reported in a technical report titled "A Criterion for Privacy Protection in Data Collection and Its Attainment via Randomized Response Procedures." Staff has also obtained substantial additional results. In particular, staff has derived certain optimal randomization procedures, namely, privacy satisfying minimax randomization designs. Staff will soon report these results in a technical report.

Subproject A.2. (Klein): In FY 2018, staff continued work on finite sample statistical analysis of synthetic data under a standard linear regression model. Staff revised a paper based on this work and the paper was accepted for publication in the *Calcutta Statistical Association Bulletin*. Staff also continued work on finite sample statistical analysis of synthetic data under a multivariate linear regression model. Staff studied properties of the proposed methodology in scenarios where some of the conditions that were used to derive the methodology do not hold. Specifically, the following scenarios were studied: (1) the scenario where the error term in the multivariate regression model is not normally distributed; (2) the scenario where the imputer who creates the synthetic data and/or the data analyst underspecify or overspecify the regression model; and (3) the scenario where the data analyst's regression model is something other than the regression of the sensitive variables on the non-sensitive variables. Under each of these three scenarios conclusions were drawn regarding the robustness of the proposed methods.

In FY 2018, staff developed theory and methodology for data analysis under differential privacy using multiple imputation. Because differentially private mechanisms usually require the original data to have bounded support, staff proposed a modified version of the standard Laplace mechanism that does not require the support of the original data to be bounded. The modified Laplace mechanism is proven to satisfy differential privacy. Staff proposed methodology where the modified Laplace mechanism is used to protect the data, and then multiply imputed versions of the original data are constructed using the protected data as input. This method allows data users to obtain valid inference under differential privacy using simple multiple imputation combining formulas. Theory and methods are developed to support the proposal. This work is currently ongoing and a manuscript based on this work is under preparation.

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: Staff continue to work on simulation study to fit propensity scores using treatment and potential outcome variables in generalized boosted regression modeling algorithm for estimating propensity scores with continuous treatments. Staff conducted a comprehensive model diagnostic to confirm that the simulation data generated using Normal and Bernoulli distributions followed and met the normality assumptions.

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 around the Census Bureau to host 2018 *SUMMER AT CENSUS* with forty-two scholars.

Staff: Tommy Wright (x31702), Joseph Engmark

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: Brett Moran, Kelly Taylor

3. PUBLICATIONS

3.1 JOURNAL ARTICLES, PUBLICATIONS

Arima, S., Bell, W. R., Datta, G. S., Franco, C., and Liseor, B. (2017). "Multivariate Fay–Herriot Bayesian Estimation of Small Area Means under Functional Measurement Error." *Journal of the Royal Statistical Society—Series A. 180 (4),* pp. 1191-1209.

Ashmead, R., Slud, E. and Hughes, T. (2017). "Adaptive Intervention Methodology for Reduction of Respondent Contact Burden in the American Community Survey," *Journal of Official Statistics, Vol. 33, No. 4*, pp. 901–919.

Bell, W. R., Chung, H. C., Datta, G. S., and Franco, C. (In Press). "Measurement Error in Small Area Estimation: Functional vs. Structural vs. Naïve Models," *Survey Methodology*.

Ben-David, E. and Slawski, M. (In Press). "Linear Regression with Sparsely Permuted Data," *Electronic Journal of Statistics.*

Chakraborty, A., Datta, G.S., and Mandal, A. (In Press). "Robust Hierarchical Bayes Small Area Estimation for Nested Error Regression Model;" *International Statistical Review*.

Datta, G.S., Rao, J.N.K., Torabi, M., and Liu, B. (In Press). "Small Area Estimation with Multiple Covariates Measured with Errors: A Linear Regression Approach of Combining Two Surveys," *Journal of Multivariate Analysis.*

de Oliveira, V., Wang, B., and Slud, E. (In Press). "Spatial Modeling of Rainfall Accumulated over Short Periods of Time," *Journal of Multivariate Analysis.*

Franco, C., Little, R. J. A., Louis, T. A., and Slud, E. V. (In Press). "Comparative Studies of Confidence Intervals for Proportions in Complex Surveys," *Journal of Survey Statistics and Methodology.*

Keller, A, Mule, V.T., Morris, D.S. and Konicki, S. (2018). "A Distance Metric for Modeling the Quality of Administrative Records for Use in the 2020 U.S. Census," *Journal of Official Statistics*, 34(3): 1-27.

Klein, M. and Datta, G. (2018). "Statistical Disclosure Control Via Sufficiency under the Multiple Linear Regression Model," *Journal of Statistical Theory and Practice, 12:1*, 100-110.

Klein, M., Zylstra, J., and Sinha, B. (In Press). "Finite Sample Inference for Multiply Imputed Synthetic Data under a Multiple Linear Regression Model," *Calcutta Statistical Association Bulletin.*

Lin, W., Huang, J., and McElroy, T. (2018). "Time Series Seasonal Adjustment Using Regularized Singular Value Decomposition." Published online, *Journal of Business and Economics Statistics*.

Livsey, J., Lund, R., Kechagias, S., and Pipiras, V. (2018). "Multivariate Integer-valued Time Series with Flexible Autocovariances and Their Application to Major Hurricane Counts," *The Annals of Applied Statistics, 12*, #1, 408-431.

Lu, Bo, and Robert Ashmead. (2018). "Propensity Score Matching Analysis for Causal Effects with MNAR Covariates," *Statistica Sinica* 28, 2005-2025.

Lu, X. (2018). "On Mini-Max Pair in Tournaments," *Graphs and Combinatorics. 34*: 613-618. https://doi.org/10.1007/s00373-018-1899-3

Martin, A., Raim, A.M., Huang, W., and Adragni, K.P. (In Press). "ManifoldOptim: An R Interface to the ROPTLIB Library for Riemannian Manifold Optimization," *Journal of Statistical Software*.

McElroy, T. (2018). "Seasonal Adjustment Subject to Accounting Constraints," *Statistica Neerlandica* 72, 574-589.

McElroy, T. (2018). "Recursive Computation for Block Nested Covariance Matrices." Published online, *Journal of Time Series Analysis, 39(3),* 299-312.

McElroy, T. and Jach, A. (In Press). "Subsampling Inference for the Autocorrelations of GARCH Processes." Published online, *Journal of Financial Econometrics*.

McElroy, T., Pang, O., and Sheldon, G. (2018). "Custom Epoch Estimation for Surveys." Published online, *Journal of Applied Statistics*.

McElroy, T. and Roy, A. (2018) "The Inverse Kullback Leibler Method for Fitting Vector Moving Averages." *Journal of Time Series Analysis, 39*, 172-191.

Mulry, Mary H., Kaputa, S, and Thompson, K. J. (2018). Setting M-Estimation Parameters for Detection and Treatment of Influential Values. *Journal of Official Statistics. 34(2),* 483–501. Doi: http://dx.doi.org/10.2478/JOS-2018-0022

Nagaraja, C. and McElroy, T. (2018) "The Multivariate Bullwhip Effect." *European Journal of Operations Research, 267*, 96-106.

Nayak, T., Zhang, C., and You, J. (2018). "Measuring Identification Risk in Microdata Release and Its Control by Post-randomization," *International Statistical Review*, 86(2), 300-321.

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

Sellers, K.F. and Morris, D.S. (2017). "Under-dispersion Models: Models that are 'Under the Radar'," *Communications in Statistics – Theory and Methods, 46 (24):* 12,075-12,086.

Steorts, R. Tancredi, A., and Liseo, B. (In Press). "Generalized Bayesian Record Linkage and Regression with Exact Error Propagation," *Privacy in Statistical Databases.*

Wildi, M. and McElroy, T. (In Press) "The Trilemma Between Accuracy, Timeliness, and Smoothness in Real-Time Signal Extraction." *International Journal of Forecasting*.

Wojciech, C., Murphy, J.M., and Weinberg, D. (2018). "Superresolution of Noisy Remotely Sensed Images Through Directional Representations," *IEEE Geoscience and Remote Sensing Letters, 99*, 1-5.

Yang, Y. and Mathew, T. (2018). "The Simultaneous Assessment of Normality and Homoscedasticity in Linear Fixed Effects Models," *Journal of Statistical Theory and Practice, 12:1*, 66-81.

Zhang, W., Liu, A., Albert, P.S., Ashmead, R.D., Schisterman, E.F., and Mills, J.L. (2018). "A Pooling Strategy to Effectively Use Genotype Data in Quantitative Traits Genome-Wide Association Studies," *Statistics in Medicine*.

Zhao, J. and Mathew, T. (In Press). "Some Point Estimates and Confidence Regions for Multivariate Inter-Laboratory Data Analysis," *Sankhya*.

3.2 BOOKS/BOOK CHAPTERS

Erciulescu, A., Franco, C., and Lahiri, P. (In Press). "Use of Administrative Records in Small Area Estimation," in A.T. Chun and M. Larsen (Eds). *Administrative Records for Survey Methodology*, New York: Wiley.

Steorts, R.J., Tancredi, A., and Liseo, B. (2018). "Generalized Bayesian Record Linkage and Regression with Exact Error Propagation" in *Privacy in Statistical Databases (Lecture Notes in Computer Science 11126)* (Eds.) Domingo-Ferrer, J. and Montes, F., Springer, 297-313.

Steorts, R.J. and Shrivastava, A. (2018). "Probabilistic Blocking with an Application to the Syrian Conflict," in *Privacy in Statistical Databases (Lecture Notes in Computer Science 11126)* (Eds.) Domingo-Ferrer, J. and Montes, F., Springer, 314-327.

Thibaudeau, Y., Slud, E., and Cheng, Y. (In Press). "Log-Linear Modelling for Estimation of Cross-Classified Small Areas in Longitudinal Surveys," in *Methodology of Longitudinal Surveys* 2, Lynn Ed., New York: Wiley Series in Survey Methodology.

Wright, T. (2018). "No Calculation When Observation Can Be Made," in A.K. Chattopadhyay and G. Chattopadhyay (Eds.), *Statistics and Its Applications*, Springer Singapore, 139-154.

3.3 PROCEEDINGS PAPERS

Joint Statistical Meetings, American Statistical Association, Baltimore, Maryland, June 30-August 3, 2017.
2017 Proceedings of the American Statistical Association

- Robert Ashmead and Eric Slud, "Small Area Model Diagnostics and Validation with Applications to the *Voting Rights Act*, *Section 203*, 1515-1523.
- Laura Bechtel, Nicole Czaplicki, Maria Garcia, and Jeremy Knutson, "Resolving Balance Complex Discrepancies in the Presence of Negative Data," 1951-1974.
- Dong, K., Trudell, T., Cheng, Y. and Slud, E., "Understanding Variance Estimator Bias in Stratified Two-Stage Sampling," Survey Research Methods Section.
- Tucker McElroy and Brian Monsell, "Issues Related to the Modeling and Adjustment of High Frequency Time Series," 1901-1925.
- Mary Mulry, Nancy Bates, and Matt Virgile, "Lifestyle Segments, Social Marketing, and Hard-to-Survey Populations: Understanding Participation in the 2015 Census Test,"678-692.
- Andrew Raim, Scott Holan, Jonathan Bradley, and Christopher Wikle, "A Model Selection Study for Spatio-Temporal Change of Support," 1524-1540.
- Eric Slud and Robert Ashmead, "Hybrid BRR and Parametric-Bootstrap Variance Estimates for Small Domains in Large Surveys," 1716-1730.
- Trudell, T., Dong, K., Cheng, Y., and Slud, E. (2018), "Domain Estimation and Successive Difference Replication," Survey Research Methods Section.
- Daniel Weinberg, Tucker McElroy, and Soumendra Lahiri, "Estimation of Locally Stationary Spatial Processes with Application to the American Community Survey," 3097-3109.

3.4 CENTER FOR STATISTICAL RESEARCH & METHODOLOGY RESEARCH REPORTS

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

RR (Statistics #2017-07): Jichong Chai and Tapan Nayak, "A Criterion for Privacy Protection in Data Collection and Its Attainment via Randomized Response Procedures," October 26, 2017.

RR (Statistics #2017-08): Tucker S. McElroy and Brian C. Monsell, "Issues Related to the Modeling and Adjustment of High Frequency Time Series," October 26, 2017.

RR (Statistics #2018-01): Tucker McElroy, Brian Monsell, and Rebecca Hutchinson, "Modeling of Holiday Effects and Seasonality in Daily Time Series," January 23, 2018.

RR (Statistics #2018-02): Ricardo Moura, Martin Klein, John Zylstra, Carlos Coelhio, and Bimal Sinha, "Inference for Multivariate Regression Model based on Synthetic Data Generated Using Plug-in Sampling," February 5, 2018.

RR (Statistics #2018-03): Tucker McElroy and Anindya Roy, "Model Identification via Total Frobenius Norm of Multivariate Spectra," March 6, 2018.

RR (Statistics #2018-04): Martin Klein, Tommy Wright, and Jerzy Wieczorek, "A Simple Joint Confidence Region for A Ranking of K Populations: Application to American Community Survey's Travel Time to Work Data," March 8, 2018.

RR (Statistics #2018-05): William Winkler, "Cleaning and Using Administrative Lists: Enhanced Practices and Computational Algorithms for Record Linkage and Modeling/Editing/Imputation," April 3, 2018.

RR (Statistics #2018-06): William Bell, Hee Chung, Gauri Datta, and Carolina Franco, "Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naïve Models," May 17, 2018.

RR (Statistics #2018-07): David Findley and Tucker McElroy, "Background and Perspectives for ARIMA Model-Based Seasonal Adjustment," May 22, 2018.

RR (Statistics #2018-08): Tucker McElroy and Marc Wildi, "The Multivariate Linear Prediction Problem: Model-Based and Direct Filtering Solutions," August 16, 2018.

RR (Statistics #2018-09): Andrew Raim, Elizabeth Nichols, and Thomas Mathew, "A Statistical Comparison of Call Volume Uniformity Due to Mailing Strategy," September 13, 2018.

RR (Statistics #2018-10): Thomas Trimbur and William Bell, "The Effects of Seasonal Heteroskedasticity in Time Series on Trend Estimation and Seasonal Adjustment," September 24, 2018.

RR (Statistics #2018-11): Cheng Zhang and Tapan Nayak, "Post-randomization for Identification Risk Limited Microdata Release from General Surveys," October 5, 2018.

4. TALKS AND PRESENTATIONS

Arizona State University, Tempe, Arizona, October 5, 2017.

- Tucker McElroy, "Model Identification via the Total Frobenius Norm of Multivariate Spectra."

Women in Statistics and Data Science, San Diego, California, October 19-21, 2017.

- Maria Garcia, Panelist on "Is there Room for Intersectional Feminism in Statistics and Data Science."
- Darcy Morris, Panelist on "The Imposter Syndrome."
- Kimberly Sellers, Panelist on "The Imposter Syndrome."

Department of Statistics, Charles III University of Madrid, Madrid, Spain, December 1, 2017.

- Carolina Franco, "Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naive Models."

11th International Conference on Computational and Financial Econometrics, London, England, December 16-18, 2017.

- Tucker McElroy, ``Real-time Signal Extraction of Vector Time Series via Multivariate Direct Filter Analysis."

International Indian Statistical Association's International Conference on Statistics, Hyderabad, India, December 27-30, 2017.

- Carolina Franco, "Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naive Models."
- Gauri Datta, "Finite Mixture Modeling in Small Area Estimation: New Alternatives to Some Popular Models."

IASSL International Conference, Colombo, Sri Lanka, December 28-29, 2017.

- Bimal Sinha, "Data Analysis Under Privacy Protection."

International Conference in Statistics and Probability, Kolkata, January 2-4, 2018.

- Tapan Nayak, "A Criterion for Protecting Privacy in Surveys and Its Attainment via Randomized Response."

International Conference on Statistics and its Applications (ICSA) with an Emphasis on Clinical and Official Statistics, Kerala, India, January 3-5, 2018.

- Andrew Raim, "An R Package for Spatio-Temporal Change of Support."
- George Ostrouchov, Nagaraj Neerchal, and Andrew Raim, "High Performance Statistical Computing using R."
- Bimal Sinha, "Data Analysis Under Privacy Protection."

Mathematics Department, Statistics Seminar, University of Maryland, College Park, Maryland, February 8, 2018.

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

Federal Committee on Statistical Methodology, 2018 Research & Policy Conference, Washington, D.C., March 7-9, 2018.

- Benmei Liu and Isaac Dompreh, "Small Area Estimation for Measures Related to Tobacco Use and Policies Using the Tobacco Use Supplement to the Current Population Survey."
- Xiaoyu Zhai and Tapan Nayak, "A Note on Multiplicative Noise Perturbation for Privacy Protection."

Public Opinion Quarterly (POQ) Special Issue Conference, Washington, DC, March 16, 2018.

- Darcy Morris, "A Modeling Approach for Administrative Records Enumeration in the Decennial Census."

5th African International Conference on Statistics, University of Botswana, Gaborone, Botswana, March 19-22, 2018.

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

Eastern North American Region (ENAR) Spring Meeting 2018, Atlanta, GA, March 25-28, 2018.

- Kimberly Sellers and Andrew Raim, "A Flexible Zero-inflated Count Model to Address Data Dispersion."

2nd Seasonal Adjustment Practitioners Workshop, Washington, DC, April 26 2018.

- James Livsey, "Applying the EM Algorithm to Multivariate Signal Extraction."
- Tucker McElroy, "A Diagnostic for Seasonality Based upon AutoRegressive Roots."
- Brian Monsell, "X-13 Stuff You Should Know."
- Osbert Pang, "Examining the Performance of Seasonality Diagnostics for Detecting Residual Seasonality."
- Thomas Trimbur, "A Modelled Approximation to the Ideal Filter for Nonstationary Time Series with Applications to Business Cycle Fluctuations."

50th Anniversary Celebration Conference (of the Statistics Department), The Pennsylvania State University, University Park, PA, May 8-10, 2018.

- Jerry Maples, "Improving Small Area Estimates of Disability: A Model-Based Approach to Combining the American Community Survey with the Survey of Income and Program Participation."

Eighth International Workshop on Innovative Statistical Methods, Daejeon, Korea, May 15-16, 2018.

- Gauri Datta, "Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naïve Models."

Statistical Society of Canada, Annual Meeting, McGill University, Montreal, Canada, June 3-6, 2018.

- Gauri Datta, "Workshop on Small Area Estimation."

2018 International Total Survey Error Workshop, Duke University, Durham, NC, MD, June 4 -6, 2018.

- Mulry, Mary H., Bates, Nancy and Virgile, Matt, "Using Lifestyle Segments to Enhance Response Propensity and Design Nonresponse Interventions."

Small Area Estimation and Other Topics of Current Interest in Surveys, Official Statistics, and General Statistics, East China Normal University, Shanghai, China, June 16-18, 2018.

- Gauri Datta, "Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naïve Models."
- Carolina Franco, "Comparative Study of Confidence Intervals for Proportions in Complex Sample Surveys."
- Jerry Maples, "Small Area Models for Over-dispersed Poisson Counts."
- Eric Slud, "Hybrid BRR and Parametric-Bootstrap Variance Estimates for Small Domains in Large Surveys."
- Yves Thibaudeau, "Fixed-Effect Log-Linear Models for Small Area Estimation in Presence of Relatively Sparse Contingency Tables."
- Tommy Wright, "No Calculation When Observation Can Be Made."

38th International Symposium on Forecasting, Boulder, CO, June 17-20, 2018.

- Thomas Trimbur, "A Modelled Approximation to the Ideal Filter for Nonstationary Time Series with Applications to Business Cycle Fluctuations."
- Patrick Joyce, "Modelling Survey Time Series Data with Flow-Observed CARMA Processes."

9th International Workshop on Applied Probability, Budapest, Hungary, June 18-21, 2018.

- Tapan Nayak, "Probabilistic Methods for Data Perturbation for Protecting Respondent's Privacy."

Conference on Multivariate Count Analysis, University of Franche-Comté, Besançon, France, July 4-6, 2018.

- Darcy Steeg Morris, "A COM-Poisson Mixed Model for Clustered Count Data."
- Kimberly Sellers, "Bivariate Conway–Maxwell–Poisson Distribution: Formulation, Properties, and Inference."

ESRI User Conference, San Diego, CA, July 9-13, 2018.

- Mary Mulry, "Viewing Participation in the U.S. Census through the Lens of Lifestyle Segments."

International Conference on Combinatorics and Graph Theory, Lyon, France, July 9-13, 2018.

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

Joint Statistical Meetings, American Statistical Association, Vancouver, Canada, July 28-August 2, 2018.

- Robert Ashmead, "A Top-Down Algorithm for Releasing Differentially Private Hierarchical Multi-Dimensional Contingency Tables with Exact Constraints."
- Martin Klein, "Finite Sample Inference for Multiply Imputed Synthetic Data Under a Multiple Linear Regression Model."
- Jerry Maples, "Small Area Population Models: Estimating the Number of Children in School Districts."
- Brian Monsell, "Benchmarking Monthly Seasonally Adjusted Series to Quarterly Adjustments."

- Darcy Steeg Morris, “A Conway-Maxwell-Multinomial Distribution for Flexible Modeling of Categorical Data.”
- Mary Mulry, “Relationship between Positive Responses to Child-Specific Probes on the 2010 Census Questionnaire and 2010 Census Coverage Measurement NonMatching Young Children.”
- Osbert Pang, “Examining the Performance of Seasonality Diagnostics for Detecting Residual Seasonality.”
- Andrew Raim, “A Statistical Comparison of Call Volume Uniformity Due to Mailing Strategy.”

Model Uncertainty: Mathematical and Statistical Workshop, Statistical and Applied Mathematical Sciences Institute, Durham, NC, August 20-24, 2018.

- Emanuel Ben-David, “Linear Regression with Merged Data Files.”

NBER/NSF Annual Time Series Conference, San Diego, CA, September 7-8, 2018.

- Thomas Trimbur, “Modeled Approximations to the Ideal Filter with Application to Time Series of Gross Domestic Product.”

5. CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY SEMINAR SERIES

Tapan Nayak, The George Washington University/U.S. Census Bureau, "A Criterion for Protecting Respondent's Privacy and Its Attainment via Randomized Response," November 28, 2017.

Yves Thibaudeau, U.S. Census Bureau, "The Versatility of Log-Linear Models in Survey Estimation," December 5, 2017.

Rebecca Steorts, Duke University/U.S. Census Bureau, "Entity Resolution with Societal Impacts in Statistical Machine Learning," December 12, 2017.

Carolina Franco, U.S. Census Bureau, "Comparative Study of Confidence Intervals for Proportions in Complex Sample Surveys," January 23, 2018.

Brandon Park, George Mason University, "Supervised Clustering via an Implicit Network for High Dimensional Data," March 22, 2018.

David Dreisigmeyer, U.S. Census Bureau, "Two Dimensionality Reduction Techniques and Their Solution Method," April 2, 2018.

Kyle Irimata, Arizona State University, "Partitioned GMM Logistic Regression Models for Time-Dependent Covariates in Longitudinal Studies," April 3, 2018.

Rebecca Steorts, Duke University/U.S. Census Bureau, "Some of Entity Resolution Workshop," May 2, 2018.

Tim McMurry, University of Virginia, School of Medicine, "Time Series Inference and Prediction through Estimation of the Autocovariance Matrix," May 8, 2018.

Jian-Guo Liu, Duke University, *SUMMER AT CENSUS,* "Analysis of Some Machine Learning Algorithms: Stochastic Gradient Descent and Online PCA," May 15, 2018.

Vladas Pipiras, University of North Carolina – Chapel Hill, "Several Aspects of Modeling Univariate and Multivariate Periodic Time Series," May 22, 2018.

Yihren Wu, Hofstra University, *SUMMER AT CENSUS,* "A Jump-Diffusion Process for Asset Price with Non-Independent Jumps," May 24, 2018.

Jerzy Wieczorek, Carnegie Mellon University/Colby College, *SUMMER AT CENSUS,* "Using the R Package 'RankingProject' to Make Simple Visualizations for Comparing Populations: Demographic Edition," May 29, 2018.

Narayanaswamy Balakrishnan, McMaster University, *SUMMER AT CENSUS,* "Cure Models," May 29, 2018.

Jerzy Wieczorek, Carnegie Mellon University/Colby College, *SUMMER AT CENSUS,* "Using the R Package 'RankingProject' to Make Simple Visualizations for Comparing Populations: Economic Edition," May 30, 2018.

John L. Eltinge, U.S. Census Bureau, "Positive-Sum Innovation," May 31, 2018.

Yeonok Choi, Statistics Korea/NORC at the University of Chicago, *SUMMER AT CENSUS,* "Enhanced Income Statistics of Korea by Combining Survey and Administrative Data," May 31, 2018.

James Livsey and Andrew Raim, U.S. Census Bureau, and Iris Gauran, University of Maryland, Baltimore County, "Advanced Statistical Programming using Rcpp Workshop," May 24 & 31, 2018.

Jared Jobe, Retired Health Scientist (formerly with U.S. Army, CDC, and NIH), *SUMMER AT CENSUS,* "Self Report Research in Health Surveys, Epidemiology, and Clinical Trials," June 5, 2018.

Huixia Judy Wang, The George Washington University, *SUMMER AT CENSUS,* "Quantile Regression and Probabilistic Prediction," June 6, 2018.

Joshua Snoke (U.S. Census Bureau Dissertation Fellow) The Pennsylvania State University, "Statistical Data Privacy Methods for Increasing Research Opportunities," June 7, 2018.

Rachel Cummings, Georgia Tech University, *SUMMER AT CENSUS,* "Differential Privacy for Growing Databases," June 12, 2018.

Richard Alba, City University of New York, *SUMMER AT CENSUS,* "The Rise of Ethno-racially Mixed Parentage – Making Sense of the Projected Majority-Minority America," June 13, 2018.

Irma Elo, University of Pennsylvania, *SUMMER AT CENSUS,* "Immigration and Longevity in the United States," June 13, 2018.

Janna Johnson, University of Minnesota, *SUMMER AT CENSUS,* "The U.S. Census Undercount of Native-Born Children: Estimates, Correlates, and Implications," June 14, 2018.

Carolyn Liebler, University of Minnesota, *SUMMER AT CENSUS,* "Patterns in the Race Responses for Young Children of All Intermarriages in the U.S., 1960-2010," June 14, 2018.

Johannis Schouten, Utrecht University, *SUMMER AT CENSUS,* "Using Mobile Devices to Collect Household Survey Data," June 18, 2018.

Alexis Santos, The Pennsylvania State University, *SUMMER AT CENSUS,* "Revisiting the Demography of Disaster: Population Estimates after Hurricane María," June 19, 2018.

Michael Wiklund, Tufts University, *SUMMER AT CENSUS,* "Lessons Learned from Applying Human Factors Engineering to Medical Technology," June 19 & 20, 2018.

Bikas Sinha, Retired Professor of Statistics, Indian Statistical Institute, Kolkata, *SUMMER AT CENSUS,* "Block Total Response Techniques for Sensitive Quantitative Features," June 25, 2018.

Marlese Durr, Wright State University, *SUMMER AT CENSUS,* "Braiding, Slicing, and Dicing: The African American Woman's Home as a Site of Work" June 25, 2018.

Richard Valliant, Research Professor Emeritus, University of Michigan & University of Maryland, *SUMMER AT CENSUS,* "Comparing Alternatives for Estimation from Nonprobabililty Samples," June 26, 2018.

Richard Valliant, Research Professor Emeritus, University of Michigan & University of Maryland, *SUMMER AT CENSUS,* "Alternatives for Estimation from Nonprobability Samples Workshop," June 26 & 27, 2018.

Malay Ghosh, University of Florida, *SUMMER AT CENSUS,* "Global-Local Shrinkage Priors for Small Area Estimation," July 10, 2018.

Becky Pettit, University of Texas, Austin, *SUMMER AT CENSUS*, "Invisible Men: Mass Incarceration and the Myth of Black Progress," July 11, 2018.

John L. Eltinge, U.S. Census Bureau, "Improving the Quality and Value of Statistical Information: Fourteen Questions on Management," July 12, 2018.

Chinhui Juhn, University of Houston, *SUMMER AT CENSUS*, "Coordinated Work Schedules and the Gender Wage Gap," July 16, 2018.

Dongchu Sun, University of Missouri-Columbia, *SUMMER AT CENSUS*, "An Objective Prior for Hyper-parameters in Normal Hierarchical Models," July 17, 2018.

Zhuoqiong He, University of Missouri-Columbia, *SUMMER AT CENSUS*, "Noninformative Priors for Multinomial and Capture-recapture Data," July 17, 2018.

Michael McDonald, University of Florida, *SUMMER AT CENSUS*, "Item Non-Response and Over-Report Bias in the Current Population Survey Voting and Registration Supplement," July 23, 2018.

Michael Traugott, University of Michigan, *SUMMER AT CENSUS*, "The Role of Rotation Group Bias in Overestimating Turnout in the Current Population Survey," July 23, 2018.

Victor de Oliveira, University of Texas, San Antonio, *SUMMER AT CENSUS*, "Gaussian Copula Models for Geostatistical Count Data," July 24, 2018.

Allyson Holbrook, University of Illinois-Chicago, *SUMMER AT CENSUS*, "Self-Reported Voter Turnout and Over-Reporting: Social Desirability, Memory Errors, and Alternative Question Wording," July 24, 2018.

Stanley Presser, University of Maryland, College Park, *SUMMER AT CENSUS*, "The Link Between Nonresponse and Social Participation: Implications for Weighting," July 24, 2018.

Barbara Entwisle, University of North Carolina at Chapel Hill, *SUMMER AT CENSUS*, "Behavioral and Social Science Insights for Big Data Research," July 30, 2018.

Nick Sinai, Harvard University, *SUMMER AT CENSUS*, "Simplifying the Census: Focus on the Data," August 7, 2018.

Xiuying Qian, Zhejiang University, *SUMMER AT CENSUS*, "Profiling Psychological Needs for Mobile App Usage through Big Data Analytics," August 8, 2018.

Rahul Mazumder, MIT, *SUMMER AT CENSUS*, "Mining Events with Declassified Diplomatic Documents," August 14, 2018.

Kala Krishna, The Pennsylvania State University, *SUMMER AT CENSUS*, "Trade and Minimum Wages in General Equilibrium: Theory and Evidence," August 16, 2018.

Ricardo Moura, Portuguese Naval Academy, *SUMMER AT CENSUS*, "Singly Imputed Synthetic Datasets as SDC Under MLR: The Feasibility of Its Analysis," August 21, 2018.

Yingchun Zhou, East China Normal University, *SUMMER AT CENSUS*, "Functional Multiple-Outcome Model in Application to Multivariate Growth Curves of Infant Data," August 22, 2018.

Zhen-Qing Chen, University of Washington, *SUMMER AT CENSUS*, "Anomalous Diffusions and Fractional Order

Differential Equations," August 23, 2018.

Martin Slawski, George Mason University, "Regression with Sparsely Permuted Data," September 4, 2018.

Jeff Jonas, Senzing, Founder and CEO, *SUMMER AT CENSUS*, "Artificial Intelligence (AI) and Entity Resolution 2.0," September 6, 2018.

Robert Belli, University of Nebraska – Lincoln, *SUMMER AT CENSUS*, "Interviewer-Respondent Interactions and Data Quality in Flexible Timeline Interviewing," September 10, 2018.

Kristen Olson & Jolene Smyth, University of Nebraska – Lincoln, *SUMMER AT CENSUS*, "How Do Interviewers Do What Interviewers Do and What Does That Tell Us About Questionnaire Design and Survey Practice?" September 10, 2018.

Leen-Kiat Soh, University of Nebraska – Lincoln, *SUMMER AT CENSUS*, "Understanding Interviewer and Respondent Behaviors in a CATI, Web-Based ATUS Instrument," September 10, 2018.

Adam Eck, Oberlin College, *SUMMER AT CENSUS*, "Sample Frame Construction from Image Data (and Other Applications of Deep Learning for Survey Research)," September 10, 2018.

Joseph Schafer, U.S. Census Bureau, "Flexible Modeling of Incomplete Multivariate Data for Inference, Imputation, and Prediction," September 12, 2018.

6. PERSONNEL ITEMS

6.1 HONORS/AWARDS/SPECIAL RECOGNITION

Bronze Medal Award, U.S. Bureau of the Census

- **Darcy Morris, … (Team Award)** – The group is recognized for developing and implementing methods and systems to use administrative records and third-party data to reduce the number of visits during the Nonresponse Follow-up operation in the 2020 Census. They developed predictive models to identify occupied, vacant, and nonexistent housing units that could receive fewer field contacts. Based on these innovations, up to 25 percent of nonresponding cases requiring fieldwork to complete their 2020 Census enumeration will receive fewer visits or none at all – saving as much as $1.4 billion and reducing burden on the respondents.

Elected Member, International Statistical Institute

- **Kimberly Sellers**

6.2 SIGNIFICANT SERVICE TO PROFESSION

Robert Ashmead

- Reviewed a paper for *The American Statistician*

Emanuel Ben-David

- Refereed papers for *Mathematical Reviews* and *American Mathematics Society Journal* (AMS).
- Member, Ph.D. Proposal Committee, George Mason University

Gauri Datta

- Associate Editor, *Sankhya*
- Associate Editor, *Statistical Methods and Applications*
- Associate Editor, *Environmental and Ecological Statistics*
- Editorial Member, *Calcutta Statistical Association Bulletin*

Carolina Franco

- Member, Ph.D. Dissertation Panel, Charles III University of Madrid
- Member, Gertrude Cox Scholarship Committee
- Refereed papers for *Statistical Science, The American Statistician, the Journal of the Royal Statistical Society, Series A,* and the *Journal of Official Statistics*

Maria Garcia

- Member, Program Committee, FCSM 2018 Research and Policy Conference
- Session Organizer: "Topics in Editing and Imputation: Automated Systems, Machine Learning, Response Propensities, and Edit Reduction", FCSM 2018 Research and Policy Conference, Washington, D.C.
- Panelist, Women in Statistics and Data Science Conference, San Diego, California
- Panelist, Infinite Possibilities Conference, Washington, D.C.

Ryan Janicki

- Refereed papers for *Environmental and Ecological Statistics, Computational Statistics & Data Analysis, Journal of the Royal Statistical Society, Series A, The American Statistician,* and *Biometrika.*

Patrick Joyce

- Refereed a paper for *Journal of Official Statistics*

Martin Klein

- Co-advisor, Ph.D. Dissertation in Statistics Committee, University of Maryland, Baltimore County
- External Examiner, Ph.D. Dissertation in Statistics Committee, George Washington University
- Refereed papers for *Journal of Official Statistics* and *Sankhya B*

James Livsey

- Reviewed papers for *The Annals of Statistics, Journal of Time Series Analysis, Stochastic Environmental Research*

and Risk Assessment, The American Statistician, Applied Stochastic Models in Business and Industry, International Statistical Review, and *Sankhya B.*

Xiaoyun Lu
- Refereed a paper for *Discussiones Mathematicae Graph Theory*

Jerry Maples
- Reviewer, Research Grant Proposal, National Science Foundation

Thomas Mathew
- Associate Editor, *Journal of the American Statistical Association*
- Associate Editor, *Sankhya*

Darcy Morris
- Panelist, Women in Statistics and Data Science Conference, San Diego, California, October 2017
- Associate Editor, *Communications in Statistics*
- Treasurer-Elect, ASA Survey Research Methods Section
- Refereed a paper for *Statistical Modelling*

Mary Mulry
- Associate Editor, *Journal of Official Statistics*
- Methodology co-Editor, *Statistical Journal of the International Association of Official Statistics*
- Member, Ph.D. Dissertation Committee, Statistics Sciences Department, Southern Methodist University

Tapan Nayak
- Associate Editor, *Communications in Statistics*
- Associate Editor, *Journal of Statistical Theory and Practice*

Ned Porter
- Reviewed papers for 5th IEEE International Conference on Data Science and Advanced Analytics

Kimberly Sellers
- Panelist, Women in Statistics and Data Science Conference, San Diego, California, October 2017
- Chairperson, American Statistical Association Committee on Women in Statistics
- Associate Editor, *The American Statistician*
- Associate Editor, *Journal of Computational and Graphical Statistics*
- Member, Advisory Board, International Black Doctoral Network Association, Incorporated
- Member, Adele's Circle of Women, University of Maryland, College Park, MD
- Refereed papers for *Statistics and Probability Letters* and *Communications in Statistics – Theory and Methods*

Eric Slud
- Associate Editor, *Biometrika*
- Associate Editor, *Journal of Survey Statistics and Methodology*
- Associate Editor, *Lifetime Data Analysis*
- Associate Editor, *Statistical Theory and Related Fields*

Rebecca Steorts
- Associate Editor, *Journal of Survey Statistics and Methodology*
- Reviewer, Grant Proposals, National Science Foundation Panel
- Area Chair, Women in Machine Learning Workshop
- Member, Organizing Committee, Workshop on Bayes, Big Data, and Social Good

Thomas Trimbur
- Refereed papers for *Journal of Forecasting* and *Statistica Nederlandica*

Daniel Weinberg
- Refereed papers for *The American Statistician* and *Annals of Statistics*

William E. Winkler

- Associate Editor, *Transactions on Data Privacy, Journal of Privacy and Confidentiality*
- Refereed papers for *Annals of Applied Statistics* and *American Political Science Review*
- Member, Computer Science Ph.D. Committee, Australia National University
- Member, Statistics Ph.D. Committee, University of Maryland, College Park
- Member, Program Committee, *Statistical Data Protection 2018*
- Discussant, Session on Record Linkage, 2017 Joint Statistical Meetings

Tommy Wright

- Associate Editor, *The American Statistician*
- Member, Board of Trustees, National Institute of Statistical Sciences
- Refereed paper for *Journal of Applied Statistics*

6.3 PERSONNEL NOTES

Gauri Datta joined our Small Area Estimation Research Group in a permanent position.

Rebecca Steorts (Statistics Faculty at Duke University) joined our Center on a Schedule A appointment.

Lauren Emanuel accepted a position in private industry.

Tucker McElroy accepted a Senior Technical position at the Census Bureau.

Joe Engmark joined our center to support research and collaboration.

Hang Kim (Statistics Faculty at University of Cincinnati) accepted an ASA/NSF/Census Research Fellowship to work on synthetic data for the Economic Census.

Kyle Irimata (Ph.D. in statistics from Arizona State University) joined our Small Area Estimation Research Group to work on the Small Area Income and Poverty Estimates (SAIPE) Program.

APPENDIX A **Center for Statistical Research and Methodology FY 2018**
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		
6250D02	Operational Design		
6550D01	Data Coding/Editing/Imputation		
6650D01	CCM Planning and Project Management		
6650D20	Evaluations-Planning and Project Management		
6750D01	Administrative Records Data		
	1. Decennial Record Linkage	William Winkler	Tom Mule
	2. Coverage Measurement Research	Jerry Maples	Tim Kennel
	3. Supplementing and Supporting Non-Response with Administrative Records	Michael Ikeda	Tom Mule
	4. Identifying "Good" Administrative Records for 2020 Census NRFU Curtailment Targeting	Darcy Steeg Morris	Tom Mule
	5. 2020 Census Communications Campaign Statistical Analyses	Mary Mulry	Nancy Bates
	6. Census Coverage Error of Young Children	Mary Mulry	Scott Konicki
	7. 2020 Census Privacy Research	Robert Ashmead	John Abowd
	8. Creation of Record-Level Public-Use File from the Census Barriers, Attitudes, and Motivators Survey Confidential Micro-data	Tommy Wright	Rob Sienkiewicz
6350D02	Address Canvassing In Field		
	9. Statistical Evaluation of the In-office Image Review Process.	Andrew Raim	April Avnayim
	10. Development of Block Data Tracking Database.	Tom Petkunas	Graham Baggett
	11. Statistical Inference on Call Volumes to Helplines	Andrew Raim	Beth Nichols
6385D70	American Community Survey (ACS)		
	12. ACS Applications for Time Series Methods	Tucker McElroy	Mark Asiala
	13. Voting Rights Section in 203 Model Evaluation and Enhancements Towards 2021 Determinations	Robert Ashmead	James Whitehorne
	14. Model-based Estimates to Improve Data Confidentiality for ACS Special Tabulations	Jerry Maples	John Abowd
	DEMOGRAPHIC		
TBA	Demographic Statistical Methods Division Special Projects		
	15. 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		
	16. Data Integration	Ned Porter	Christopher Boniface
7165018	Social, Economic, and Housing Statistics Division Small Area Estimation Projects		
	17. Research for Small Area Income and Poverty Estimates (SAIPE)	Carolina Franco	Wes Basel
	18. Small Area Health Insurance Estimates (SAHIE)	Ryan Janicki	Wes Basel
	19. Sub-County Estimates of Poverty from Multi-year ACS Data	Jerry Maples	Wes Basel
	ECONOMIC		
1183X01	Economic Statistical Collection		
	20. Research on Imputation Methodology for the Monthly Wholesale Trade Survey	Martin Klein	Joe Schafer
	21. Use of Big Data for Retail Sales Estimates	Darcy Steeg Morris	Carma Hogue
2270D10	Economic Census/Survey Engineering: Time Series Research; Economic Missing Data/Product Line Data; Development/SAS		
	22. Seasonal Adjustment Support	Brian Monsell	Kathleen McDonald-Johnson
	23. Seasonal Adjustment Software Development and Evaluation	Brian Monsell	Kathleen McDonald-Johnson
	24. Research on Seasonal Time Series: Modeling & Adjustment Issues	Tucker McElroy	Kathleen McDonald-Johnson
	25. Supporting Documentation & Software	Brian Monsell	Kathleen McDonald-Johnson
TBA	Investigation of Alternative Methods for Resolving Balance Complex Failures in StEPS		
	26. Developing Economic Census Synthetic Microdata	Maria Garcia	Jenny Thompson
	CENSUS BUREAU		
7225010	*27. National Cancer Institute Tobacco Use Survey/CPS*	Isaac Dompreh…	Benmei Liu

APPENDIX B

FY 2018 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 twentieth year to provide *seven measures of performance*, this time for the fiscal year 2018. For FY 2018, the *measures of performance* for our center are:

Measure 1. Overall, Work Met Expectations: Percent of FY 2018 Program Sponsored Projects/Subprojects where sponsors reported that work met their expectations.

Measure 2. Established Major Deadlines Met: Percent of FY 2018 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 2018 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight.

Measure 3b. Plans for Implementation: Of the FY 2018 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 2018 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 2018.

Measure 6. Proceedings Publications: Number of proceedings publications documenting research that appeared in FY 2018.

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 Present thru December 3, 2018 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 2018 ***(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 2018 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 2018 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 2018 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
Rebecca Steorts (Duke U.)

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
Carolina Franco
Kyle Irimata
Ryan Janicki

Time Series Research
Brian Monsell
Osbert Pang
Brian Monsell (Acting)
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
Joe Engmark
Hang Kim (F)
Michael Hawkins
VACANT

(F) ASA/NSF/Census Research Fellow

September 30, 2018