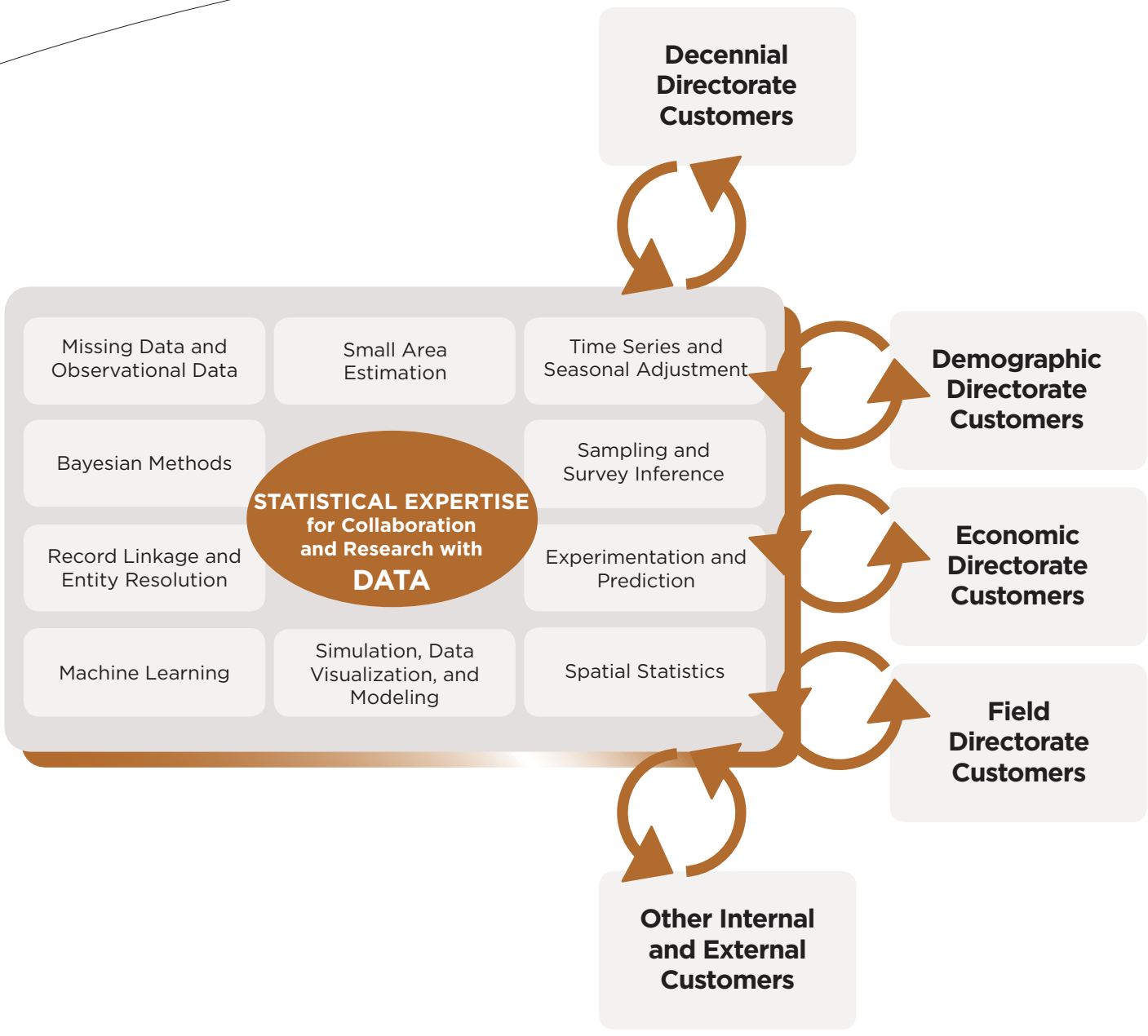
Annual Report of the Center for Statistical Research and Methodology

Research and Methodology Directorate

Fiscal Year 2019



United States® Census Bureau

U.S. Department of Commerce
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 2019, this report is an accounting of our work and our results.

Center for Statistical Research & Methodology
https://www.census.gov/topics/research/stat-research.html

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 2019 follow, and more details are provided within subsequent pages of this report:

- *Demonstrated with Statistical Models How American Community Survey Data Can Improve Statistical Estimates from Smaller Sample Surveys*: (1) Analyzed an application using American Community Survey data to improve statistical estimates of disability from the Survey of Income and Program Participation; (2) Showed that impressive reductions in MSEs/posterior variances can be achieved by using bivariate statistical models rather than the direct statistical estimators.

- *Enhanced Record-Linkage Capability at the Census Bureau:* (1) Provided technical leadership, training, and research for the evaluation of record-linkage software and multi-list methodology used with administrative records for the 2020 Census; (2) Led a series of sessions for Center for Statistical Research and Methodology and Center for Optimization and Data Science staff to increase knowledge of various record linkage software packages; (3) Conducted and published statistical research to assess regression analysis in the presence of linkage error; and (4) Researched the problem of merging noisy databases to remove duplicate entities where typically a unique identifier is not known.

- *Statistical Analyses for 2020 Census Communications Campaign*: (1) Investigated with others, the use of a lifestyle segmentation of the U.S. population to gain insight about variation in self-response in censuses and sample surveys to determine whether the segments could be useful in the 2020 Census Communications Campaign; (2) Developed statistical models that showed that adding lifestyle segments to the Low Response Scores improves the prediction of self-response rates at the tract level.

- *Advanced Small Area Estimation Methodology*: (1) Expanded the statistical theory behind the moment based MSE estimation approach in small area estimation; (2) Developed a bootstrapping method for approximating MSE with moment based estimators for the model parameters; (3) Developed and fitted two independent share statistical models to estimate school district to county share of both the poor and non-poor school age children.

- *With Economic Directorate, Used Statistical Models and Commercial Data to Investigate the Feasibility of Producing Statistical State Level Retail Estimates*: (1) Gained acceptance of a manuscript which describes industry-level profiles and a statistical methodology for producing geographically granular (at state level) retail sales experimental statistical estimates for select industries using commercial monthly retail sales data. Challenges are documented.

- *Conducted Statistical Research to Improve Seasonal Adjustment of Economic Data:* (1) Assessed the impact of weather regressors on seasonal adjustment, modeled and analyzed daily time series, and assessed new results on seasonality diagnostics based upon autoregressive roots; (2) Continued statistical research that documents currently available seasonal adjustment diagnostics, as part of work with the Bureau of Economic Analysis (BEA) on residual seasonality in GDP.

- *Developed Hybrid Design-based and Model-based Methodology for Small-cell Gross Flow Estimates at the State Level in the Current Population Survey*.

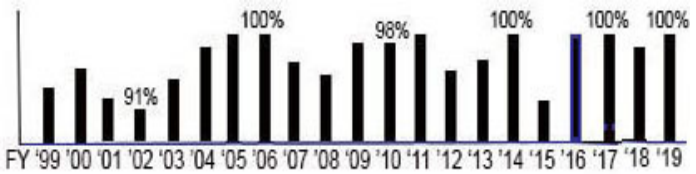
- *SUMMER AT CENSUS:* Sponsored, with divisions around the Census Bureau, scholarly, short-term visits by 29 researchers/leaders who collaborated extensively with us and presented seminars on their research. For a list of the 2019 *SUMMER AT CENSUS* scholars, see *https://www.census.gov/topics/research/events/summer-at-census.html*.

How Did We[1] Do...

For the 21st year, we received feedback from our sponsors. Near the end of fiscal year 2019, 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 21 fiscal years):

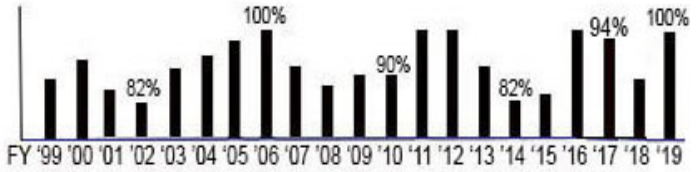
Measure 1. Overall, Work Met Expectations

Percent of FY2019 Program Sponsored Projects/Subprojects where sponsors reported that overall work met their expectations (agree or strongly agree) (27 out of 27) ……..…….... 100%



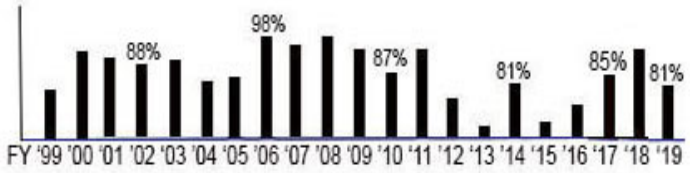
Measure 2. Established Major Deadlines Met

Percent of FY2019 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met (16 out of 16 responses) …….....…...…...…... 100%



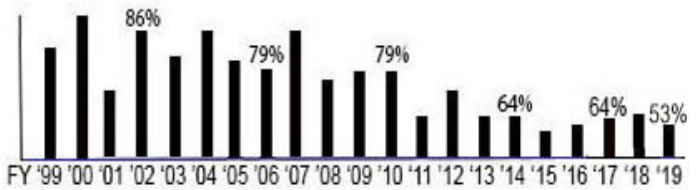
Measure 3a. At Least One Improved Method, Developed Technique, Solution, or New Insight

Percent of FY2019 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight (22 out of 27 responses) ……..…… 81%



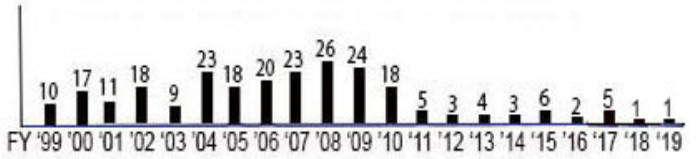
Measure 3b. Plans for Implementation

Of these FY2019 Program Sponsored Projects/Subprojects reporting at least one improved method, technique developed, solution, or new insight, the percent with plans for implementation (9 out of 17 responses) ……..…….. 53%



Measure 4. Predict Cost Efficiencies

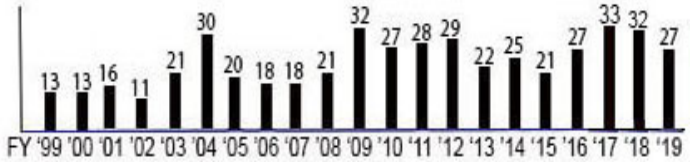
Number of FY2019 Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency" ……..…… 1



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

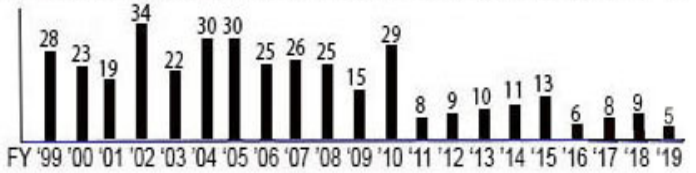
Measure 5. Journal Articles, Publications

Number of peer reviewed journal publications documenting research that appeared (19) or were accepted (8) in FY2019 …….... 27



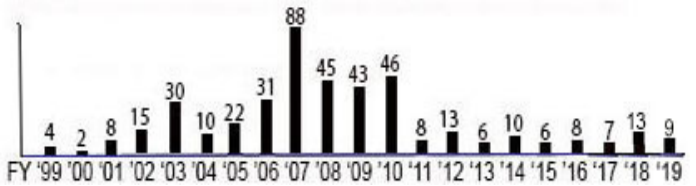
Measure 6. Proceedings, Publications

Number of proceedings publications documenting research that appeared in FY2019 ……..……. 5



Measure 7. Center Research Reports/Studies, Publications

Number of center research reports/studies publications documenting research that appeared in FY2019 ……..…… 9



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 ADDRESS CANVASSING IN OFFICE (DECENNIAL Project 6350E01)

1.2 ADVERTISING CAMPAIGN (Decennial Project 6450E20)

1.3 DATA CODING/EDITING/IMPUTATION (Decennial Project 6550E01)

1.4 REDISTRICING DATA PROGRAM (Decennial Project 6550E06)

1.5 CCM PLANNING & PROJECT MANAGEMENT (Decennial Project 6650E01)

1.6 2020 EVALUATIONS – PLANNING & PROJECT MANAGEMENT (Decennial Project 6650E20)

1.7 2030 PLANNING & PROJECT MANAGEMENT (Decennial Project 6650E25)

1.8 ADMINISTRATIVE RECORDS DATA (Decennial Project 6750E01)

A. 2020 Census Communications Campaign Statistical Analyses - I

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 2019, 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 sample 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 sample 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. The models show that adding lifestyle segments to the LRS improves the prediction of self-response rates at the tract level. The team submitted a manuscript that describe the results to a journal and completed a second revision of the paper. The team submitted the new version to the journal for further consideration.

Staff: Mary Mulry (682-305-8809)

A. 2020 Census Communications Campaign Statistical Analyses - II

Description: The U.S. Census Bureau is preparing to field the 2020 Census Communications Campaign to encourage participation in the 2020 Census. Similar campaigns aided in maintaining high self-response rates for the 2000 and 2010 Censuses. To prepare, the U.S. Census Bureau fielded the 2020 Census Barriers, Attitudes and Motivators Survey (CBAMS) to collect data on attitudes and knowledge about the U.S. Census. Data from over 17,000 respondents was used to classify individuals into one of six mindsets that could be used in developing messages to persuade individuals to respond. Our research examines the feasibility of assigning a mindset to each record in a third-party dataset containing over 250 million adult records and ultimately to households. The 2020 CBAMS variables used in determining the mindsets are not present on the third-party dataset although the-dataset does contain approximately 500 variables that reflect demographics, socioeconomic status, attitudes and behavior. Our approach links the 2020 CBAMS records to the third-party dataset and then uses multinomial logistic regression with independent variables from the third-party dataset to predict the probabilities of the mindsets.

Highlights: During FY 2019, staff collaborated with staff in the Research and Methodology Directorate and the Center for Behavioral Science Methods in research that is

innovative in that it explores the feasibility of developing a 'typing tool' that assigned one of six Census mindsets to households for the delivery of both mail and digital advertising to encourage response to the 2020 Census. The Census mindsets are being used in developing products for the 2020 Census Communications Campaign. The research demonstrated that linking the 2020 CBAMS survey records and the NHF (National Household File acquired from a third-party) records is feasible and produces enough linked individual records to conduct the analyses. The procedure to link records in the 2020 CBAMS survey records with the NHF records required multiple steps and had its challenges. The two types of records did not contain exactly the same identifying information. In addition, the research demonstrates an approach to variable reduction for fitting a six-level multinomial logistic regression model. The strategy involved fitting a logistic regression model for each pair of mindsets, which resulted in 15 models. The research also investigated how the proposed assignment of the mindset with the highest estimated probability to the household and found the approach somewhat problematic. These results appear in the "Variable Selection for multinomial logistic regression modeling to assign one of six Census mindsets using Big Data," which has been submitted to the *2019 JSM Proceedings.*

Staff: Mary Mulry (682-305-8809)

B. 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 2019, staff continued 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. Staff compared results where the tax year 2009 IRS 1040 household is based on returns received in weeks 4-17 of 2010 to results where the tax year 2009 IRS 1040 household is based on returns received in weeks 4-21 of 2010. Staff reran the programs after modifying one portion of the model. As before, results with and without the extra four weeks of IRS returns produce similar predicted values for corresponding IDs and similar population estimates for corresponding cutoffs. Staff summarized the results of the analysis above in an overview that was made available to the Administrative Records Modeling subteam. Full documentation of the results was added to the overview and the tables were rounded to better fit with Disclosure Review Board rules. The overview has started through the process to become a Decennial Statistical Studies Division internal memorandum. Staff also made minor revisions to the draft overview of an earlier comparison of modeling with and without the additional four weeks of IRS 1040 returns. The revisions resulted from minor modifications to some model runs which caused minor changes to some tables but no substantive changes to the analysis results. Staff revised the draft document (co-authored with Darcy Morris) that summarizes the results of the analysis of modeling with and without state data from specific programs targeted toward low-income households. The results with and without state data were generally similar. The revisions include additional results showing continued similarity in a slightly different context. A References section was added, citations were added to earlier sections, and assorted minor editorial changes were made. Results from the draft document were an input into the decision to not use state data in modeling for AR removals from nonresponse follow-up. The draft document has started through the process to become a DSSD internal memorandum. Staff wrote a draft overview of results of an exploratory analysis looking at field status of units identified as administrative records (AR) deletes in the evaluation sample of the end-to-end test. The evaluation sample sent non-response follow-up units through the full field process regardless of AR modeling results, allowing field status to be compared to the AR modeling results. The overview was provided to members of the AR modeling subteam. It found several variables that could be potentially useful in refining predictions of field status.

Staff: Michael Ikeda (x31756)

C. 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 2019, staff studied how well disclosure avoidance methodology of the 2010 Decennial data products worked by implementing a reconstruction and re-identification of person data from the 2010 Decennial Census using public and commercially purchased data. Results will be published in a forthcoming paper. Staff also contributed to a successful 2018 "End-to-End" test running the Disclosure Avoidance System (DAS) algorithm with test data. In addition, staff researched the feasibility and accuracy of proposed 2020 data products and continued to improve and scale the DAS algorithm.

Staff: Robert Ashmead (x31564), Brett Moran, Michael Ikeda, Ned Porter, Philip Leclerc (CED), Simson

Garfinkel (CED), William Sexton (CED), John Abowd (ADRM), Daniel Kifer (CED), Casey Blalock (CODS)

D. 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 2019, staff read literature on latent class models – specifically related to their use for multi-source data – and worked on latent class analysis of Census test housing unit status outcomes to provide further insight about discrepancies in field response and AR outcomes. Staff continued to attend and participated in meetings of 2018 test updates and 2020 production preparations.

Staff: Darcy Steeg Morris (x33989), Yves Thibaudeau

E. 2010 Congressional District Summary File for the 116th Congress
Description: The 2010 Congressional District Summary File contains data compiled from the 2010 Census about people and housing units. The file contains 333 subject tables at many different levels of geography, and the subject content of these tables is identical to the tables in the 2010 Census Summary File 1. The geographic content of the 2010 Congressional District Summary File focuses primarily on Congressional Districts within each State, and certain subdivisions of the Congressional Districts in a hierarchical sequence, for example (but not limited to), State-Congressional District-County-Census Tract. The 116th Congress convened on January 3, 2019, and redistricting has caused geographic changes in Congressional District boundaries between the 115th Congress and the 116th Congress. The Census Bureau plans to release a 2010 Congressional District Summary File for the 116th Congress that will contain data corresponding geographically to the Congressional Districts as they are in place for the 116th Congress. There are concerns that retabulating the 2010 Census data for new Congressional Districts may pose a disclosure risk because of the possibility of "slivered" geography. Because of disclosure risk concerns, staff were asked to investigate the extent to which the 2010 Congressional District Summary File for the 116th Congress can be produced based on data that are already publicly available.

Highlights: During FY 2019, staff examined the tables of the 2010 Congressional District Summary File. The 333 subject tables consist of 177 population tables labeled with prefix "P", 84 population tables labeled with prefix "PCT", 10 population tables labeled with prefix "PCO", 58 housing tables labeled with prefix "H", and 4 housing tables labeled with prefix "HCT". In the 2010 Census Summary File 1, the "P" and "H" tables are provided down to census block level, the "PCT" and "HCT" tables are provided down to census tract level, and the "PCO" tables are provided down to county level. Because the "P" and "H" tables are shown in 2010 Summary File 1 at census block level, while the "PCT", "HCT", and "PCO" tables are not shown down to the block level, it appears that only the "P" and "H" tables (235 subject tables each at many levels of geography) can be reliably produced within the given constraints. A proposal to produce a Congressional District Summary File for the 116th Congress consisting of only the "P" and "H" subject tables using systems that are already in place, and have been previously used and tested, is under review. We understand that the proposal was accepted, and this project is completed.

Staff: Martin Klein (x37856), Tommy Wright

F. Experiment for Effectiveness of Bilingual Training
Description: Training materials will be available for enumerators in the 2020 Census to communicate with non-English speaking households. Previously, such situations were left to the enumerator's discretion, and intended census messaging may not have been conveyed uniformly. The Census Bureau would like to measure the effect of this new training on response rate and other key metrics. The goal of this project is to prepare a statistical experiment to be embedded in the census, subject to operational constraints such as dynamic reassignment of cases and the potential for both trained and untrained enumerators to visit the same households.

Highlights: During FY 2019, staff considered models to support an experiment to measure the effectiveness of the bilingual training module. Staff pursued mixed effects models based on the continuation-ratio logit parameterization of the multinomial distribution to compare response rates by treatment (training vs. no training) and also to adjust for enumeration attempt (first attempt, second attempt, etc). A continuation-ratio logit model was recommended for data analysis, which can make use of dynamic case assignments once recorded in field operations. A simpler model was recommended for tractable sample size calculation to aid in experimental design. Staff worked out a method to study the power of a general linear hypothesis test with fixed effects only. Staff extended the method to mixed effects models using

several alternatives for approximate likelihood calculation. Staff carried out simulation studies to compare the alternatives, and began preparation of a manuscript based on the work.

Staff: Andrew Raim (x37894), Thomas Mathew, Kimberly Sellers, Renee Ellis (CBSM), Mikelyn Meyers (CBSM), Luke Larson (CBSM)

G. Unit-Level Modeling of Master Address File Adds and Deletes

Description: This line of research serves as part of the 2020 Census Evaluation Project on Reengineered Address Canvassing authored by Nancy Johnson. Its aim is to mine historical Master Address File (MAF) data with the overall goal of developing a unit-level predictive model by which existing MAF units may be added or deleted from the current status of live residential housing units for purpose of sampling (e.g., in the American Community Survey sampling universe) or decennial census coverage. There has never been such a predictive model at unit level, nor a concerted effort to mine historical unit-level MAF records for predictive information, and the search for such a unit-level model promises new insights for which MAF units outside the filtered HU universe are most likely to (re-)enter that universe, and also might suggest useful ways to decompose the MAF population in assessing the effectiveness of in-office canvassing procedures.

Highlights: During the first half of FY 2019, staff did background reading and consulted with Decennial Statistical Studies Division (DSSD) and Geography Division staff on MAF data; prepared SAS macros to preprocess ACS versions of MAF files; and began analysis in R of recoded variables using tree-structured machine learning techniques. During Quarter 3 of FY 2019, staff continued predictive analysis of MAF ACS-universe indicators in terms of recoded MAF history-variables, using tree-structured machine learning techniques as well as conventional logistic regression models. The findings up to this point are that, using the current MAF variables other than the ACS-universe variable itself, the algorithm used by DSSD for defining new ACS-universe adds can be reconstructed with reasonably high accuracy using the R tree-based recursive-partitioning package **rpart**. In genuine step-ahead prediction of ACS-universe adds, that is, of the rare event of existing addresses being newly recognized by DSSD at the next Delivery Sequence File updates as legitimate MAF ACS-universe addresses, the accuracy of correct prediction is low. There was no further progress on this project in Quarter 4.

The objective of these analyses is not so much to predict as to define subpopulations of MAF non-ACS addresses with unusually high rate of conversion to ACS addresses. Research on these techniques is continuing.

Staff: Eric Slud (x34991), Daniel Weinberg, Nancy Johnson (DSSD)

H. Decennial Record Linkage

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

Highlights: During FY 2019, staff met with a new group formed by our Center that is reviewing methods of record linkage and available software that has been developed within the Census Bureau. Staff received the CARRA PVS SAS matcher that has been used for ten years for administrative records matching.

Staff in our Center and the Center for Optimization and Data Science have used the FEBRL package to generate some test data.

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

I. 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 2019, staff has attended and participated in relevant specification walkthroughs in preparation to finalizing the specs for the Coverage Measurement operations for 2020. Staff have also been reviewing final versions of documents in preparation for the 2020 Census Coverage operations.

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

J. 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 2019, the In-Office Address Canvassing (IOAC) 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. Reviewers completed digital interactive review for all 11 million Census blocks. It is reported that IOAC will cut field workload by two-thirds.

Staff: Tom Petkunas (x33216)

K. A Deterministic Retabulation of Colorado and Pennsylvania Congressional District Profiles from the 115th Congress to the 116th Congress

Description: During the 115th Congress and before the start of the 116th Congress, congressional districts in Colorado and Pennsylvania were redrawn. Changes in the new boundaries for Colorado were extremely minor. The desire was to retabulate 116th Congress (2019-2021) congressional district profiles for the newly drawn seven districts of Colorado and the eighteen districts of Pennsylvania. The request was to produce these congressional district profiles without using the usual estimation methodology based on American Community Survey (ACS) sample microdata. This was to be done only using data that were already available to the public. For each congressional district, these profiles provide estimates of age distribution, median age, sex (Female/Male) distribution, number of Veterans 18 years and older, race distribution, proportion Hispanic/Latino, distribution of occupied/vacant housing units, and distribution of occupied housing units according to owner occupied and renter occupied.

Highlights: During FY 2019, a deterministic method was used which accounts for "movement" of population and housing units [and their data vintage of the 2015 ACS 1-year data grounded in 2010 Census/Summary File 1 (SF1) data] from old districts to the newly drawn districts in Pennsylvania. (Because changes in Colorado were extremely minor, we recommended that no new profiles be produced.) The methodology "adjusts" 2010 Census (SF1) population and housing units counts at the block level of old districts to reflect growth from 2010 to 2015 using 2015 ACS 1-year profiles, moves their blocks to the new districts, and retabulates distributions in profiles for new districts with these adjusted numbers of housing units and persons (population). A draft report is being prepared for review.

Staff: Tommy Wright (x31702), Martin Klein, Eric Slud

L. Assessing Variability of Data Treated by TopDown Algorithm for Redistricting

Description: Data from the most recent decennial censuses are used by the U.S. Department of Justice (DOJ) to clear some new proposed redistricting plans of U.S. House of Representatives congressional districts, as well as for some new state level districts. The objective of this study is to assess the variability of data results from the application of a disclosure avoidance randomization algorithm to 2010 Census Edited File Data (CEF) for Rhode Island.

Highlights: During FY 2019, staff met with and obtained background information from DOJ to determine the quantitative criteria it uses to clear any new proposed redistricting plan that it reviews. Our specific focus regarding variability was whether or not data results from a TopDown Algorithm for disclosure avoidance meet the needs of the DOJ when it performs analyses to clear a new proposed redistricting plan. Our approach has two parts: (1) to report observations on variability of results of 25 runs of the TopDown Algorithm for Rhode Island and (2) to report observations on variability between the results among the 25 runs of the TopDown Algorithm and the published 2010 Public Law 94-171 data for Rhode Island. While we used epsilon=0.25 throughout our study, we obtain results for other values of epsilon from 0.01 to 10 and observe that variability decreases as epsilon increases and that this decrease in variability seems to level off for values of epsilon greater than or equal to 3. We observe that variability in the TopDown Algorithm increases as we consider decreasing levels of geography: from congressional districts, to upper chamber districts, to lower chamber districts, and finally to tract level for Rhode Island data. The following draft report has been prepared: "Variability Assessment of Data Treated by the TopDown Algorithm for Redistricting."

Staff: Tommy Wright (x31702), Kyle Irimata

1.9 AMERICAN COMMUNITY SURVEY (ACS) (Decennial Project 6385E70)

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 2019, staff researched how the Cayley-Hamilton transformation could be used to parameterize CARMA processes, and completed calculations for some of the basic examples of sampled continuous-time processes.

Staff: Tucker McElroy (x33227; ADRM), Patrick Joyce

B. 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 [S ee

The Ranking Project: Methodology Development and Evaluation, Research Section under Project 0331000].

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

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

D. Voting Rights Section 203 Model Evaluation and Enhancements Towards 2021 Determinations
Description: Section 203 of the *Voting Rights Act (VRA)* 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 2019, staff reviewed and documented script files in order to better catalogue what was done for the 2016 determinations and prepare for the 2021 determinations. Staff prepared documentation for the migration of VRA project files to the new IRE server environment, and clarified the directory structure and purged necessary files in preparation for the 2021 *Voting Rights Act* project.

Staff: Carolina Franco (x39959), Eric Slud, Robert Ashmead, Mark Asiala (DSSD)

E. 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 2019, staff improved the spatial multivariate model to estimate the ACS special tabulation of young children in Minnesota. Staff has been able to incorporate finite mixture components in both the fix effects and spatial effects of the model. The open theoretical issues remain on how to select the number of mixture components which seem to vary by table. Staff compared two different methods to implement the choice of the number of mixture components in a Bayesian framework. Results were presented to the American Community Survey Office staff in the summer. Staff is currently preparing a manuscript to document the methodology.

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

1.10 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 2019, staff met with CPS collaborators to discuss new Joint Statistical Meetings research of CPS staff on Balanced Replication variance estimation, including techniques of matching to assess and improve CPS variance estimation for Non Self-Representing Primary Sampling Units (PSUs) and additional topics on Successive Difference Replication (SDR) for Self-Representing PSUs. In related longitudinal-methodology research concerning CPS data, staff completed a draft chapter (Thibaudeau, Y., Slud, E., and Cheng, Y., in press) for a forthcoming volume on Longitudinal Survey Methods on two-period model-based longitudinal analysis of multi-outcome survey data. The goal was to provide a summary of methodology extending the hybrid design- and model-based 2-outcome published methods of Thibaudeau, Slud and Gottschalck (2017) to three employment-status outcomes, and to describe fixed- and random-effect approaches to small area estimation (state-level employment gross flows) for CPS. In addition, staff (Slud and Thibaudeau) wrote and submitted a follow-on journal paper on fixed- and random-effect analyses of longitudinal CPS employment data for the two months of June and July 2017. Here again the goal was to develop hybrid design- and model-based methodology for small-cell gross flow estimates at the state level in CPS.

Regular research discussions continued with CPS staff. Final revisions were prepared for the paper of Slud and Thibaudeau (2019, "Multi-outcome Longitudinal Small Area Estimation -- A Case Study"), and it appeared online in the journal Statistical Theory and Related Fields.

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

B. Research on Biases in Successive Difference Replication Variance Estimation in Very Small Domains

Description: In various small-area estimation contexts at the Census Bureau, current methods rely on the design-based sample survey estimates of variance for survey-weighted totals, and in several major sample surveys including the American Community Survey (ACS) and Current Population Survey. These variance estimates are made using Successive Difference Replication (SDR). One important application of such variance estimates based on ACS is the *Voting Rights Act (VRA)* small-area estimation project supporting the Census Bureau determinations of jurisdictions mandated under VRA Section 203(b) to provide voting language assistance. The current research project is a simulation-based study of the degree of SDR variance-estimation bias seen in domains of various sizes.

Highlights: During FY 2019, staff coded and tested R functions to conduct the simulation and wrote a preliminary draft summarizing SDR background and methods. Staff resumed discussions concerning the design of simulation studies for SDR biases in small-domain variance estimation. The systematic simulation runs and tabulation of results to answer research questions have yet to be performed.

Staff: Eric Slud (x34991), Robert Ashmead, Tim Trudell (DSMD)

1.11 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 2019, staff ported software for cleaning matching and standardizing files to the IRE. Staff developed software for analyzing 1940 Census data that had formal disclosure avoidance methods applied. Staff developed a novel approach to find false matches in a Census Duplicate Person Identification and submitted for a public presentation.

Staff: Ned Porter (x31798)

1.12 SOCIAL, ECONOMIC, AND HOUSING STATISTICS DIVISION SMALL AREA ESTIMATION PROJECTS (Demographic Project 7165018/7165019)

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 2019, staff continued comparison of alternative small area estimation models for estimation of county poverty rates of school-aged children in poverty using the artificial population. Staff compared the Binomial Logit Normal Model to a Fay-Herriot (FH) model on rates and to a log-transformed FH model similar to that used in production, using design-based estimates of sampling variances. Staff implemented the GVF described in Franco and Bell (2013) in artificial population and compared its performance to the direct estimates of sampling variances, as well as their performance when used to estimate effective sample sizes. Staff also implemented models above using the GVF estimates of sampling variances and effective sample sizes, and compared their performance. Staff compared how the results differ depending on which estimate of sampling variance is used for fitting the model, and to the (artificial) case where sampling variances are known, to understand the impact of estimating sampling variances on the resulting prediction mean squared errors. Staff wrote a technical report on this which is currently undergoing internal review.

Staff also began implementation of the Dirichlet-Multinomial Share models to model the total pop and poverty of school-aged children in school districts. The model uses the 5 years of Federal Tax data (2008-2012) to form the predictors and the 5-year 2012 ACS data to form the direct estimates of school district piece to county shares. The method employs a pair of Dirichlet-

Multinomial Share models to model the share of school-aged children in poverty and not-in-poverty. The estimated shares and the county estimates of school-aged children in poverty and total number of school-aged children are combined to form the count estimates and their measure of uncertainty. The predictions from the model showed a median reduction of 40% and 63% for school district poverty and population estimates compared to the direct survey estimates only using the ACS data. This work was presented 2019 Joint Statistical Meetings in Denver, CO and at a CSRM seminar in September 2019.

Staff also explored several bivariate models without covariates using SAIPE county-level data on school-aged children in poverty, as well as model diagnostic and comparison tools to aid in the selection of univariate and bivariate models with very different assumptions. Staff wrote a paper including many of the results and presented them in two conferences, as well as in a seminar in Madrid, Spain and a CSRM seminar.

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.

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.

Highlights: During FY 2019, staff performed a comprehensive literature review of unit-level small area modeling strategies for situations in which sample survey data are collected under an informative sampling scheme. A selection of these approaches were compared by fitting the models on real American Community Survey (ACS) data, as well as on simulated data, which was drawn using an informative survey design from a subset of ACS data. A review paper, which includes the literature review, as well as the comparisons of different methods, was written and published to the ArXiv. Staff also worked on new unit-level small area modeling strategies by developing design-weighted Bayesian pseudo-likelihood models, which can be used with non-Gaussian survey data. A paper documenting this work was completed and submitted for journal publication.

Staff: Ryan Janicki (x35725)

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

Highlights: During FY 2019, staff has worked on developing a joint model for both the number of relevant school-aged children in a school district and the number of relevant school-aged children in poverty in a school district. This pair of estimates is difficult to construct appropriate models for because both the total number of children and poor children must be calibrated to the estimated county values, but also there cannot be any district where the number of poor children exceeds the total number of children. Staff developed an approach based on two independent share models to estimate the school district to county shares of both the poor and non-poor school age children. Each of the share models are based on a modified Dirichlet-Multinomial distribution. Two models, one for the share of the children in poverty and one for the share of children not in poverty, are fitted independently. The population estimate comes from the sum of the two subgroups. Staff implemented this model in R and fitted this model to the 2012 5-year ACS and 2008-2012 Federal Tax Data. Staff presented results at the Joint Statistical Meetings (poster presentation and proceedings paper) and in a CSRM seminar.

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

1.13 GENERAL ECONOMIC STATISTICAL SUPPORT (Economic Project 1183X01)

A. Research on Imputation Methodology for the Monthly Wholesale Trade Survey
Description: In the previous phase of this project, staff conducted a simulation study to investigate new imputation methodology for the Monthly Wholesale Trade Survey (MWTS). In this phase of the project, staff are creating a more realistic simulated wholesale trade population and investigating improved MWTS estimators. The MWTS is a longitudinal sample survey that provides month-to-month information on sales and inventories of U.S. merchant wholesalers. Key

estimates produced from this sample survey include total sales, month-to-month relative change in sales, total inventories, and month-to-month relative change in inventories (overall and within industry subclasses). There are a number of challenges when developing estimators for the MWTS, including variables with highly skewed distributions, missing values in predictor variables from the Economic Census, and sample survey variables with trends that differ across industry classes. The longitudinal information in addition to a rich set of frame data available from the Economic Census can be used to build Bayesian models that address these challenges. It is expected that this model will be applicable to other business sample surveys.

Highlights: During FY 2019, staff began preparing a manuscript to describe the methods used to construct the realistic, artificial population that was developed as a tool for evaluating the performance of statistical methodology applied to the Monthly Wholesale Trade Survey (MWTS) or other similar surveys. Due to staffing limitations, work was suspended.

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

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: During FY2019, staff finished working on exploratory data analysis of a new delivery of First Data monthly retail sales data based on raw aggregates, specifically investigating suppression rates and measures of coverage. Staff worked with statisticians and subject matter experts in the Economic Directorate to understand trends in data availability by industry, state and region and determined industries and geographies with acceptable quality data for producing experimental estimates. An adaptation of the final report was submitted and published in the *Journal of Big Data*. This paper, "Using Electronic Transaction Data to Add Geographic Granularity to Official Estimates of Retail Sales," describes industry-level "quality profiles" and the methodology for producing geographically granular retail sales experimental estimates for select industries. Staff also presented this work at a joint Center for Statistical Research and Methodology and Economic Statistical Methods Division Seminar.

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

C. 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 2019, staff provided seasonal adjustment and software support for users within and outside the Census Bureau, and continued drafting recommendations on assessing residual seasonality in economic time series.

Specific seasonal adjustment and software support for users within and outside the Census Bureau included: RCF Economic & Financial Consulting, Irish Central Statistics Office, Statistics Norway, UC Berkeley, Stats Canada, Statistics Centre Abu Dhabi, Beijing, JP Morgan, Shandong University of Science and Technology, Woodstat, Bureau of Labor Statistics, World Bank Group, Informed Portfolio Management (Sweden), Statistical Office of Argentina.

Staff members continued to participate in an inter-agency group (ISAT) to study and make recommendations on mitigating residual seasonality in estimates of GDP, which has developed into a discussion group for research into seasonality diagnostics and residual seasonality.

Staff: Tucker McElroy (x33227; ADRM), James Livsey, Osbert Pang, William R. Bell (ADRM)

D. 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 2019, and in collaboration with the Center for Optimization and Data Science, staff released a debugged version of X-13ARIMA-SEATS, planned the architecture for a new software based on RegCMPNT, and added new functionality to Ecce Signum.

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

E. 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 2019, staff made progress on several research projects: (a) assessed the impact of weather regressors on seasonal adjustment; (b) modeling and analysis of daily time series (New Zealand immigration data and credit card transaction data); (c) new results on seasonality diagnostics based upon autoregressive roots; (d) analysis of weekly factors for improved monthly seasonal adjustment; (e) new results on seasonal vector form, useful for understanding the properties of down-sampled time series, which assists the analysis of frequency aggregation; (f) continued modeling and analysis on calendarization; (g) developed and tested benchmarking optimization methodology for improved indirect seasonal adjustment of quarterly time series; (h) examined the presence of residual seasonality in large components of GDP using seasonality diagnostics; (i) further explored the use of the EM algorithm in conjunction with signal extraction methods to fit multivariate time series models.

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

F. 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 2019, staff continued writing a paper that documents currently available seasonal adjustment diagnostics, as part of work with the Bureau of Economic Analysis (BEA) on residual seasonality in GDP. Staff developed documentation of X-13ARIMA-SEATS source code, and presented code walks for interested employees. Staff began crafting documentation for Ecce Signum.

Staff: Tucker McElroy (x33227; ADRM), James Livsey, Osbert Pang, William R. Bell (ADRM)

G. Redesign of Economic Sample Surveys (Stratification)

Description: Following the recommendations of a National Academy of Sciences panel, work was begun to redesign the economic sample surveys into a common sampling and estimation system. This process seeks to take several economic sample surveys and reformulate the surveys as part of a singular sampling and estimation operation. Separate research teams were formed with different research tasks. The work of this project involves a single research team focused upon the construction of a stratification methodology for the common economic survey execution.

Highlights: During FY 2019, background information was compiled on the stratification methods of the various economic sample surveys within the scope of the redesign. Literature in the area of high-tailed and certainty-sampling methods were collected and reviewed to determine the current state of academic knowledge. Notions of stratification with regards to multiple variables as well as discussion of machine

learning techniques were discussed. Emphasis was also placed on the discussion of the notion of respondent burden reduction and that notion's relationship with sampling and estimation methods.

Staff: Patrick Joyce (x36793), Eric Slud

1.14 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: During FY 2019, we planned for the migration of research1 and research2 projects to the Integrated Research Environment (IRE). Chad Russell (CSRM) gave a presentation on the overall migration strategy to users explaining the IRE data access model and how to prepare for the eventual migration by: 1) deleting or archiving unnecessary files, 2) moving files from "personal" to "project-based" directories, and 3) ensuring the existence and correctness of a DMS record for each project to be migrated. Chad Russell and David White (CODS) created a "mapping" which mapped everything that needs to be migrated from research1 and research2 to the appropriate target workspaces on IRE. The first batch of projects was migrated in late September, and each week a batch of projects is migrated, with all of the projects expected to be migrated by the end of FY 2019.

Work continued on the Cloud Research Environment (CRE) prototype. AWS GovCloud instances, both Linux and Windows, have been set up with the desired software for research computing. A graphical desktop interface is provided via NoMachine Enterprise Terminal Server (ETS) along with standard Linux and Windows tools. Currently each compute server has elastic block storage (EBS) volumes which hold read-only "source" data and writable "project" data. These EBS volumes are initially populated from S3 storage. One problem with the current scheme is managing multiple copies of the data while ensuring consistency. A shared filesystem would alleviate some of these problems. The Computer Services Division (CSVD) has developed a Network File System (NFS) solution to address this issue for the time being. Much of the CRE work in FY 2019 was focused on obtaining the required "Authorization to Operate" (ATO) from the Office of Information Security (OIS). The team continues to prepare for the ATO assessment which is scheduled for December.

Staff: Chad Russell (x33215)

1.15 BUREAU OF ECONOMIC ANALYLSIS (Census Bureau Project TBA)

A. Business Cycle Movements in National Accounts Series
Description: Staff collaborated with researchers from the Bureau of Economic Analysis (BEA) on this project, which aims to analyze business dynamics revealed in Real Gross Domestic Product (GDP) and its major components. It provides an updated set of empirical regularities that are based on recent data and use an adaptive approach that accounts for series' different properties.

Highlights: During FY 2019, staff met regularly with BEA staff to develop a new set of so-called "stylized facts," meant to summarize cyclical properties and inter-relationships among GDP and major components such as Investment and Consumption of Durables. One finding has been the different dynamic characteristics of GDP and various components. Adaptive methods were used to avoid spuriousness in the resulting extracted cycles. Much previous work makes use of fixed, non-parametric filters that involve serious risk of distortions in the output and hence statistics and quantities computed on the basis of such output.

In particular, cross-correlation plots were analyzed that show correlations between cycles and different leads and lags. Additionally, an extensive set of Granger causality tests were performed that quantitatively examine the temporal dependences among the GDP components. Significant differences in results were obtained in several cases for the widely used Hodrick-Prescott and Baxter-King filters.

Staff: Thomas Trimbur (x36864), Baoline Chen (BEA)

1.16 FEDERAL HIGHWAY ADMINISTRATION (Census Bureau Project TBA)

A. Modeling and Signal Extraction of Pedestrian and Bicycle Crash Data
Description: The Census Bureau provided statistical expertise to the Federal Highway Administration (FHWA). The research involves modeling and signal extraction of pedestrian and bicycle crash data. In this

event level crash data, trends and cycles are obscured by the large amount of seasonality present as well as the sparsity of events at fine levels of geography and time. This collaboration uses Census Bureau demographic data as aggregation factors to amplify the signals then considers signal extraction methodology to account to seasonality.

Highlights: During FY 2019, staff and FHWA worked together to construct a homogenized dataset of all pedestrian and bicycle crashes, along with 117 covariates, for the entire state of North Carolina. The time series staff provided FHWA multiple R functions that aggregate and model differing levels of geography and time. A primary finding was decreasing crash occurrences in higher income counties. This finding along with signal extraction methodology was presented to 40 States' Safety Engineers at The Evaluation of Low Cost Safety Countermeasures Pooled Fund Study (ELCSI-PFS) 15th annual technical advisory committee meeting.

Staff: James Livsey (x33517), Roya Amjadi (FHWA)

1.17 NATIONAL CANCER INSTITUTE

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. We were asked to help provide the county level data for NCI.

Highlights: During FY 2019, final results of the TUS-CPS Small Area Estimation project were released to the sponsor at National Cancer Institute after a comprehensive disclosure review process. A research paper documenting work on this project and poster section presented in Minneapolis, Minnesota on Conference on Tobacco Health is near completion. This project has been completed.

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

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 2019, staff collaborated in resolving difficulties using non-parametric Bayesian methods developed by Kim et al. (2017) to edit, impute and synthetize economic census data. To resolve these issues, Kim implemented updates to his software tailored to Economic Census data for a Census Bureau version of the software. The research team edited, multiply imputed, and synthesized data for sample industries in seven separate trade areas using the updated Census Bureau version. The project team is working on evaluating the risk/utility tradeoff and effectiveness of the synthetic data developed using this methodology. Staff collaborated with other team members to write a user's guide for the Census Bureau's version of the software. Staff reviewed reports on research and evaluation of administrative records to impute ACS demographic items and work on selective editing for periodic, continuous economic data.

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. Some focus on record linkage has been added.

Highlights: During FY 2019, we tracked the ecosystem of BigMatch. BigMatch was written by talented programmers in the Statistical Research Division in the late 1990's and early 2000's. Unfortunately BigMatch never was integrated into a formal software development lifecycle. It was entirely homemade and it appears there never were specifications written. The programming decisions are not documented, let alone justified. We showcased BigMatch and provided metrics of its performance along with the SAS (PVS) Matcher. Both matchers performed well using the FEBRL file simulator. BigMatch was more stable and progressive in reaching pre-determined coverage as a function of precision. The SAS matcher was reaching slightly better coverage at given cutoffs along the curve vs. precision, but was also more volatile, meaning eliciting a wrong cutoff can lead to more substantial loss of coverage. The methodology of BigMatch is of interest and the possibility of integrating the BigMatch methodology as part of a comprehensive record-linkage platform at the Census Bureau is being examined.

In an effort in developing comprehensive log-linear methodology for the the multitude applications at the Census Bureau, staff researched and implemented specific parameterization of hierarchical log-linear model designed to track the condition of the parameter space under multidimensional contingency tables in the presence of general patterns of zeros. The explicit parameterization track the various collapsing patterns of the parameter space, in particular collapsing patterns for the conditional probability. This is based on an asymmetric parameterization, which parse the contingency variables as "conditioning" and

"conditioned, as opposed to the symmetric parameterization of Fienberg and Rinaldo (12). This approach makes it simple to track the collapsing patterns and to derive the residual models based on the collapsed space.

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 2019, we continued collaboration with an interdivisional team researching the implementation of a nonparametric Bayesian hierarchical model proposed by Kim et al. (2017) to edit, multiply impute, and generate synthetic, industry-level, economic census microdata, that can be shared with the public in place of suppressed estimates. Each synthetic dataset is a replicate from a given posterior predictive distribution; the posterior distribution is the distribution of the parameters given the multiply imputed data obtained in the editing/imputation phase of this research. Team members used the methodology to developed multiply-imputed synthetic data for seven separate economic census trade areas. The team researched methods to evaluate the synthetic data, developed generalized code for estimation and variance estimation, and developed generalized code for calculating quality metrics and multivariate comparisons. The team calculated quality metrics and measures for variances using Rubin's multiple imputation formulae, confidence intervals, correlations, comparison graphics, etc. We continue researching measures to estimate attribute disclosure risk for economic census synthetic data including adaptation of methodology developed by Hu, Savitzky, and Williams (2018). Staff collaborated in writing a User Guide for Dr. Kim's Census Bureau's package "EditImputeEconCensus," specifically designed for editing, imputing, and synthesizing Economic Census microdata. The User Guide describes requirements identified by the research team that are specific to running the package with Census Bureau continuous economic data that must satisfied ratio and balance edits.

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. Linear Regression with Sparsely Mismatched Data
Description: Statistical analysis with linked data may suffer from an additional source of non-sampling error that is due to linkage error. For example, when predictive models are of interest, in the linkage process, the response variable and the predictors may be mismatched or systematically excluded from the sample. In this research, we focus on the cases where responses reside in one file and predictors reside in another file. These variables are then paired up using an error-prone record linkage process. We nevertheless assume that only a small fraction of these pairs is mismatched. The goal of the research is then to develop efficient methodologies for adjusting the statistical analyses for bias or inconsistency introduced by linkage error.

Highlights: During FY 2019, the staff worked on two different approaches for linear regression with linked data. In the first project, we consider a multi-response case, in which we propose a two-stage method. In the first stage,

we estimate the regression parameters and subsequently we identify the mismatches. In the second project, we propose a pseudo-likelihood method with a two-component mixture density. We use Expectation-Maximization schemes for estimating the parameters. An advantage of the later proposal is that it tolerates larger fraction of mismatches than that of other approaches. It also provides estimates of the noise level as well as the fraction of mismatches.

Staff: Emanuel Ben-David (x37275)

B. 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: During FY 2019, staff noted two main pieces of work for scaling prior work from Steorts (2015), which are available on CRAN and known as blink. First, 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. This paper is still under review. Second, in a recent piece of work, we are working on extending Steorts (2015) to have more flexible prior distributions, namely using non-parameteric distributions on the linkage structure. We are investigating this under synthetic and conflict data. In addition, we are building a package that will replace blink on CRAN for more scalable computing.

Workshops/training have been presented to staff at the Census Bureau, and testing has started to see if dblink can be utilized in house for the 2020 Census.

Staff: Rebecca C. Steorts (919-485-9415), David Brown (CES)

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. 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 (NHIS), the Survey of Income and Program Participation, and the Current Population Survey.

Highlights: During FY 2019, staff continued analysis of application of health insurance non-coverage, where ACS state estimates are used to improve NHIS state estimates. Staff also analyzed two other examples, that of using ACS state estimates to improve the estimates of disability from SIPP, and that of using the ACS five year county estimates to improve the corresponding one-year estimates, using the variable of school-aged children in poverty as an example. Implemented and compared several bivariate and univariate models. Analyzed the benefits of using a bivariate model or univariate model rather than the direct estimates. Showed in all three applications that impressive reductions in MSEs/posterior variances can be achieved by using bivariate models rather than the direct estimators or simple univariate shrinkage models. Studied model comparison tools. Created contour plots that illustrate the theoretical decrease from using bivariate models rather than univariate models under different circumstances. The analytical results were consistent with the empirical results. Wrote a paper on the subject, and gave two presentations at conferences and two seminars.

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

B. Bootstrap Mean Squared Error Estimation for Small Area Means under Non-normal Random Effects

Description: The empirical best linear unbiased predictor (EBLUP) is often used to produce small area estimates under the assumption of normality of the random effects. The exact mean squared error (MSE) for these approaches are unavailable, and thus must also be approximated. Staff will explore the use of estimating equations to obtain estimates of model parameters and the use of asymptotic expressions with a nonparametric bootstrap method to approximate the MSEs.

Highlights: During FY 2019, staff expanded the theory behind the moment based MSE estimation approach. Staff developed a bootstrapping method for approximating MSE with moment based estimators for the model parameters. One important feature of this work was the evaluation of various approaches for truncating random effects variance to ensure positive estimates. Staff conducted simulation studies to compare the approach to existing approaches, such as those proposed by Prasad and Rao and by Fay and Herriot. Staff saw that the bootstrap approach performed comparatively better in the case of large sampling variance, relative to the random effects variance.

Staff: Gauri Datta (x33426), Jerry Maples, Kyle Irimata, Eric Slud

C. Random Coefficient Extension to the Fay-Herriot Model

Description: In the standard Fay-Herriot Model it is assumed that the regression relationship is the same across all areas for each predictor covariate. In model checking and validation, this assumption can sometimes be shown to be false. A few causes for this can be a missing interaction term or an unknown clustering effect between the areas. We propose extending the Fay-Herriot model to include random coefficients where some of the regression effects can vary between the areas. This project is motivated by model evaluations done with the SAIPE county production model where differential effects of tax data has been detected for various subsets of the data.

Highlights: During FY 2019, staff finalized the proof for the propriety of the posterior distribution for the proposed prior and implemented the MCMC method in R. Staff conducted a simulation study showing the robustness of the random coefficient model when the true model was the standard Fay-Herriot. The model is being applied to the 1989 state median household income for 4-person households. The predictor variable is the state median household income for 3-person households. Staff wrote and submitted the manuscript titled "Estimation of Median Incomes for American states: Bayesian Estimation of Means of Subpopulations" and was accepted for publication as a book chapter in *Opportunities and Challenges in Development: Essay for Sarmila Banerjee.* This project was completed.

Staff: Gauri Datta (x33426), Jerry Maples

Survey Sampling-Estimation and Modeling

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

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

The Decennial Census is such a massive undertaking that

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

Research Problems:

• How can methods making additional use of administrative records, such as model-assisted and balanced sampling, be used to increase the efficiency of household surveys?

• Can non-traditional design methods such as adaptive sampling be used to improve estimation for rare characteristics and populations?

• How can time series and spatial methods be used to improve ACS estimates or explain patterns in the data?

• Can generalized weighting methods be implemented via optimization procedures that allow better understanding of how the various steps relate to each other?

• Some unusual outlying responses in the surveys of retail trade and government employment are confirmed to be accurate, but can have an undesired large effect on the estimates - especially estimates of change. Procedures for detecting and addressing these influential values are being extended and examined through simulation to measure their effect on the estimates, and to determine how any such adjustment best conforms with the overall system of estimation (monthly and annual) and benchmarking.

• What models aid in assessing the combined effect of all the sources of estimable sampling and nonsampling error on the estimates of population size?

• How can administrative records improve census coverage measurement, and how can census coverage measurement data improve applications of administrative records?

• What analyses will inform the development of census communications to encourage census response?

• How should a national computer matching system for the Decennial Census be designed in order to find the best balance between the conflicting goals of maximizing the detection of true duplicates and minimizing coincidental matches? How does the balance between these goals shift when modifying the system for use in other applications?

• What can we say about the additional information that could have been obtained if deleted census persons and housing units had been part of the Census Coverage Measurement (CCM) Survey?

Potential Applications:

• Improve estimates and reduce costs for household surveys via the introduction of additional design and estimation procedures.

• Produce improved ACS small area estimates through the use of time series and spatial methods.

• Apply the same weighting software to various surveys.

• New procedures for identifying and addressing influential values in the monthly trade surveys could provide statistical support for making changes to weights or reported values that produce more accurate estimates of month-to-month change and monthly level. The same is true for influential values in surveys of government employment.

• Provide a synthesis of the effect of nonsampling errors on estimates of net census coverage error, erroneous enumerations, and omissions and identify the types of nonsampling errors that have the greatest effects.

• Describe the uncertainty in estimates of foreign-born immigration based on American Community Survey (ACS) used by Demographic Analysis (DA) and the Postcensal Estimates Program (PEP) to form estimates of population size.

• Improve the estimates of census coverage error.

• Improve the mail response rate in censuses and thereby reduce the cost.

• Help reduce census errors by aiding in the detection and removal of census duplicates.

• Provide information useful for the evaluation of census quality.

• Provide a computer matching system that can be used with appropriate modifications for both the Decennial Census and several Decennial-related evaluations.

A. Household Survey Design and Estimation

[See Demographic Projects]

B. Sampling and Estimation Methodology: Economic Surveys

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

Highlights: During FY 2019, staff's collaboration with staff in the Economic Directorate has been on hold. The team's work has focused on the development of methodology for the treatment of influential observations in the estimation of sales and inventories for the Monthly Retail Trade Survey (MRTS) and Monthly Wholesale Trade Survey (MWTS). The team previously developed methodology for parameters settings for the M-estimation algorithm and prepared software for a side-by-side test with MWTS as it is collected. The team has been awaiting the creation of files with the data collected in a side-by-side test with MWTS that ran from the summer of 2018 through December 2018.

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: During FY 2019, staff extended and updated the methodology. Mostly notably, the method was shown to be easily adaptable to a Bayesian setting. Staff updated a draft paper; and is preparing to re-submit.

Staff: Tommy Wright (x31702), Martin Klein (FDA), 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.

Sample Allocation

Highlights: During FY 2019, staff provided an elementary derivation of Kadane's dynamic sampling plan by first directly finding the sample allocation that minimizes a decomposed weighted objective function with costs. We then prove that the sample allocation also minimizes the sampling variance. Cost methodology was extended to the case when there are minimum sample size constraints on the strata. Software was shared with the public.

Staff: Tommy Wright (x31702), Andrew Raim

Apportionment

Highlights: During FY 2019, staff developed a clear and explicit decomposition of an objective function which leads directly to the current method for apportioning the U.S. House of Representatives, as well as many others with varying constraints on number of seats possible for each state.

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: During FY 2019, staff revised and improved on the methodology and analysis used to fit daily response propensity modeling. Staff has revised a formal report of the analysis and estimation documenting work on NCVS daily response propensity modeling and methodology.

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

F. Consistent Estimation of Mixed-Effect Superpopulation-Model Parameters in Complex Surveys with Informative Sampling

Description: This research studies the problem of design-consistent model-assisted estimation for regression and variance-component parameters within parametric models based on complex survey data. Starting from seminal work of Binder (1983) on `pseudo-likelihood', it has been known how to design- and model- consistent inference from survey data, based only on observed data and single-inclusion weights, when units are independent under the superpopulation model. However, it has largely been an open problem since first studied in papers of Pfeffermann et al. (1998), Korn and Graubard (2003) and Rabe-Hesketh and Skrondal (2006), how – or if it is even possible -- to do consistent survey-weighted inference based on single-inclusion weighted survey data when

data share random effects within clusters and sampling may be informative.

Highlights: In the present research, conducted throughout FY 2019, an EM method was developed based on augmented-data pseudo-likelihood (where `augmented' means that the loglikelihood is for observable unit data plus unobservable random effects) and argued to be consistent when cluster-sampling may be informative but sampling within clusters is assumed noninformative. Research on this topic is continuing in FY2020.

Staff: Eric Slud (x34991)

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 2019, staff made progress on several research projects: (a) 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 by implementing an improved seasonality diagnostic; (b) continued work on signal extraction diagnostics methodology, including new diagnostics for seasonal adjustment based on autoregressive roots. Simulation studies were completed and theory developed, with extensions to the case of seasonal unit roots and the problem of over-adjustment; (c) 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) 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) continued writing code and text for a book on multivariate real-time seasonal adjustment and forecasting. Recent work includes extensions to co-integrated processes and mixed-frequency data; (f) 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.

Staff: Tucker McElroy (x33227; ADRM), James Livsey, 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 2019, staff made progress on several projects: (a) continued algorithmic work on computing Gaussian orthant probabilities, with applications to multivariate analysis; (b) continued additional simulation work for a Frobenius norm methodology for fitting and comparing multivariate time series models in frequency domain. The project gives a new distribution theory for a broad range of statistics for vector time series; (c) completed simulation work to model the behavior of forecasters over multiple horizons; (d) 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; (e) obtained new results for computing and estimating multivariate inverse autocovariances, which is useful for fitting vector moving averages; (f) continued research into methods for count time series, including multivariate modeling that allows for sparse parameterizations; (g) continued research into new models for business cycles.

Staff: Tucker McElroy (x33227; ADRM), James Livsey, Osbert Pang, Anindya Roy, Thomas Trimbur

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: During FY 2019, staff revised and resubmitted the manuscript for review with a journal; the manuscript currently remains under review.

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 2019.

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

B. 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 2019, (1) Staff published a research manuscript in the *Journal of Statistical Computation and Simulation* that developed 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. (2) Staff continued to derive the form of a multivariate CMP distribution via the compounding method. Various theoretical properties have been determined. (3) Staff derived a generalization of the multinomial distribution extended via the CMP distribution, thus establishing a Conway-Maxwell-multinomial (CMM) distribution. Staff derived properties, studied estimation procedures and applied the distribution to clustered categorical data. Staff submitted a JSM 2018 proceedings paper on this research titled "Introducing a Conway-Maxwell-Multinomial Distribution for Flexible Modeling of Categorical Data," presented a poster at JSM 2019 -- "Fitting Flexible Count Data Distributions Derived from the COM-Poisson Distribution" – describing the distribution and its computational challenges, presented this research in a CSRM seminar titled "Conway-Maxwell Distributions for Flexible Modeling of Count Data", and submitted a manuscript for internal review in preparation to send to a peer-reviewed journal. (4) Staff presented a poster at the *Conference on Statistical Practice*, and continued working on a paper and associated programming to incorporate normal- and conjugate-distributed random effects in a clustered/longitudinal COM-Poisson model.

Staff: Kimberly Sellers (x39808), Darcy Steeg Morris, Andrew Raim

C. 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 2019, staff revised the stcos R package to expose useful functions (such as computation of areal basis functions and overlap matrices) directly to the user; this potentially supports use of the package outside of the Bradley, Wikle, and Holan (2015) model. The package was submitted to CRAN for public use. Staff completed the accompanying article and submitted to a journal for peer review.

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

Simulation and Statistical Modeling

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

Research Problems:

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

Potential Applications:

- Simulating data collection operations using Monte Carlo techniques can help the Census Bureau make more efficient changes.
- Use noise multiplication or synthetic data as an alternative to top coding for statistical disclosure control in publicly released data. Both noise multiplication and synthetic data have the potential to preserve more information in the released data over top coding.
- Rigorous statistical disclosure control methods allow for the release of new microdata products.
- *Tea* provides modeling and editing flexibility,

especially with a focus on incorporating administrative data.

• Using an environment for simulating complex surveys, statistical properties of new methods for missing data imputation, model-based estimation, small area estimation, etc. can be evaluated.

• Model-based estimation procedures enable efficient use of auxiliary information (for example, Economic Census information in business surveys), and can be applied in situations where variables are highly skewed and sample sizes are not sufficiently large to justify normal approximations. These methods may also be applicable to analyze data arising from a mechanism other than random sampling.

• Variance estimates and confidence intervals in complex surveys can be obtained via the bootstrap.

• Modeling approaches with administrative records can help enhance the information obtained from various sample surveys.

A. Development and Evaluation of Methodology for Statistical Disclosure Control

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

Highlights, Subproject A.1 (Nayak): During FY 2019, staff conducted research on randomized response methods for protecting respondent's privacy and data confidentiality. Staff published a paper titled "A Criterion for Privacy Protection in Data Collection and Its Attainment via Randomized Response Procedures" in *Electronic Journal of Statistics.* This paper proposes a rigorous privacy protection criterion, derives all privacy preserving methods and compare their data utility. Staff completed a project on deriving optimal (minimax) methods and reported the findings in a technical report titled "Minimax Randomized Response Methods for Providing Local Differential Privacy." In another technical report titled "A Local l-Diversity Mechanism for Privacy Protected Categorical Data Collection," staff developed a new randomized response scheme that is easy to use and offers a better trade-off between privacy protection and statistical efficiency than an optimal procedure under local differential privacy. Staff developed a new method for controlling identification risks in microdata release. It guarantees that any unique match in released data, after perturbation, would be a correct match only with a small probability. Staff is working on writing a paper based on this research.

Highlights, Subproject A.2 (Klein): During FY 2019, staff continued work on the development of methodology for multiple imputation based parametric inference under a differentially private Laplace mechanism. Staff consider the scenario where continuous microdata are noise infused using a differentially private Laplace mechanism for the purpose of statistical disclosure control. It is assumed that the original data are independent and identically distributed, having distribution within a parametric family of continuous distributions. A modification of the standard Laplace mechanism is employed that uses the notion of truncation to allow differential privacy to be attained even if the range of the original data is unbounded. Under this scenario, staff developed methodology to impute the original data, based on the noise infused data, using the framework of multiple imputation for missing data. This approach allows the data user to analyze the released multiply imputed data as if it were original, i.e., not noise infused, and then to obtain inference that accounts for the noise infusion mechanism using standard multiple imputation combining formulas. Staff derived the methodology for univariate data, and some simulation studies were implemented to evaluate the performance of the proposed methodology. Staff also derived an extension of the proposed methodology for multivariate data. A manuscript describing this work is under preparation.

Staff worked on developing methodology for inference based on singly imputed synthetic data generated under plug-in sampling when the original data are distributed as multivariate normal. Based on the observed synthetic dataset, staff derived a statistical test for the generalized variance, a test for sphericity, a test for independence between two subsets of variables, and a test for the regression of one set of variables on the other. The procedures are based on finite sample theory. Some simulation studies were conducted to confirm that the proposed procedures perform as expected. A manuscript describing this work is under preparation.

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

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

Description: Staff is currently using general research methodology to work on a simulation study to describe how to produce a generalized boosted regression modeling algorithm for estimating propensity scores with bootstrap and continuous treatment methods. For this simulation study, staff is looking into how to estimate the generalized propensity scores for causal inference, potential outcomes and treatment outcomes for this research. As part of this research, staff continues to conduct comprehensive model diagnostics to confirm

that the simulation data generated using parametric distributions meet the normality assumptions. Three or more simulation scenarios are being considered for treatment and potential outcome variables by using both parametric and nonparametric statistical modeling for estimation of the generalized propensity scores.

Highlights: During FY 2019, staff worked on bootstrap resampling and simulation methods to fit parametric and non-parametric bootstrap simulation files (a) to check sampling variability and to compute standard errors; (b) to compute relative risk ratio estimates; (c) to check the accuracy of the usual Gaussian based methods; (d) to compute accurate confidence intervals for a variety of statistics and (e) for a variety of complex sampling methods, and how to perform significance tests with p-values in bootstrap resampling and sampling estimation methods. A draft report documenting work on the Generalized Propensity Scores Research is in progress.

Staff: Isaac Dompreh (x36801)

Summer at Census

Description: For each summer since 2009, recognized scholars in the following and related fields applicable to censuses and large-scale sample surveys are invited for short-term visits (one to three 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: During FY 2019, staff facilitated all the details and background with staff around the Census Bureau to host 2019 *SUMMER AT CENSUS* with twenty-nine scholar nominations.

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: Joe Engmark, Brett Moran, Kelly Taylor

3. PUBLICATIONS

3.1 JOURNAL ARTICLES, PUBLICATIONS

Baker, S., McElroy, T., and Sheng, X. (In Press). "Expectation Formation Following Large and Unpredictable Shocks," *Review of Economics and Statistics*.

Barseghyan, L., Molinari, F., Morris, D.S., and Teitelbaum (In Press). "The Cost of Legal Restrictions on Experience Rating," *Journal of Empirical Legal Studies*.

Bell, W.R., Chung, H., Datta, G.S., and Franco, C. (2019). "Measurement Error in Small Area Estimation: Functional Versus Structural Versus Naive Models," *Survey Methodology*, 45, 61–80.

Chai, J. and Nayak, T.K. (2018). "A Criterion for Privacy Protection in Data Collection and Its Attainment via Randomized Response Procedures," *Electronic Journal of Statistics*, 12, 4264-4287.

de Oliveira, V., Wang, B., and Slud, E. (2018). "Spatial Modeling of Rainfall Accumulated over Short Periods of Time," *Journal of Multivariate Analysis, 166*, 129-149.

Dumbacher, B., Morris, D.S., and Hogue, C. (2019). "Using Electronic Transaction Data to Add Geographic Granularity to Official Estimates of Retail Sales." *Journal of Big Data, 6:80*.

Durrant, M., Durrant, L., and McElroy, T. (2019). "Establishing a Common Instantaneous Center of Rotation for the Metatarso- Phalangeal and Metatarso-Sesamoidal Joints: a Theoretical Geometric Model Based on Specific Morphometrics," *Journal of Orthopedic Surgery and Research*, 14:107.

Franco, C., Little, R.J.A., Louis, T.A. and Slud, E.V. (2019). "Comparative Study of Confidence Intervals for Proportions in Complex Surveys," *Journal of Survey Statistics and Methodology, 7.3*, 334-364.

Hyatt, H. and McElroy, T. (In Press). "Labor Reallocation, Employment, and Earnings: Vector Autoregression Evidence," *LABOUR: Review of Labour Economics and Industrial Relations*.

Janicki, R. (In Press). "Properties of the Beta Regression Model for Small Area Estimation of Proportions with Application to Estimation of Poverty Rates," *Communications in Statistics – Theory and Methods.*

Klein, M., Moura, R., and Sinha, B. (In Press). "Multivariate Normal Inference Based on Singly Imputed Synthetic Data under Plug-In Sampling," *Sankhya*.

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

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

Lu, X. (2019). "A Note on Min-Max Pair in Tournaments," *Graphs and Combinatorics, 35*, 1139-1145.

Martin, S., 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. and Jach, A. (2019). "Testing Collinearity of Vector Time Series," Published online, *The Econometrics Journal, 22*, 97-116.

McElroy, T., Pang, O., and Sheldon, G. (2019). "Custom Epoch Estimation for Surveys," *Journal of Applied Statistics*, 46, 638-663.

McElroy, T. and Penny, R. (2019). "Maximum Entropy Extreme-Value Seasonal Adjustment," *Australian New Zealand Journal of Statistics*, 61(2), 152-174.

McElroy, T. and Politis, D. (In Press). *Time Series: A First Course with Bootstrap Starter*, Chapman Hall.

O'Hara, B., Medalia, C., and Maples, J. (2019). "Modeling a Bridge When Survey Questions Change: Evidence from the Current Population Survey Health Insurance Redesign," *Journal of Official Statistics, Vol. 35, Issue 1,* 189-202.

Roy, A., McElroy, T., and Linton, P. (2019). "Estimation of Causal Invertible VARMA Models." *Statistica Sinica, 29, 455-478.*

Sellers, K.F. and Young, D. (2019) "Zero-inflated Sum of Conway-Maxwell-Poissons (ZISCMP) Regression," *Journal of Statistical Computation and Simulation*, 89 (9): 1649-1673.

Slawski, M. and Ben-David, E. (2019). "Linear Regression with Sparsely Permuted Data," *Electronic Journal of Statistics, Vol 13, No. 1*, 1-36.

Slud. E. and Thibaudeau, Y. (2019). "Multi-Outcome Longitudinal Small Area Estimation – A Case Study," *Statistical Theory and Related Fields, 3*, 136-149. DOI: 10.1080/24754269.2019.1669360 Published online September 26, 2019.

Slud, E., Vonta, I., and Kagan, A. (2018). "Combining Estimators of a Common Parameter across Samples," *Statistical Theory and Related Fields 2*(2), 158-171, published online.

Wildi, M. and McElroy, T. (2019). "The Trilemma between Accuracy, Timeliness, and Smoothness in Real-Time Signal Extraction," *International Journal of Forecasting*, 35, 1072-1084.

Wright, T., Klein, M., and Wieczorek, J. (2019). "A Primer on Visualizations for Comparing Populations, Including the Issue of Overlapping Confidence Intervals," *The American Statistician, Vol 73, No. 2,* 165-178.

3.2 BOOKS/BOOK CHAPTERS

Chung, H.-C., Datta, G.S., and Maples, J. (2019). "Estimation of Median Incomes of the American States: Bayesian Estimation of Means of Subpopulations" in S. Bandyopadhyay and M. Dutta (Eds.), *Opportunities and Challenges in Development Esays in Honor of Sarnila Banerjee*, Springer Nature.

Findley, D. and McElroy, T. (2018). "Background and Perspectives for ARIMA Model-Based Seasonal Adjustment." in G. Mazzi and D. Ladiray (Eds.), *Handbook on Seasonal Adjustment*, pp. 215--253. Luxembourg: Publications Office of the European Union.

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

Thibaudeau, Y., Slud, E., and Cheng. Y. (2020). "Log-Linear Modeling for Estimation of Cross-Classified Small Areas in Longitudinal Surveys," in P. Lynn (Ed.), *Methodology of Longitudinal Surveys 2*, New York, NY: Wiley Series in Survey Methodology.

Thibaudeau, Y. and Slud, E. (2020). "Small-Area Estimation of Cross-Classified Gross Flows Using Longitudinal Survey Data." in P. Lynn (Ed), *Methodology of Longitudinal Surveys*, 2 New York, NY: Wiley Series in Survey Methodology.

3.3 PROCEEDINGS PAPERS

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

- Khoa Dong, Timothy Trudell, Yang Cheng, and Eric Slud, "Understanding Variance Estimator Bias in Stratified Two-Stage Sampling," 1638-1645, Survey Research Methods Section.
- Jerry Maples, "Small Area Population Models: Estimating the Number of Children in School Districts," Survey Research Methods Section.
- Darcy Morris, Andrew Raim, and Kimberly Sellers, "Introducing a Conway-Maxwell-Multinomial Distribution for Flexible Modeling of Categorical Data," Biometrics Section.

- Mary Mulry, “Relationship between Positive Responses to Child-specific Probes on 2010 Census Questionnaires and 2010 Census Coverage Measurement Nonmatching Young Children,” 949-965, Survey Research Methods Section.
- Rebecca Steorts, “Entity Resolution with Societal Impacts in Statistical Machine Learning,” Survey Research Methods Section.

3.4 CENTER FOR STATISTICAL RESEARCH & METHODOLOGY RESEARCH REPORTS
<http://www.census.gov/srd/csrmreports/byyear.html>

RR (Statistics #2018-12): Eric Slud, Robert Ashmead, Patrick Joyce, and Tommy Wright, “Statistical Methodology (2016) for Voting Rights Act, Section 203 Determinations,” December 13, 2018.

RR (Statistics #2018-13): James Livsey and Tucker S. McElroy, “A Time Series Model Information Criterion Based Upon Residual Entropy,” December 17, 2018.

RR (Statistics #2019-01): Thomas Trimbur and Tucker McElroy, “Modelled Approximations to the Ideal Filter with Application to GDP and its Components,” January 30, 2019.

RR (Statistics #2019-02): Thomas Trimbur and Tucker McElroy, “A Class of Multivariate Filters for Trend Extraction and Statistical Analysis of Multiple Related Time Series,” February 25, 2019.

RR (Statistics #2019-03): Tommy Wright, “Direct Proof of Exact Sample Allocation Optimality with Cost Constraints,” March 6, 2019.

RR (Statistics #2019-04): Jichong Chai and Tapan K. Nayak, “Minimax Randomized Response Methods for Providing Local Differential Privacy,” March 20, 2019.

RR (Statistics #2019-05): Martin Klein and Bimal Sinha, “Multiple Imputation for Parametric Inference Under a Differentially Private Laplace Mechanism,” May 7, 2019.

RR (Statistics #2019-06): Martin Klein, Ricardo Moura, and Bimal Sinha, “Multivariate Normal Inference on Singly Imputed Synthetic Data under Plug-in Sampling,” May 7, 2019.

RR (Statistics #2019-07): Hang J. Kim, Joerg Drechsler, and Katherine J. Thompson, “Synthetic Microdata for Establishment Surveys Under Informative Sampling,” July 26, 2019.

4. TALKS AND PRESENTATIONS

Shenandoah Undergraduate Mathematics and Statistics Conference, James Madison University, October 13, 2018.

- Kimberly Sellers, Plenary Speaker, "CSI: Count on Statistics for Investigation."

Women in Statistics and Data Science, Cincinnati, OH, October 18-20, 2018.

- Kimberly Sellers, Panelist on "Implicit Bias and Power Dynamics: How We Can Change the Culture."
- Kimberly Sellers, Panelist on "She Leads with Statistics and Data Science Research."

Big Surv18, Barcelona, Spain, October 25-27, 2018.

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

The Ohio State University Statistics Departmental Seminar, Columbus, OH, November 8, 2018.

- Kimberly Sellers, "Introducing a Flexible Bivariate Distribution for Dispersed Discrete Data."

North Carolina State University, Raleigh-Durham, North Carolina, November 15, 2018.

- Tucker McElroy, "Nonlinear Prediction via Hermite Transformation."

Tenth International Triennial Calcutta University Symposium on Probability and Statistics, Kolkata, India, December 27-30, 2018.

- Bimal Sinha, "Finite Sample Inference for Multiply Imputed Synthetic Data under a Multiple Linear Regression Model."

Haverford College, Haverford, Pennsylvania, January 28, 2019.

- Tucker McElroy, "Variable Targeting and Reduction in Large Vector Autoregressions with Applications to Workforce Indicators."

Conference on Statistical Practice (CSP), New Orleans, LA, February 14-16, 2019.

- Darcy Morris, "Don't Count on Poisson: COM-Poisson Regression Models for Count Data."

In Likelihood Free Methods of Inference: A PRIN Workshop, University of Padua, Padua, Italy, February 19-20, 2019.

- Gauri Datta, "Small Area Estimation under Non-normal Random Effects."

Department of Statistics, Invited Speaker Series Seminar, University of Miami, Florida, February 21, 2019.

- James Livsey, "Applying the EM Algorithm to Multivariate Signal Extraction."

Mathematics and Statistics Department, Statistics Colloquium, University of Maryland, Baltimore County, Maryland, March 1, 2019.

- Martin Klein, "Statistical Analysis of Noise Multiplied Data Using Multiple Imputation."

University of North Carolina, Chapel Hill, North Carolina, March 1, 2019.

- Tucker McElroy, "Casting Vector Time Series: Forecasting, Imputation, and Signal Extraction in the Context of Big Data."

Statistics Seminar, University of Maryland, College Park, Maryland, March 7, 2019.

- Eric Slud, "A Case Study in Comparing Bayes Estimated Fixed Effects with Frequentist Estimated Random Effects."

Data Institute San Francisco Annual Conference (DSC019), University of San Francisco, California, March 10-12, 2019.

- James Livsey, "Augmenting Analysis with U.S. Census Data."

New Techniques and Technologies for Statistics, Brussels, Belgium, March 12-14, 2019.

- Tucker McElroy, "A Diagnostic for Seasonality Based Upon Autoregressive Roots."

Department of Statistics, East China Normal University, Shanghai, China, March 19, 2019.

- Bimal Sinha, "Statistical Analysis of Noise Multiplied Data Using Multiple Imputation."

Department of Statistics and Probability, Peking University, Beijing, China, March 21, 2019.

- Bimal Sinha, "Statistical Analysis of Noise Multiplied Data Using Multiple Imputation."

STRATA Data Conference, San Francisco, CA, March 25-28, 2019.

- Yves Thibaudeau, "New Directions in Record Linkage."

Office for National Statistics, Newport, United Kingdom, March 26, 2019.

- Tucker McElroy, "Casting Vector Time Series: Forecasting, Imputation, and Signal Extraction in the Context of Big Data."

Department of Statistics Seminar, University of Virginia, Charlottesville, VA, March 29, 2019.

- Emanuel Ben-David, "A Two-stage Approach to Multivariate Linear Regression with Sparsely Mismatched Data."

Department of Statistics Seminar, George Washington University, Washington, D.C., April 26, 2019.

- Emanuel Ben-David, "Linear Regression with Linked Data."

13th Annual Probability & Statistics Day, University of Maryland, Baltimore County, Maryland, April 27, 2019.

- Tommy Wright, "Lagrange's Identity and Five BIG Results."

Mathematics & Statistics Special Colloquium, University of Washington, Seattle, Washington, April 29, 2019.

- Tommy Wright, "Lagrange's Identity and Apportionment of the U.S. House of Representatives."

Statistics Seminar, Charles III University of Madrid, Madrid, Spain, June 3, 2019.

- Carolina Franco, "Borrowing Strength from Larger Surveys to Improve Related Estimates from Smaller Surveys Using Bivariate Small Area Estimation Models."

The Sixth Italian Conference on Survey Methodology, Florence, Italy, June 5 2019.

- Carolina Franco, "Borrowing Strength from Larger Surveys to Improve Related Estimates from Smaller Surveys Using Bivariate Small Area Estimation Models."

Summer Program in Research and Learning, American University, June 6, 2019.

- Kimberly Sellers, "The Conway-Maxwell-Poisson Regression for Dispersed Count Data."

Conference for African-American Researchers in the Mathematical Sciences (CAARMS), Princeton University, June 19-21, 2019.

- Kimberly Sellers, "Flexible Regression Models for Dispersed Count Data."

Third International Conference on Econometrics and Statistics, National Chung Hsing University, Taichung, Taiwan, June 25-27, 2019.

- Gauri Datta, "Estimation of Median Incomes of the American States: Bayesian Estimation of Means of Subpopulations."

Population Dynamics & Health Program, University of Michigan, Ann Arbor, Michigan, July 10, 2019.

- Rebecca Steorts, "An Introduction to Entity Resolution."

Joint Statistical Meetings, American Statistical Association, Denver, Colorado, July 27-August 1, 2019.

- Carolina Franco, "Using American Community Survey Data to Improve Estimates from Smaller Surveys through Bivariate Small Area Estimation Models."
- Patrick Joyce, "Overdispersed Binomial Small Area Models with Application to Poverty Rate Estimation."
- James Livsey, "Applying the EM Algorithm to Multivariate Signal Extraction."
- Jerry Maples, "Small Area Estimates of the Child Population and Poverty in School Districts Using Dirichlet-Multinomial Models."

- Tucker McElroy, “Quadratic Prediction of Time Series via Auto-Cumulants.”
- Darcy Steeg Morris, “Fitting Flexible Models for Count Data COM-Poisson Regression Bivariate Multinomial and Mixed Models.”
- Mary Mulry, “Variable Selection for Multinomial Logistic Regression Modeling to Assign One of Six Census Mindsets to Database Records.”
- Osbert Pang, “Seasonal Adjustment Subject to Frequency Aggregation Constraints.”
- Andrew Raim, “Statistical Assessment of Bovine Body Weight via Functional Gait Data.”
- Martin Slawski and Emanuel Ben-David, “A Two-Stage Approach to Multivariate Linear Regression with Sparsely Mismatched Data.”
- Eric Slud, “Model-assisted Estimation of Mixed-Effect Model Parameters in Complex Surveys.”
- Tommy Wright, “An Elementary Derivation of Kadane’s Optimal Dynamic Sampling Plan.”

Current Trends in Survey Sampling, National University of Singapore, Singapore, August 13-16, 2019.

- Gauri Datta, “Small Area Estimation under Non-normal Random Effects.”

Workshop on Population and Health Program, University of Michigan, Ann Arbor, Michigan, August 22, 2019.

- Emanuel Ben-David, “Record Linkage Part II.”

5. CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY SEMINAR SERIES

Dobrislav Dobrev, Federal Reserve Board, "A Randomized Missing Data Approach to Robust Filtering with Applications to Economics and Finance," November 1, 2018.

Brian Dumbacher (ESMD) & Darcy Morris (CSRM), U.S. Census Bureau, "Using Electronic Transaction Data to Add Geographic Granularity to Official Estimates of Retail Sales," November 7, 2018.

Jingchen Hu (ASA/NSF/BLS Research Fellow), Vassar College, "Bayesian Pseudo Posterior Synthesis for Data Privacy Protection," March 12, 2019.

Tommy Wright, U.S. Census Bureau, "An Elementary Derivation of Kadane's Optimal Dynamic Sampling Plan," April 16, 2019.

Chloe East, University of Colorado – Denver, *SUMMER AT CENSUS,* "The Effect of Increasing Immigration Enforcement on the Labor Supply of High-Skilled Citizen Women," May 21, 2019.

Partha Lahiri, University of Maryland – College Park, "Small Area Estimation in Presence of Linkage Errors," May 28, 2019.

Katharine Donato, Georgetown University, *SUMMER AT CENSUS,* "The Legal Landscape of U.S. Immigration in the 21st Century," June 11, 2019.

Martin Slawski, George Mason University, *SUMMER AT CENSUS,* "Regression with Unknown Permutation and Adjustment for Linkage Error in Merged Data," June 11, 2019.

Singdhansu Bhusan Chatterjee, University of Minnesota, *SUMMER AT CENSUS,* "On Analysis of Repeated Cross Sectional Data," June 13, 2019.

Robert Lund, Clemson University, *SUMMER AT CENSUS,* "Stationary Count Time Series," June 18, 2019.

David Swanson, University of California, Riverside, *SUMMER AT CENSUS*, "Cohort Change Ratios and their Applications," June 19, 2019.

Michael Schober, The New School, *SUMMER AT CENSUS*, "Overview on Social Media and Surveys," June 20, 2019.

Fred Conrad, University of Michigan, *SUMMER AT CENSUS*, "Social Media as an Alternative to Surveys of Opinions about the Economy," June 20, 2019.

Robyn Ferg, University of Michigan, "An Introduction to Obtaining Tweets and Analyzing their Sentiment," June 20, 2019.

Jacob Bastian, University of Chicago, *SUMMER AT CENSUS*, "The Rise of Working Mothers and the 1975 Earned Income Tax Credit," June 25, 2019.

Amos Golan, American University, Santa Fe Institute, Pembroke College, University of Oxford, *SUMMER AT CENSUS*, "Info-Metrics for Modeling and Inference: a Four-Lecture Series," June 25-27, 2019.

Elizabeth Stuart, Johns Hopkins University, *SUMMER AT CENSUS*, "Using Propensity Scores to Estimate Causal Effects in Non-Experimental Studies," July 8, 2019.

Dongchu Sun, University of Missouri-Columbia, *SUMMER AT CENSUS*, "Bayesian Analysis of One-way Multivariate ANOVA Models," July 9, 2019.

Zhuoqiong He, University of Missouri-Columbia, *SUMMER AT CENSUS*, "Bayesian Smoothing Spline Model and Its Application in Current Population Survey," July 9, 2019.

Xiao-Li Meng, Harvard University, *SUMMER AT CENSUS*, "Data Science: What is it not?," July 10, 2019.

Xiao-Li Meng, Harvard University, *SUMMER AT CENSUS*, "A Trio of Inference Problems That Could Win You a Nobel Prize in Statistics (if you help fund it)," July 11, 2019.

Abel Kho, Northwestern University, *SUMMER AT CENSUS*, "The Building Blocks of Inter-operability: A Multisite Analysis of Patient Demographic Attributes Available for Matching," July 15, 2019.

Jesse Frey, Villanova University, *SUMMER AT CENSUS*, "Improved Confidence Intervals for a Proportion Using Ranked-Set Sampling," July 16, 2019.

Bimal Sinha, University of Maryland, Baltimore County/U.S. Bureau of the Census, "Data Analysis under Differential Privacy," July 16, 2019.

Trent Buskirk, Bowling Green State University, *SUMMER AT CENSUS*, "Applications of Supervised and Unsupervised Machine Learning Methods for Sample Design and Analysis," July 17, 2019.

Robert Phillips, American Board of Family Medicine Foundation, *SUMMER AT CENSUS*, "How Other Countries Use Deprivation Indices – and Why the United States Desperately Needs One," July 25, 2019.

Paul Biemer, RTI International/University of North Carolina at Chapel Hill, "Total Error Frameworks for Integrated Survey and Found Data," August 5, 2019.

David Autor, MIT and Anna Salomons, Utrecht School of Economics, *SUMMER AT CENSUS*, "Documenting the Emergence of New Work in the United States over 1900-2015," August 6, 2019.

Mauricio Sadinle, University of Washington, *SUMMER AT CENSUS*, "Bayesian Propagation of Record Linkage Uncertainty into Population Size Estimation," August 7, 2019.

Oksana Honchar, Australian Bureau of Statistics, "Measuring Impact of Transformation in Official Statistics," August 8, 2019.

Richard J. Harris, The University of Texas at San Antonio, *SUMMER AT CENSUS*, "Selected Research Applications that Potentially Would Benefit from New Measurement Strategies for the New Race and Ethnicity Questions in the 2020 Census," August 13, 2019.

Amy Kind, University of Wisconsin, *SUMMER AT CENSUS*, "Making Neighborhood-Disadvantage Metrics Accessible – The Neighborhood Atlas," August 26, 2019.

Zhen-Qing Chen, University of Washington, *SUMMER AT CENSUS*, "High Density Multi-Type Population Growth Model," August 27, 2019.

Jian-Guo Liu, Duke University, *SUMMER AT CENSUS*, "Random Batch Method and Its Application to Sampling," August 27, 2019.

Joshua Gottlieb, University of Chicago, *SUMMER AT CENSUS*, "What Does Health Care Billing Cost, and Why Does It Matter?," August 28, 2019.

Carolina Franco, U.S. Bureau of the Census, "Using American Community Survey Data to Improve Estimates from Smaller Surveys through Bivariate Small Area Estimation Models," August 28, 2019.

Terrance Savitsky, U.S. Bureau of Labor Statistics, "Topic 1: Bayesian Dependent Functional Mixture Estimation for Area and Time-indexed Data and Topic 2: Bayesian Estimation under Informative Sampling," August 29, 2019.

Changrong Yuan, Fudan University, *SUMMER AT CENSUS*, "Construction and Evaluation of a Smartphone-Based Information Support Framework for Breast Cancer Care," September 3, 2019.

Emanuel Ben-David, U.S. Bureau of the Census, "Some Approaches to Remedy Regression Estimates Working with Linked Data Subject to Linkage Error," September 3, 2019.

James Livsey, U.S. Bureau of the Census, "Multivariate Signal Extraction," September 4, 2019.

Jerry Maples, U.S. Bureau of the Census, "Small Area Estimates of the Child Population and Poverty in School Districts using Dirichlet-Multinomial Models," September 4, 2019.

Darcy Steeg Morris, U.S. Bureau of the Census, "Conway-Maxwell Distributions for Flexible Modeling of Count Data," September 5, 2019.

Yves Thibaudeau, U.S. Bureau of the Census, "Small-Area Estimation of Cross-Classified Gross Flows using Semi-Longitudinal Survey Data," September 5, 2019.

Justin Grimmer, Stanford University, *SUMMER AT CENSUS*, "Supervised Learning Methods Applied to Text as Data," September 10, 2019.

Jiming Jiang, University of California-Davis, *SUMMER AT CENSUS*, "Big Data in Surveys: Challenges, Opportunities, and Strategies," September 25, 2019.

Jiming Jiang, University of California-Davis, *SUMMER AT CENSUS*, "Sumca: Simple, Unified, Monte-Carlo Assisted Approach to Second-order Unbiased MSPE Estimation with Application to Small Area Estimation," September 26, 2019.

6. PERSONNEL ITEMS

6.1 HONORS/AWARDS/SPECIAL RECOGNITION

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

- **Maria Garcia** – "Maria Garcia supported and expanded the Standard Economic Processing System (StEPS), which is the main computational tool to impute, edit, and balance financial data and statistics. Dr. Garcia programmed and coded parts of StEPS designed to balance financial statistics subject to complex constraints in the *Financial Quarterly Report*."

Poster Award, Survey Research Methods Section, American Statistical Association, 2019 Joint Statistical Meetings

- **Jerry Maples** – "Small Area Estimates of the Child Population and Poverty in School Districts Using Dirichlet-Multinomial Models."

6.2 SIGNIFICANT SERVICE TO PROFESSION

Robert Ashmead

- Refereed a paper for *Journal of Official Statistics*

Emanuel Ben-David

- Refereed papers for *Mathematical Reviews*, *Wiley Interdisciplinary Reviews: Computational Statistics* and *Journal of Computational and Graphical Statistics*

Gauri Datta

- Associate Editor, *Sankhya*
- Guest Co-Editors, *Sankhya*, Special Issue (Memorial volume for J.K. Ghosh)
- Associate Editor, *Statistical Methods and Applications*
- Associate Editor, *Environmental and Ecological Statistics*
- Editorial Member, *Calcutta Statistical Association Bulletin*

Carolina Franco

- Member, Gertrude Cox Scholarship Committee
- Refereed two papers for *the Journal of the Royal Statistical Society, Series A*
- Member, International Relations of Statistics Committee, American Statistical Association

Kyle Irimata

- Reviewed textbook proposal for Springer

Ryan Janicki

- Refereed papers for *International Statistical Review* and *The American Statistician*

Martin Klein

- Refereed a paper for *Journal of Survey Statistics and Methodology*

James Livsey

- Refereed papers for *Journal of Statistical Software, The American Statistician, Stochastic Environmental Research and Risk Assessment, Journal of Multivariate Analysis and Environmetrics, Applied Stochastic Models in Business and Industry,* and *International Statistical Review*
- Session Organizer, 2019 Joint Statistical Meetings, Denver, Colorado

Xiaoyun Lu

- Refereed a paper for *Australasian Journal of Combinatorics*

Jerry Maples

- Refereed papers for *Sankhya* and *Journal of Statistical Theory and Related Fields*

Thomas Mathew

- Associate Editor, *Journal of the American Statistical Association*
- Associate Editor, *Sankhya*
- Associate Editor, *Journal of Multivariate Analysis*

Tucker McElroy

- Refereed papers for the *Annals of Statistics, IEEE, International Statistical Review, Journal of Time Series Analysis, Journal of Applied Econometrics, and Journal of Official Statistics*

Darcy Morris

- Associate Editor, *Communications in Statistics*
- Treasurer, Survey Research Methods Section, American Statistical Association
- Reviewed papers for *Filomat* and *Journal of Modern Applied Statistical Methods*

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, *Journal of Statistical Theory and Practice*
- Session Organizer, International Conference on Statistical Distributions and Applications, October 10-12, 2019, Grand Rapids, Michigan

Osbert Pang

- Reviewed a book proposal for CRC Press

Ned Porter

- Reviewed papers for *Formal Methods for Statistical Software* (NIST)

Andrew Raim

- Refereed a paper for *Communications in Statistics*

Kimberly Sellers

- Chairperson, American Statistical Association Committee on Women in Statistics (through December 2018)
- Associate Editor, *The American Statistician*
- Associate Editor, *Journal of Computational and Graphical Statistics*
- Organizing Committee member, Women in Statistics and Data Science Conference, Cincinnati OH, October 2018
- Commissioning Editor, *WIREs Computational Statistics*
- Refereed a paper for *Statistics and Probability Letters*
- Advisory Board member, Summer Program in Research and Learning (SPIRAL), American University

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*
- Chair, NAS Steering Committee for Workshop on Using Models to Estimate Hog Production by National Agricultural Statistical Service (NASS)
- Member, FDA Pharmaceutical Sciences and Clinical Pharmacology Advisory Committee
- Member, Working Group on National Research Program Design, Internal Revenue Service
- Refereed papers for *Journal of Survey Statistics and Methodology*, *Journal of the American Statistical Association*, and *Journal of Royal Statistical Society, Series B*

Rebecca Steorts

- Associate Editor, *Journal of Survey Statistics and Methodology*
- Associate Editor, *Journal of the American Statistical Association, Applications and Case Studies*
- 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
- Organizing Committee, International Indian Statistical Association (IISA) International Conference on Statistics

Yves Thibaudeau

- Refereed a paper for *Statistical Theory and Related Fields*

Thomas Trimbur

- Refereed papers for *Journal of Forecasting* and *Journal of Official Statistics*

William E. Winkler

- Associate Editor, *Transactions on Data Privacy, Journal of Privacy and Confidentiality*
- 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*
- Refereed three papers for IEEE DINA2019

Tommy Wright

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

6.3 PERSONNEL NOTES

Brian Monsell, following 36+ years of service as member of the Census Bureau's Time Series Research Group, retired.

Robert Ashmead accepted a position with the Ohio Colleges of Medicine Government Resource Center, The Ohio State University.

Martin Klein accepted a position with the Food and Drug Administration.

William Winkler, following 39 years of federal service (most) as member of the Census Bureau's Machine Learning & Computational Statistics Research Group, retired.

APPENDIX A **Center for Statistical Research and Methodology FY 2019**
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		
6350E01	Address Canvassing In Office		
6450E20	Advertising Campaign		
6550E01	Data Coding/Editing/Imputation		
6550E06	Redistricting Data Program		
6650E01	CCM Planning and Project Management		
6650E20	2020 Evaluations-Planning and Project Management		
6650E25	2030 Planning and Project Management		
6750E01	Administrative Records Data		
	1. 2020 Census Communications Campaign Statistical Analyses-I	Mary Mulry	Gina Walejko
	2. 2020 Census Communications Campaign Statistical Analyses-II	Mary Mulry	Gina Walejko
	3. Supplementing and Supporting Non-Response with Administrative Records	Michael Ikeda	Tom Mule
	4. 2020 Census Privacy Research	Robert Ashmead	Rob Sienkiewicz
	5. Identifying "Good" Administrative Records for 2020 Census NRFU Curtailment Targeting	Darcy Morris	Tom Mule
	6. 2010 Congressional District Summary File for the 116th Congress	Tommy Wright	James Whitehorne
	7. Experiment for Effectiveness of Bilingual Training	Andrew Raim	Renee Ellis
	8. Unit-Level Modeling of Master Address File Adds and Deletes	Eric Slud	Nancy Johnson
	9. Coverage Measurement Research	Jerry Maples	Tim Kennel
	10. Development of Block Data Tracking Database.	Tom Petkunas	Graham Baggett
	11. A Deterministic Retabulation of Colorado and Pennsylvania Congressional District Profiles from the 115th Congress to the 116th Congress	Tommy Wright	James Whitehorne
	12. Assessing Variability of Data Treated by TopDown Algorithm for Redistricting	Tommy Wright	John Abowd
6385E70	American Community Survey (ACS)		
	13. Voting Rights Section in 203 Model Evaluation and Enhancements Towards 2021 Determinations	Eric Slud	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 on Balanced Repeated Replication and Other Variance Estimation Techniques for Use with Current Population Survey	Eric Slud	Tim Trudell
0906/1444X00	Demographic Surveys Division (DSD) Special Projects		
	16. Data Integration	Edward Porter	Christopher Boniface
7165018/ 7165019	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. Use of Big Data for Retail Sales Estimates	Darcy Morris	Brian Dumbacher
	21. Seasonal Adjustment Support	Tucker McElroy	Kathleen McDonald-Johnson
	22. Seasonal Adjustment Software Development and Evaluation	Tucker McElroy	Kathleen McDonald-Johnson
	23. Research on Seasonal Time Series - Modeling & Adjustment Issues	Tucker McElroy	Kathleen McDonald-Johnson
	24. Supporting Documentation & Software for X-13ARIMA-SEATS	Tucker McElroy	Kathleen McDonald-Johnson
	25. Redesign of Economic Sample Surveys (Stratification)	Patrick Joyce	Justin Z. Smith
	BUREAU OF ECONOMIC ANALYSIS		
	26. Business Cycle Movements in National Accounts	Thomas Trimbur	Baoline Chen
	FEDERAL HIGHWAY ADMINISTRATION		
	27. Modeling and Signal Extraction of Pedestrian and Bicycle Crash Data	James Livsey	Roya Amjadi

APPENDIX B

FY 2019 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 2019. For FY 2019, the *measures of performance* for our center are:

Measure 1. Overall, Work Met Expectations: Percent of FY 2019 Program Sponsored Projects/Subprojects where sponsors reported that work met their expectations.

Measure 2. Established Major Deadlines Met: Percent of FY 2019 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 2019 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight.

Measure 3b. Plans for Implementation: Of the FY 2019 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 2019 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 2019.

Measure 6. Proceedings Publications: Number of proceedings publications documenting research that appeared in FY 2019.

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, 2019 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 2019 ***(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 2019 where a 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 2019 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 2019 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
VACANT

Record Linkage & Machine Learning Research Group
Yves Thibaudeau
- Emanuel Ben-David
- Xiaoyun Lu
- Rebecca Steorts (Duke U.)
- Dan Weinberg

Missing Data & Observational Data Modeling Research Group
Darcy Morris
- Maria Garcia
- Jun Shao (U. of WI)

Research Computing Systems & Applications Group
Chad Russell
- Tom Petkunas
- Ned Porter

Simulation, Data Science, & Visualization Research Group
Tommy Wright (Acting)
- Isaac Dompreh
- Brett Moran
- Bimal Sinha (UMBC)
- Nathan Yau (FLOWINGDATA.COM)

MATHEMATICAL STATISTICS AREA
Eric Slud

Sampling & Survey Inference Research Group
Eric Slud (Acting)
- Mike Ikeda
- Patrick Joyce
- Mary Mulry
- Tapan Nayak (GWU)

Small Area Estimation Research Group
Jerry Maples
- Gauri Datta
- Kyle Irimata
- Ryan Janicki

Carolina Franco

Time Series & Seasonal Adjustment Research Group
James Livsey
- Osbert Pang

Tucker McElroy (Acting)
- Soumendra Lahiri (Washington U.)
- Anindya Roy (UMBC)
- Thomas Trimbur

Experimentation, Prediction, & Modeling Research Group
Tommy Wright (Acting)
- Thomas Mathew (UMBC)
- Andrew Raim
- Kimberly Sellers (Georgetown U.)

OFFICE OF THE CHIEF
Tommy Wright
- Kelly Taylor
- Joe Engmark
- Hang Kim (F)
- Michael Hawkins

(F) ASA/NSF/Census Research Fellow

January 2020