## STATISTICAL RESEARCH DIVISION FY 2010 FIRST QUARTER REPORT –October, November, December 2009–

## **COLLABORATION**

## **DECENNIAL DIRECTORATE**

# Decennial Management Division/Decennial Statistical Studies Division/American Community Survey Office (Sponsors)

Project	Project	
Number	Title	FTEs
5210001	Forms Development	
Α.	Census Questionnaire Design Features (Other than Race and Ethnicity)	
В.	Development of Race and Ethnicity Questions	
5210003	Language Planning and Development	1.00
5310001	Data Collection Planning and Development	1.00
Α.	Accessible Web Surveys (Research)	
В.	Decennial Reinterview Internet Testing - Usability Input	
5310008	Special Place/Group Quarters (GQ) Planning and Development	
5610002	Statistical Design and Estimation	6.67
Α.	Decennial Editing and Imputation	
В.	Decennial Record Linkage	
С.	Decennial Disclosure Avoidance	
<i>D</i> .	Census Unduplication Research	
Ε.	Statistical Design for Experiments and Evaluations	
5610003	Coverage Measurement Planning and Development	1.39
А.	Coverage Measurement Research	
В.	Accuracy of Coverage Measurement	
С.	Questionnaire Wording and Automation Team	
5610005	Coverage Improvement Planning and Development	1.00
5610006	Evaluation Planning Coordination	3.50
Α.	Development of Questionnaires for Decennial Coverage Improvement	
В.	2010 CPEX Experimental Overcount Booklet	
С.	Evaluations, Experiments, and Assessments Operational Integration Team (EEA OIT)	
<i>D</i> .	Evaluation of CCM Interviews	
Ε.	Investigation of Study Methods for the Census Coverage Measurement (CCM) on Group	
	Quarters (GQ) Population	
F.	2010 Census Language Study (CPEX)	
<i>G</i> .	2010 Census Behavior Coding Evaluation	
Н.	Comparative Ethnographic Studies of Enumeration Methods and Coverage in Race/Ethnic	
	Groups	
Ι.	Explaining How Census Errors Occur through Comparing Census Operations History with	
	Census Coverage Measurement Results	
<i>J</i> .	2011 Relationship Survey	
5385060	American Community Survey (ACS)	2.23
Α.	ACS Missing Data and Imputation	
В.	ACS Group Quarters Item Imputation and Micro Data Disclosure Avoidance Research	
С.	ACS Applications for Time Series Methods	
<i>D</i> .	ACS Variances	
Е.	ACS Small Area Estimation for Selected Characteristics	
F.	ACS Small Area Estimation for Group Quarters	
G.	ACS Data Issues	

5385095	American Community Survey (ACS)/Methods Panel	3.32
Α.	ACS Language Research	

- B. ACS Data Reliability Indicator Project
- C. ACS Messaging Project
- D. ACS Internet Testing Usability Input
- E. ACS Internet Testing Cognitive Input
- F. ACS Internet Test Experimental Design Team
- G. ACS Iterative Testing of the Web Site

## **DEMOGRAPHIC DIRECTORATE**

#### **Population Division (Sponsor)**

Project	Project	
Number	Title	FTEs
TBA	Current Population Survey (CPS)/Annual Social and Economic Supplement (ASEC) Tables	TBA

#### Demographic Surveys Division (Sponsor)

Project	Project	
Number	Title	FTEs
0906/737	4 Demographic Surveys Division Special Projects	1.16
Α.	Data Integration	
В.	Using Survey Paradata to Manage Surveys in the Field and Estimate Survey Error	
С.	Usability Testing of the National Survey of College Graduates	
1465001	Quick Turnaround Pretesting of Household Surveys	1.06
Α.	Rental Housing Finance Survey	
В.	National Crime Victimization Survey (NCVS)	
С.	Development of the CARI Behavior Coding System	
Housing	and Household Economic Statistics Division (Sponsor)	
Project	Project	
Number	Title	FTE

Number	Title	FTEs
1465444	Re-Engineered Survey of Income and Program Participation Methodological Research	1.75

#### Data Integration Division (Sponsor)

Project	Project	
Number	Title	FTEs
7165000	Data Integration Division Small Area Estimation Projects	
Α.	Research for Small Area Income and Poverty Estimates (SAIPE)	
В.	Small Area Health Insurance Estimates (SAHIE)	

## **ECONOMIC DIRECTORATE**

Project	Project	
Number	Title	FTEs
2370054	Editing Methods Development	
	Investigation of Selective Editing Procedures for Foreign Trade Programs	
2470051	Disclosure Avoidance Methods	1.14

2370052	Time Series Research	. 2.11
Α.	Seasonal Adjustment Support	
В.	Seasonal Adjustment Software Development and Evaluation	
С.	Research on Seasonal Time Series - Modeling and Adjustment Issues	
<i>D</i> .	Supporting Documentation and Software for X-12-ARIMA and X-13A-S	
TBA	Survey of Research and Development in Industry, Imputation and Sampling Research and	
	Software Design	TBA

## STRATEGIC PLANNING AND INNOVATION

Project	Project	
Number	Title	TEs
0359999	Remote Access - Microdata Analysis System	1.72

## **CENSUS BUREAU**

Project	Project	
Number	Title	FTEs
0381000	Program Division Overhead	12.25
Α.	Division Leadership and Support	
В.	Research Computing	

# GENERAL RESEARCH AND SUPPORT

Project	Project	
Number	Title	FTEs
0351000	General Research and Support	. 27.26
1871000	General Research	3.26
STATISTI	ICAL METHODOLOGY	
Α.	Disclosure Avoidance	
В.	Disclosure Avoidance for Microdata	
С.	Seasonal Adjustment	
<i>D</i> .	Household Survey Design and Estimation	
Ε.	Sampling and Estimation Methodology: Economic Surveys	
F.	Research and Development Contracts	
G.	Small Area Estimation	
STATISTI	CAL COMPUTING METHODOLOGY	
Α.	Record Linkage and Analytic Uses of Administrative Lists	
<i>B.1</i> .	Editing	
<i>B.2</i> .	Editing and Imputation	
С.	Developed Software Support – General Variance Estimation Development and Support	
<i>D</i> .	Missing Data and Imputation: Multiple Imputation Feasibility Study	
Ε.	Modeling, Analysis and Quality of Data	
SOCIAL d	& BEHAVIORAL SCIENCES SURVEY METHODOLOGY	
Α.	Usability Research and Testing	
В.	Questionnaire Pretesting	
С.	Questionnaire Design Experimental Research Survey 2006 (QDERS)	
<i>D</i> .	Language: Interdisciplinary Research on Language and Sociolinguistic Issues Relevant to	
	Survey Methodology	
Е.	Training for Cognitive Interviewing	
F.	Research on Cognitive Testing of Non-English Language Survey Instruments	
G.	Interviewer-Respondent Interactions	
Н.	Q-Bank: A Database of Pretested Questions	
Ι.	Health Insurance Measurement	

- J. Emerging Social Trends on Household Structure and Living Situations, Race/Ethnicity, and Linkages to Enumeration Method and Coverage
- K. Using Vignettes to Explore Survey Concepts
- L. Retrieval Effects on Judgments about Knowledge

RESEARCH SUPPORT AND ASSISTANCE

## **PUBLICATIONS**

- Journal Articles, Publications
- Books/Book Chapters
- Proceedings Papers
- Statistical Research Division Research Reports
- Statistical Research Division Studies
- Other Reports

## TALKS AND PRESENTATIONS

## STATISTICAL RESEARCH DIVISION SEMINAR SERIES

## **PERSONNEL ITEMS**

- Honors/Awards/Special Recognition
- Significant Service to Profession
- Personnel Notes

## **1. COLLABORATION**

### 1.1 FORMS DEVELOPMENT (DECENNIAL PROJECT 5210001)

# A. Census Questionnaire Design Features (Other than Race and Ethnicity)

*Description*: This project involves participation in decennial content team meetings, including the Content and Forms Design Integrated Product Team, the Housing Unit Operational Integration Team, the Nonresponse Followup Instrument Subteam, the Mode Consistency Subteam, and the Census Program for Evaluations and Experiments (CPEX) Implementations Team. It also includes cognitive pretesting of census questionnaires.

*Highlights:* This quarter, staff assisted Field Division with updating training on how to deal with erasures on the Enumerator Questionnaire forms. Staff also provided expert guidance on the development of and testing plan for the experimental Internet census form for the 2010 Census Quality Survey.

*Staff*: Jennifer Hunter Childs (x34927), Leticia Fernández, Nathan Jurgenson, George Higbie, Anissa Sorokin, Mikelyn Meyers, Dawn Norris, Matt Clifton, Lorraine Randall

#### **B.** Development of Race and Ethnicity Questions

*Description*: Staff conducted cognitive pretesting of five alternative versions of the race and ethnicity questions used in the Decennial Census for the 2010 CPEX panels. Staff also conducted cognitive pretesting of race and ethnicity questions used in a reinterview which will be conducted in the 2010 Census.

*Highlights:* Staff revised the reintervew cognitive pretesting report and conducted more analysis of reinterview cognitive pretesting data.

*Staff:* Rodney Terry (x35475), Leticia Fernández, Jennifer Hunter Childs, Patricia Goerman, Terry DeMaio, Yuling Pan, Matt Clifton, George Higbie, Mikelyn Meyers, Nathan Jurgenson

## 1.2 LANGUAGE PLANNING AND DEVELOPMENT (Decennial Project 5210003)

*Description*: Staff members participate in the interdivisional Decennial Task Force, or language team, which focuses on developing and planning the Language Program for the 2010 Census, pre-census tests, and the Dress Rehearsal. In addition, staff members in our division provide consultation and technical support in the design, development and conduct of research for Decennial language-related projects.

*Highlights:* Staff continued ad hoc review of Spanish and Chinese 2010 Census promotion materials.

*Staff:* Patricia Goerman (x31819), Yuling Pan, Leticia Fernández, Virginia Wake Yelei, Matt Clifton, Anissa Sorokin

## 1.3 DATA COLLECTION PLANNING AND DEVELOPMENT (Decennial Project 5310001)

#### A. Accessible Web Surveys (Research)

*Description:* There is much for Web survey designers to keep in mind when designing surveys to conform to *Section 508* regulations. The regulations require persons with disabilities to have access comparable to the access available to others. This means individuals with visual deficits who use a screen-reader to read text must have the same visual sequence of questions, answer choices, skip patterns, and instructions.

*Highlights:* This quarter, staff began research into which Web survey tools available as COTS software or developed in-house conforms to *Section 508* regulations.

*Staff:* Larry Malakhoff (x33688), Temika Holland, Andrew Zukerburg (NCES)

## **B.** Decennial Reinterview Internet Testing – Usability Input

*Description:* The usability team is working with the cognitive team on the development of an online instrument for the Decennial Reinterview Project. The project will consist of two rounds of usability and cognitive testing of prototypes of the internet instrument and the letters associated with the paper forms. Staff will provide test plans and formal reports for each round of testing as well as provide input at regular team meetings during the development of the ACS online instrument and its subsequent field testing.

*Highlights:* Staff attended regular meetings and assisted the sponsor team in developing wireframe screens for the Internet data collection interface. Also, two rounds of usability and cognitive testing were planned and are expected to take place later in FY2010.

*Staff:* Kathleen Ashenfelter (x34922)

### 1.4 SPECIAL PLACE/GROUP QUARTERS (GQ) PLANNING AND DEVELOPMENT (Decennial Project 5310008)

[See Projects 5610005 and 5610006 (F).]

## 1.5 STATISTICAL DESIGN AND ESTIMATION (Decennial Project 5610002)

#### A. Decennial Editing and Imputation

[See Projects 0351000 and 1871000 (B), General Research - Statistical Methodology]

#### **B. Decennial Record Linkage**

[See Projects 0351000 and 1871000 (A), General Research - Statistical Computing Methodology]

#### **C. Decennial Disclosure Avoidance**

*Description:* The purpose of this research is to develop disclosure avoidance methods to be used for Census Bureau publicly available decennial census and American Community Survey (ACS) data products. Emphasis will be placed on techniques to implement disclosure avoidance at the stage of processing. Disclosure avoidance research will be conducted on alternative methods to protect both tabular data and microdata from the decennial census and the ACS. Methods will be developed, tested, evaluated, and documented. We will also aid in the implementation of the methods.

*Highlights:* An official confidential memo was sent from the disclosure avoidance staff to staff from Decennial Systems and Processing Office (DPSO) that had, as an attachment, the Disclosure Avoidance Specifications for Census 2010. This confidential attachment describes the way swapping is implemented, e.g., it describes the input and output files, the way the SAS program should be run, and the way the results should be analyzed, and lists the code for the program.

Staff continued the development of data swapping software for the ACS 5-year data products. The goal is to redesign the swapping procedure so that the data products will provide disclosure protection at the block group level and above. Staff evaluated the swapping procedure for the state of New Jersey at the county level. A report is forthcoming.

Staff are investigating another way of protecting tabular data for small geographic areas and are applying measures of data utility and disclosure risk to compare this newly developed method of imputing geographies with the current swapping methods. A report is forthcoming.

Staff are developing software for 2010 data swapping for the Island Area and created the input files for the program based on the text files. The Disclosure Review Board continued work on developing disclosure rules for ACS five-year base tables.

Additionally, Staff worked on creating a fully synthetic Decennial microdata file that closely matches the published SF1 tables. Using published SF1 tables as margins for log-linear models, staff has shown theoretically how to develop an appropriate model for this situation through specification of an appropriate design matrix and parameter constraints. Staff have proposed computational solutions for fitting such loglinear models, and the evaluation of these computational solutions is underway. Staff are developing a computational method for model fitting to large scale geographies that involves subdividing the problem and fitting smaller models tract-by-tract and in parallel. Algorithms for log-linear model fitting to sparse tables are being developed and evaluated in order to be able to fit large models in a reasonable amount of time.

*Staff:* Laura Zayatz (x34955), Asoka Ramanayake, Jason Lucero, Paul Massell, Shira Appelbaum, Julie Tsay, Marlow Lemons, Rob Creecy, Martin Klein

#### **D.** Census Unduplication Research

*Description:* The goal of this project is to conduct research to guide the development and assessment of methods for conducting nationwide matching and unduplication in the 2010 decennial census. One of the major problems is how to incorporate the effects of name frequency into the unduplication procedures. Staff also provides assistance in specifying and reviewing output from the matching and unduplication procedures for test censuses and eventually for Census 2010. We began this project in May 2004.

*Highlights:* Staff continued examining results from running the four-pass matching system on the Census 2000 data. Based on this examination, some revisions were made to the list of invalid names for matching in the 2010 Decennial Census.

Staff: Michael Ikeda (x31756), Ned Porter

#### **E. Statistical Design for Experiments and Evaluations** *Description:* The overall objective of this project is to provide statistical expertise in experimental design for decennial-related experiments and evaluations.

In the design and evaluation of experiments to estimate the measurement error resulting from responses based on a census internet questionnaire, in comparison with those from a census paper questionnaire, latent class analysis (LCA) will be used. Three responses will be available for this purpose: those from census paper questionnaires, internet re-interviews, and mail re-interview questionnaires. This will result in an identifiable LCA model. Simple response variances will be estimated based on the latent class model, in order to compare the different data collection modes. The simple response variance is a function of the latent class model parameters. Methodologically, a loglinear representation of the latent class model can be used to estimate the model parameters based on maximum likelihood.

*Highlights:* For applying latent class analysis to survey data, the survey conditions need not be identical for the three interviews. Thus latent class analysis permits the estimation of simple response variances under assumptions that are more general than those made in traditional analysis.

*Staff:* Thomas Mathew (x35337), Aref Dajani

## 1.6 COVERAGE MEASUREMENT PLANNING AND DEVELOPMENT (Decennial Project 5610003)

#### A. Coverage Measurement Research

*Description*: Staff members conduct research on modelbased small area estimation of census coverage, and they consult and collaborate on modeling coverage measurement.

Highlights: An evaluation to include random effects in the logistic coverage models as a way to safeguard against model error is nearly complete. The aim of the study was to determine the possible gains in using random effects in the coverage measurement context and to determine the usefulness of available software. For small models. SAS PROC nlmixed works fine and exhibits a "correction" to estimates that may be missing interaction terms. For a larger, more realistic coverage model, it has not been possible to successfully implement the SAS procedure. Using SAS IML has been successful but the empirical Bayes approach implemented has been ineffectual due to near-zero MLE for the variance components in the model. Currently, a full Bayes procedure is being implemented, in order to get a more complete picture of what may be useful.

Implementation of a "standard" random effects small area coverage model was begun for discussion purposes. The particular model being implemented postulates a small area component correlated among match rate, correct enumeration rate and data-defined rate. The model also includes a "model error" random effect for these three components. Standard "rule-of-thumb" design effect adjustment to the binomial model is currently being used with the aim of investigating the worthiness of this assumption later. Data from the 2006 Travis County test site is being used with census tract as the small area.

Staff continued to work on a recursive partitioning procedure as part of a small area model development using the 2006 Census Coverage test data in Travis County, TX. Recursive partitioning was used to allow

many fixed-effect interaction terms into the logistic regression models of match, correct enumeration status and data-defined rate with the aim of investigating design effects after taking into account the fixed effects model. Staff is also preparing a SAS program template to automate the recursive partitioning procedure for future use.

Staff also worked on another, similar recursive partitioning procedure, solely for racial grouping the large number of multi-racial/ethnicity cells into smaller subsets that are still internally homogeneous with respect to match rate and erroneous enumeration rate.

Work continues on developing a nonignorable/ignorable model hybrid that consists of a mixture of saturated models. This model is designed to work with tabular data. Although no plans for implementation of this model to coverage data is planned, it will be used to help assess other nonignorable models in the literature.

Staff attends weekly meetings of each of the three Decennial Statistical Studies Division coverage measurement subgroups: the CCM estimation group, the component missing data group, and the small area estimation group.

*Staff:* Don Malec (x31718), Aaron Gilary, Ryan Janicki, Jerry Maples, Julie Tsay, Partha Lahiri

#### **B.** Accuracy of Coverage Measurement

*Description*: 2010 Census Coverage Measurement (CCM) Research conducts the research necessary to develop methodology for evaluating the coverage of the 2010 Census. This includes planning, designing, and conducting the research, as well as analyzing and synthesizing the results to evaluate their accuracy and quality. The focus is on the design of the Census Coverage Measurement survey and estimation of components of coverage error with secondary emphasis on the estimation of net coverage error. The estimation of overcount and undercount separately has not been done for previous censuses because of the difficulty of obtaining adequate data for unbiased estimates. The first attempt to implement the new methodology is with data from the 2006 Census Test.

*Highlights:* Staff provided technical expertise and experience in the planning and implementation of coverage measurement research for the 2010 Census. This included serving on three teams formed to plan and implement Census Coverage Measurement (CCM) research for the 2010 Census.

Staff continued previous work on the error structure for estimates of components of census coverage error by refining the structure and identifying sources of the errors. The focus was on the types of errors in the CCM P Sample and analysis of how these errors affect the estimates of net error and the component estimate of omissions.

Staff continued work on designs for some of the CCM evaluation studies in the 2010 Census Program for Evaluations and Experiments (CPEX). The combination of CCM CPEX projects is designed to provide information about the basic types of errors that may affect the CCM implementation.

#### Staff: Mary Mulry (x31759)

#### C. Questionnaire Wording and Automation Team

Description: The purpose of this project is to design the coverage measurement survey instruments for the 2010 Census. These instruments will gather enough data to measure both person and household coverage of the 2010 Census. In preparation for 2010, there will be a 2006 test of the coverage measurement operation in specific sites in conjunction with the 2006 Census Test. For 2006, there will be automated person interview (PI) collecting an independent roster of people living at pre-selected sample addresses in the sites and their residency. There will also be a paper-based person followup (PFU) questionnaire which collects additional residency information about some people collected in the census or the independent roster, but for whom we did not collect enough residency information to determine where they should have been counted for the census. Both these instruments will be used to measure person coverage. Our immediate goals are to create and test these two instruments given requirements from other teams working on coverage measurement planning. This team is further tasked with developing the independent housing unit listing booklet (ILB), and housing unit followup (IHUFU) forms in order to measure housing unit coverage in 2008/2010.

*Highlights:* Staff analyzed and reported on the respondent debriefings conducted as part of the two Person Followup operational tests. Respondent debriefings were conducted in two small-scale field tests of the PFU operation. The field tests were conducted to test modifications to the form to assist interviewers in finding the most knowledgeable respondent for the interview. (This proved problematic in recent Census Tests.)

The first field test was conducted in March 2009 with expired Current Population Survey (CPS) sample in the Washington, DC metropolitan area. Results showed that the interviewers did not ask the scripted questions to determine whether the respondent was knowledgeable about where the followup person lived in 2008. Thus, the questions needed more revision. Following changes to the content and layout of the questions and also to the training of the interviewers, a second field test was conducted in August, 2009, again with expired CPS sample. Interviews were conducted in New York, NY and several counties in the Washington, DC metropolitan area. The revised questions were more successful. Interviewers asked them correctly and navigated properly through the form. The debriefing questions showed that proxy respondents did not have a good idea of the kinds of questions that would be asked. This occurred in instances where the proxy said they had heard of the follow-up person but did not know them well enough to answer questions; it also occurred when the proxy said they did know them well enough, and then did not have enough information to answer the questions. The results of these field tests will be incorporated in the 2010 PFU operation, which will be conducted in January 2011. Findings are documented in Nichols, B., and Childs, J. (2009). "2008 and 2009 PFU Operational Tests Respondent Debriefing Results." *DSSD 2010 Census Coverage Measurement Memorandum Series #2008-D7-17*, October 27, 2009.

Staff observed the 2010 Census Coverage Measurement listing operation. The observation is documented in Nichols, B. (2009). "Observation of the 2010 Census Coverage Measurement Independent Listing Operation in Prince Georges County, Maryland on October 27, 2009," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-F-14, November 17, 2009.

*Staff:* Beth Nichols (x31724), Jennifer Hunter Childs, Terry DeMaio, Nathan Jurgenson

### 1.7-1.8 COVERAGE IMPROVEMENT PLANNING AND DEVELOPMENT/ EVALUATION PLANNING COORDINATION (DECENNIAL PROJECTS 5610005 AND 5610006)

#### A. Development of Questionnaires for Decennial Coverage Improvement

*Description*: We will consult on the development of questions and questionnaires designed to improve within household coverage in the Decennial Census. We will participate in the development and pretesting of household and individual-level coverage questions in the decennial short form and the Coverage Followup (CFU) reinterview instrument.

*Highlights:* Staff assisted in the development of and completed cognitive testing on an alternative version of the Coverage Followup (CFU) instrument, called the Targeted CFU, which will be fielded in 2011 using cases that have been identified as duplicates in the 2010 Census. Staff presented results of the cognitive testing and corresponding recommendations to team. Staff also assisted the Decennial Statistical Studies Division in planning a large-scale cognitive test and qualitative study of the TCFU and duplicates in the 2010 Census.

*Staff*: Jennifer Hunter Childs (x34927), Matt Clifton, Anissa Sorokin, Nathan Jurgenson, Mikelyn Meyers, George Higbie, Lorraine Randall

#### **B. 2010 CPEX Experimental Overcount Booklet**

Description: The purpose of this project is to develop and test an alternative mailout census booklet with special coverage questions to compare to the standard census form in terms of coverage in the Census 2010 Alternative Questionnaire Experiment split-panel test. Both forms include a question asking whether each person in the household sometimes lives or stays somewhere else, and for what reason. On the standard census form, this question functions as a flag for later phone followup to get more complete coverage data. The alternative mailout booklet converts this question into a screener for a new set of questions on the mailout form itself to identify persons' alternative addresses and where to count them. If it works, the alternative approach has the potential to improve coverage as well as cut the costs and time involved in conducting followup operations.

*Highlights:* We wrote the draft final report for the Census 2010 CPEX Overcount Cognitive Testing Project and sent it to the sponsor for review and received comments. We are revising the report and will post it in our division's report series. In this project, we identified serious problems with the skip and the other address question and noted that the sequence may not work for tenuously attached persons with more than two addresses. We recommended format and wording changes, some of which were accepted by the Decennial Statistical Studies Division. We also documented that the experimental overcount question sequence we developed did succeed in identifying, from information on the census form itself, the correct residence for 75% of the persons in the household who sometimes live or stay somewhere else. This suggests that the experimental form may work well in a live census to resolve overcount coverage questions from data on the form itself, thereby reducing the costs and time of later phone follow-ups. The revised booklet will be sent out as an experimental panel in the 2010 Census "Avoid Followup Experiment." Results are in our draft final report, "Cognitive Testing of the Census 2010 Experimental Overcount Questions for the Census 2010 Alternative Questionnaire Experiment: Final Project Report."

*Staff*: Laurie Schwede (x32611), Anissa Sorokin, Virginia Wake Yelei

#### C. Evaluations, Experiments, and Assessments Operational Integration Team (EEA OIT)

*Description*: The purpose of the EEA OIT is to facilitate planning and timely implementation of 2008 Census Dress Rehearsal and 2010 Census evaluations, experiments, and assessments. The group guides and monitors the development, implementation, and reporting of the 2010 evaluations, experiments and assessments. It ensures that program integration and implementation of the 2010 Census Program of Evaluations and Experiments (CPEX) meets the guidance provided by the Census Integration Group and prepares and monitors the 2010 Census Program for Evaluations and Experiments Master Plan.

*Highlights:* Staff serves as an advisor to the Evaluations, Experiments, and Assessments OIT and as a Co-Advocate for Coverage Improvement Evaluations and Experiments. She attends meetings, giving feedback on other research related to CPEX evaluations, and keeping researchers in our division informed of DMD/DSSD updates.

*Staff*: Laurie Schwede (x32611)

#### **D. Evaluation of CCM Interviews**

Description: The 2010 Census Program for Evaluations and Experiments (CPEX) includes studies that focus on the quality of the data collected in Census Coverage Measurement Program (CCM). In particular, the focus is on two CCM interviews, the Person Interview (PI) and the Person Followup (PFU) in 2010. The primary methodologies used to evaluate the PI and PFU are respondent debriefing studies and recall bias studies. These studies will provide information about how well the CCM instruments capture the members of the household at each housing unit on CCM interview day and the usual residence of each household member and/or followup person on Census Day. The recall bias study also investigates the quality of the reporting of dates that respondents moved and the reporting regarding previous residents of the housing units. Additionally, these studies will highlight the causes and possible remedies within the questionnaire for any errors of usual residence and household membership.

*Highlights:* Staff drafted a study plan for the respondent debriefing project and assisted Decennial Statistical Studies Division (DSSD) in acquiring the 2006 Questionnaire Design and Experimental Research Survey (QDERS) specification, output programs and training. DSSD will use a modified 2006 QDERS instrument to conduct the recall bias study for CCM. Staff are in the initial planning stages for an independent memory study of landmark events. This will be a separate project not connected with the CCM recall bias study.

*Staff*: Beth Nichols (x31724), Mary Mulry, Jennifer Hunter Childs

#### E. Investigation of Study Methods for the Census Coverage Measurement (CCM) on Group Quarters (GQ) Population

*Description*: This project undertakes research and studies before and during the 2010 Census to ultimately develop potential methods for assessing the group quarters population coverage accuracy in the 2020 CCM program. Study methods for the 2010 research includes: field observations, in-depth interviews, focus groups, cognitive pretesting, ethnography, respondents debriefings, and a pilot small scale post-enumeration CCM-like survey with student population residing at university housing in 2010. Staff will document the success and difficulties for conducting a 2010 ethnographic study on the coverage measurement evaluation of each of the eight broad types of group quarters population and a pilot field test of a CCM-like survey with the student population.

*Highlights:* During the first quarter of FY 2010, staff completed 15 in-depth interviews with GQ administrators and the 2010 GQ project study plan. The plan was presented to and approved by the Census Integration Group. Staff recruited six university researchers to conduct the ethnography study of group quarters in the 2010 Census. Each researcher will be studying one specify broad types of group quarters.

*Staff*: Anna Chan (x38462), George Higbie, Temika Holland, Stephen Lubkemann

#### F. 2010 Census Language Study (CPEX)

*Description:* We will conduct systematic/structured observations of Nonresponse Followup (URFU) interviews in areas with heavy concentration of linguistically isolated (LI) households from various national origins. The aim of this research is to (a) observe how enumerators in the 2010 Census environment approach LI households, (b) observe what measures are taken by enumerators to collect the required census data from these households, and (c) based on our observations, determine what changes, if any, are needed to improve the conduct of in-person interviews with LI households.

*Highlights:* Staff developed the study plan for this research and received comments from critical reviewers on the study plan. Staff also worked on study design and theoretical framework to guide the research and conducted research to identify ethnographers to lead seven language teams to conduct the study.

Staff: Yuling Pan (x34950)

#### G. 2010 Census Behavior Coding Evaluation

*Description*: In order to learn how well census enumerators/ interviewers ask, and how well respondents answer, census questions, behavior coding studies will be conducted for all interviewer-administered instruments (e.g., NRFU, CFU, CCM) in 2010. The purpose is to calibrate how well survey instruments are administered by interviewers, and to identify problems with how interviewers ask and respondents answer questions. By conducting behavior coding for all intervieweradministered instruments, this study will tell us whether census questions are being asked as intended and will identify problems with the questions and with interviewer training. This study can further help the Census Bureau interpret apparent disparities in data that may arise between different operations.

*Highlights:* This quarter, staff completed study plans and schedules for the 2010 Behavior Coding of NRFU, CFU

and CCM PI. Staff also began coordinating data collection and analysis with other divisions.

Staff: Jennifer Hunter Childs (x34927), Nathan Jurgenson

#### H. Comparative Ethnographic Studies of Enumeration Methods and Coverage in Race/Ethnic Groups

*Description*: Staff will conduct comparative ethnographic research on enumeration methods and coverage in four to nine race/ethnic communities during Census 2010. The aim is to identify ways to improve census enumeration methods and coverage for race/ethnic populations, some of which have been categorized as hard-to-enumerate groups in previous censuses. This field study will involve accompanying enumerators to observe, tape, and debrief respondents during three 2010 operations involving personal visit census data collection: Update/Enumerate, Followup, and Census Nonresponse Coverage Measurement. We will identify and explore three sets of issues affecting the completeness and accuracy of the census: 1) enumeration methods, 2) questionnaire issues, and 3) residence rule/coverage issues. An additional component to explore factors respondents use in selfidentification of race is under consideration.

Highlights: The evaluation author drafted the official 2010 evaluation study plan and incorporated comments from supervisory reviewers as well as official and informal critical reviewers from Decennial Statistical Studies Division (DSSD), Population Division, Field Division (FLD), and our division. Staff had several meetings with DSSD reviewers to negotiate changes to the methodology to mitigate DSSD concerns about possible contamination of the independence of the census and CCM. We met with FLD reviewers to work out changes to the study plan to reduce possible researcher effects on interviewers that might alter their future behavior. We incorporated comments from all reviewers and sent the revised draft study plan to Decennial Management Division. We gave the official study plan briefing to the Census Integration Group and met again with DSSD to work out some remaining operational issues and compromises.

We used several methods to identify outside ethnographers to do our field studies. We talked with colleagues at meetings of the American Anthropological Association and the Washington Association of Professional Anthropologists. We arranged with outside colleagues to have a flyer posted on several professional practitioner websites. We sent the flyer out on multiple listservs to anthropologists and sociologists and we targeted information to some experts on very small race/ethnic groups to try to ensure we had enough applicants with previous research experience with those groups. We worked closely with DSSD sampling experts to get sampling lists for use in research site selection. We also worked with special data from a prior communications study to identify tracts with high Planning Data Base hard-to-count scores and high proportions of our target race/ethnic groups to aid in selecting the race/ethnic research sites for this study.

Staff: Laurie Schwede (x32611), Rodney Terry, Matt Clifton

#### I. Explaining How Census Errors Occur through Comparing Census Operations History with Census Coverage Measurement (CCM) Results

*Description*: The goal of this project is to help us understand what sorts of errors tend to be associated with the different Census operations. We want to merge Census files from the various stages of Census operations for a subsample of CCM areas and compare them to the CCM results. This comparison is intended to help find patterns of errors in Census operations and provide insights into ways to avoid these errors.

*Highlights:* Staff submitted a revised version of the evaluation schedule to Decennial Management Division (DMD). Staff also answered a DMD questionnaire on data sources for the evaluation.

Staff: Michael Ikeda (x31756), Mary Mulry

#### J. 2011 Relationship Survey

Description: Recent changes in the American legal and social landscape with respect to family composition, relationships, and same sex marriages have potential impacts on the content of the relationship and marital status questions in the censuses and surveys conducted by the Census Bureau. This project involves a program of research and testing that will guide the development of revised questions. Exploratory focus groups will be conducted across the United States with members of cohabiting couples to collect qualitative information about alternative terms, definitions, categories and/or questions that most accurately measure relationship status and partnership situations. Cognitive interviews will be conducted to evaluate questions developed based on the results of the focus groups. The revised questions will be subjected to further testing and evaluation before being implemented by the Census Bureau.

*Highlights:* Staff commented on and finalized the Statement of Work for a contract to conduct 17 focus groups in eight cities across the nation. We reviewed the nine proposals submitted and made recommendations for awarding the contract. The schedule calls for the contract to be awarded in early January 2010.

Staff: Terry DeMaio (x34894), Nancy Bates

## 1.9 AMERICAN COMMUNITY SURVEY (ACS) (Decennial Project 5385060)

#### A. ACS Missing Data and Imputation

*Description*: This project undertakes research and studies on missing data and imputation for the American Community Survey and aims to impute missing socioeconomic data in the National Assessment of Educational Progress (NAEP) data files using Census long form and American Community Survey (ACS) data.

*Highlights:* Staff provided basic programming support to the National Center for Education Statistics contractor.

*Staff*: María García (x31703), Yves Thibaudeau

## **B.** ACS Group Quarters (GQ) Item Imputation and Micro Data Disclosure Avoidance Research

*Description*: American Community Survey group quarters microdata and tabulations are protected from identity disclosures via synthetic data methods. This project coordinates staff in our division, Decennial Statistical Studies Division (DSSD), Population Division (POP), and Housing and Household Economic Statistics Division (HHES) to generate production code (in the R language) for this purpose. Staff will also ascertain the effectiveness of using synthetic data methods as an alternative to hot deck allocation in ACS group quarters.

*Highlights:* Staff altered ACS group quarters disclosure avoidance program to operate via a configuration script, incorporated DISCRETE editing program to ensure edit consistency of synthetic data, and implemented an improved flagging procedure for identifying at-risk records.

*Staff*: Laura Zayatz (x34955), Paul Massell, Rolando Rodríguez, Jason Lucero, Asoka Ramanayake, Lisa Singh, Bimal Sinah, Tapan Nayak

#### C. 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.

*Highlights:* Staff organized a new project to examine discrepancies between multi-year estimates and related single year quantities. Staff also consulted with staff from the Veterans Administration regarding establishing benchmark time periods for multi-year estimates.

*Staff*: Tucker McElroy (x33227), Natalya Titova, Chaitra Nagaraja

#### **D. ACS Variances**

*Description*: Work under this heading this year concerned two research projects: (i) Completion of a project providing design-based superpopulation consistency theory along with linearized variance formulas for variances of estimators of totals from complex surveys to which replication methods like BRR can be compared. (ii) Development of a method of simultaneous nonresponse adjustment, calibration to achieve population controls, and weight smoothing or truncation.

*Highlights:* This quarter, staff prepared an ASA Proceedings paper out of the JSM 2009 paper, "Simultaneous calibration and nonresponse adjustment." An expanded paper with the same title was also produced for the division's *Research Report Series*.

Staff also reviewed the draft final report (authors M. Beaghan, M. Asiala and K. Albright) for The Alternative Population Controls project in which staff participated through much of the preceding year.

*Staff*: Eric Slud (x34991), Yves Thibaudeau

# E. ACS Small Area Estimation for Selected Characteristics

*Description:* This project aims to propose, develop, and evaluate small area estimation methodologies to produce ACS estimates for selected characteristics in geographies with small populations. The characteristics of initial interest are unemployment, income, and poverty.

*Highlights:* An interdivisional group was formed to address this topic. Meetings were held to present and discuss background material. The group was divided into two separate subgroups due to the different requirements for small area estimates of Group Quarters and the total population. A staff member is looking into what variables might be available for use from the StARS and the LEHD administrative record systems.

Staff: Lynn Weidman (x34902), Patrick Joyce, Don Malec

# F. ACS Small Area Estimation for Group Quarters (GQ)

*Description:* This project aims to propose, develop, and evaluate small area estimation methodologies to produce ACS estimates for the GQ population (totals and characteristics) for substate geographies, including counties, places, block groups, and tracts.

*Highlights:* An interdivisional group was formed, initially to investigate methods to be applied with ACS estimation processing in the spring of 2011. The mandate for the group changed to developing and preparing a method for implementation for 2010 production of estimates for the 2005-2009 GQ population. The goal of the methodology is to impute person records and corresponding weights into enough GQ facilities not in sample so that at least one GQ of each major type per tract including that major type contains data. Furthermore, the method must fit into the existing ACS system for producing estimates and variance estimates.

The methodology will be tested and evaluated using simulated samples from the Census 2000 GQ records that contain only short form variables. Software was written to select 25 4-year samples using ACS GQ sampling specifications. Evaluation will look at the properties of estimates of demographic characteristics across these samples. The sample simulation includes assigning a number of people to each GQ in each year. These numbers are assigned randomly based on the relationship between expected size and actual size determined from analysis of data on ACS GQ sampling universe and weighting files. The donor pool for a GQ requiring imputation includes all persons from GQs of the same specific type that are geographically 'close', or if no such GQs are available, from GQs of the same major type. Some of the relevant pieces of software have been tested and run in a reasonable amount of time.

*Staff:* Lynn Weidman (x34902), Chandra Erdman, Patrick Joyce, Chaitra Nagaraja

#### G. ACS Data Issues

*Description:* Various issues related to the quality and presentation of ACS estimates were discussed and investigated by small interdivisional teams or division staff. The goal of these investigations was to make recommendations to aid in resolving the issues.

*Highlights:* Data analysis was completed for the draft report "Developing Guidelines Based on CVs for when Three-Year Estimates Can Be Used Instead of Five-Year Estimates in the American Community Survey." This extends previous work which looked at ACS one-year estimates.

A staff member participated in discussions of an interdivisional team on the next steps in testing a data quality indicator and whether and how it could be implemented in the new version of the American FactFinder.

A staff member reviewed material on multiple comparisons and drafted a technical document describing what procedures should be used for various comparisons and tests using ACS data. Several examples of use of the methods were proposed and a request for data with specified characteristics to be used in these examples was prepared.

Staff: Lynn Weidman (x34902), Michael Ikeda

### 1.10 AMERICAN COMMUNITY SURVEY (ACS)/METHODS PANEL (Decennial Project 5385095)

#### A. ACS Language Research

*Description*: This project provides technical and research support for addressing language issues in ACS data

collection instruments and supporting documents. Staff members serve on inter-divisional working groups and provide consultation and technical support in the design and development of language research for the ACS.

*Highlights:* Staff members continue to be active members of the ACS Language Team and the ACS Content Council.

During the first quarter of FY 2010, staff worked closely with the ACS Language Team to plan and review Statements of Work for three new projects: 1) Translation review and cognitive testing of translated ACS language assistance guides in Chinese and Korean, 2) Project D, Part II, which is the cognitive testing of the remaining half of the Spanish CATI/CAPI instrument, and 3) cognitive testing of the ACS CAPI letters in five languages (Project C). Staff worked as Census Bureau research analysts, providing technical guidance for the contractor (RTI). Staff participated in the kickoff meeting for these three projects with the contractor and the ACS language team members to provide specific guidance and requirements for the project. Staff reviewed and provided comments on the recruitment plans, drafted English cognitive interview protocols, and translated cognitive interview protocols in two languages (Spanish and Chinese), and Spanish and Chinese translated questions for these three projects. Staff also worked on developing a sociolinguistic framework for translation-review and on principles of influence from a psychological perspective for ACS letters and interview recruitment.

In addition, staff continued to work as reviewers and technical support for the cognitive testing of the Spanish and English content of the upcoming ACS Content Test. During this quarter, staff attended regular meetings about the project schedule, methodology and progress.

Finally, staff developed and presented a seminar on translation methodology to the ACS Research Team, summarizing findings of the multilingual cognitive testing projects.

*Staff*: Yuling Pan (x34950), Patricia Goerman, Leticia Fernández, Virginia Wake Yelei, Rodney Terry, Matt Clifton, George Higbie

#### **B. ACS Data Reliability Indicator Project**

*Description*: The usability team designed a series of usability evaluations of a new method of displaying the ACS data tables. The new feature to be tested was a color-coded indicator of the reliability of the data. The purpose of the testing was to examine how well the datareliability indicator worked for users (especially as compared to the current ACS data tables without the indicator) and to identify any problems that actual users might have with the data tables. The data reliability indicator was based on the Coefficient of Variation (CV), which is defined as the standard error of an estimator divided by the estimate. Another purpose of this testing was to examine whether users would notice and use the Margin of Error (MOE) when answering questions about the estimates from the table. This second testing goal was based on the observation that although the MOE is currently provided with each estimate, the MOE is routinely ignored by ACS data users.

*Highlights:* Two rounds of usability testing have already been completed and the third is planned for the second quarter of FY2010.

Staff: Kathleen Ashenfelter (x34922), Victor Quach

#### **C. ACS Messaging Project**

*Description*: The purpose of this project is to develop and test new messages on ACS letters and a brochure to alert ACS respondents in 2010 that they are required to respond to both ACS and census questionnaires. In 2000, ACS response rates were affected by the 2000 Census environment. Until March 2000, ACS response rates rose as a result of census publicity, but they fell for the rest of the year after respondents also received their census forms, particularly around Census Day. The aim of this project is to try to avoid these drops in response rates in 2010 by informing ACS respondents that they will be receiving both forms and need to complete both.

*Highlights:* Staff attended ACS Messaging Project team meetings. One staff member did additional analysis and gave a presentation to the team, "More Results on the ACS Messaging Project." Using observational data collected during the cognitive interviews, this report revealed strong associations between how respondents removed materials from the initial cover letter package and the extent to which they found and read the cover letter in the package, as well as other findings from this study. This paper forms the basis of our new 2010 AAPOR paper, "Do Respondents Read Those Key Messages in our Questionnaire Package Cover Letters?: What Comes Out of the Envelope First Matters."

Staff: Laurie Schwede (x32611), Anissa Sorokin

#### **D.** ACS Internet Testing – Usability Input

*Description*: The usability team is leading our division's contribution to the development of an online instrument for the American Community Survey. The multi-year project will consist of several rounds of usability testing of prototypes of the internet instrument. Staff will provide test plans and formal reports for each round of testing as well as provide input at regular team meetings during the development of the ACS online instrument and its subsequent field testing.

*Highlights:* The letters and wireframe screens for this project were nearing completion at the end of the first quarter. Combined usability and cognitive testing is planned for the third quarter.

*Staff:* Kathleen Ashenfelter (x34922), Victor Quach, Larry Malakhoff

#### **E.** ACS Internet Testing – Cognitive Input

*Description*: The cognitive lab is participating in the development of an online instrument for the American Community Survey. The multi-year project will consist of questionnaire development and several rounds of cognitive and usability testing of prototypes of the internet instrument in both English and Spanish. Staff will provide questionnaire development and mode consistency expertise at regular team meetings during the development of the ACS online instrument and its subsequent field testing. Staff will also participate in the joint cognitive/usability testing sessions led by the usability lab.

*Highlights:* Staff provided expert guidance on question wording and mode consistency at weekly development meetings. Staff prepared experimental versions of roster screens for the instrument for cognitive and usability testing and assisted preparing the cognitive and usability test plan. Staff also began working on a CATI Reinterview survey for the internet experiment.

*Staff*: Jennifer Hunter Childs (x34927), Elizabeth Nichols, George Higbie

#### F. ACS Internet Test Experimental Design Team

*Description*: Staff is contributing methodological expertise and input to the development of the experimental design that will be used for the field testing of the ACS Internet form that is currently being developed and is planned for early 2011. The design of this methodology includes such considerations as sampling, pre-notification letters, mailing schedule, panel design, and planned analysis of the results.

*Highlights:* Meetings for this team concluded during the first quarter of FY 2010 because the experimental design of the Internet test is complete.

*Staff*: Kathleen Ashenfelter (x34922)

## G. Iterative Testing of the American Community Survey Web Site

*Description*: The American Community Survey (ACS) Web site is making major re-designs to their interface. The usability lab has been asked to participate in the iterative user-centered design development of the site. This effort encompasses the full spectrum of usercentered design activities, from iterative low-fidelity paper prototype testing to high-fidelity testing with a working prototype. The purpose of the testing is to identify usability issues. Recommendations made to resolve the issues are intended to improve the usability of the Web site for all users.

*Highlights:* Staff began the process of recruiting users and identifying appropriate tasks for the first click study.

Staff met with ACSO staff to discuss the upcoming study and to review the prototypes.

Staff: Erica Olmsted-Hawala (x34893), Jennifer Chen, Temika Holland

## 1.11 CURRENT POPULATION SURVEY (CPS) / ANNUAL SOCIAL AND ECONOMIC SUPPLEMENT (ASEC) TABLES (Demographic Project TBA)

*Description*: Staff provided technical consultation services and programming support for the redesign and content of SAS programs that produce the table packages for the 2007 Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) that will feature information at the national and regional levels for special population/topics.

*Highlights:* Staff updated and tested table shell generating software in Visual BASIC. Staff coached personnel in the Population Division to generate more than one hundred new 2009 race, older, and gender tables accurately and efficiently.

Staff: Aref Dajani (x31797), Tom Petkunas

## 1.12 DEMOGRAPHIC SURVEYS DIVISION (DSD) SPECIAL PROJECTS (Demographic Project 0906/7374)

#### 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:* Using the Matcher, staff linked files from the Publically Purchased Lists to the PUMS file (Public use Microdata file from the ACS) to look for outliers based on similar blocking criteria. The Matcher was modified to produce outliers (block sizes were less than five) and not output every match. The outliers should in theory be easier to attack. Staff set up spider programs to gain additional knowledge of individuals in the PUMS file.

Staff: Ned Porter (x31798), Lisa Singh, Rolando Rodríguez

# **B.** Using Survey Paradata to Manage Surveys in the Field and Estimate Survey Error

*Description:* This project seeks to understand how paradata (survey process data) are currently used and where they are stored throughout the Census Bureau,

particularly in the Demographic and Field Divisions. The broader goal of the project is to modernize project management through the use of graphical representations of paradata that are displayed in interactive, real-time dashboards. This improvement in paradata access will allow managers to make quick and better project management decisions, and will equip the Census Bureau with the ability to carry out responsive design.

*Highlights:* Staff has been gaining familiarity with Census Bureau survey tracking and reporting systems for production, sampling, and financial information by meeting with people who maintain and use these systems. Staff is also gaining familiarity with current uses of paradata to manage surveys, specifically the Current Population Survey (CPS).

Staff reviewed literature on paradata use in survey research and use of operations data for error reduction more broadly, and attended survey cost estimation training for survey managers presented by the Field Division (FLD).

*Future Plans:* Staff plans to use IBM Cognos to reproduce some of FLD's CPS reports (used by Ning), which are currently done in Excel, and to get input from survey managers in DSD, managers in FLD, and Regional Office staff regarding their paradata needs for survey management. From these requests, staff will prototype and test the use of different charts and dashboards with various users.

*Staff:* Matt Jans (x36724), Kathy Creighton (Contractor, DSD), Chris Laskey (DSD), Cheryl Landman (DSD), Chris Stringer (DSD)

# C. Usability Testing of the National Survey of College Graduates

*Description:* The National Survey for College Graduates (NSCG) is an online data collection Web site that collects education and job information from diverse users who have received bachelor's degrees from American schools or abroad.

*Highlights:* Staff evaluated the usability of the NSCG online survey. The testing evaluated the success and satisfaction of eight participants who attempted to complete the survey developed by the Special Surveys Branch (SSB) of the Demographic Surveys Division (DSD). Usability testing revealed a few usability problems, including illogical placement of survey items and difficulty entering information. Changes are currently being made to the instrument, and a second round of testing is expected to begin in February 2010.

Staff: Jennifer Romano (x33577), Jennifer Chen

### 1.13 QUICK TURNAROUND PRETESTING OF HOUSEHOLD SURVEYS (Demographic Projects 1465001)

*Description*: This project involves pretesting new or revised series of questions for insertion into household surveys. The projects are of the short-term, quick turnaround variety rather than long-term research efforts to redesign a survey. Methods used include cognitive testing and other techniques as appropriate.

#### A. Rental Housing Finance Survey

*Description*: This project involves cognitive testing of a new questionnaire that collects information from property owners and managers of multi-family housing units.

*Highlights:* Staff held several meetings with staff from HHES, HUD, and DSD to discuss the scope of the project. Cognitive interviews are scheduled to begin in July.

*Staff*: Terry DeMaio (x34894)

### **B.** National Crime Victimization Survey (NCVS)

*Description:* The NCVS asks respondents to report the race of the offender for any crimes committed against them. Currently the response categories (White, Black, Other) do not match the categories included in Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. This project includes a review of survey methods and social scientific literature relevant to identification of the race of offenders, as well as designing and implementing experimental research on this topic.

*Highlights:* Staff met with staff from the Demographic Surveys Division and the Bureau of Justice Statistics to discuss the outlines of the project and the scope of the literature review.

Staff: Terry DeMaio (x34894), Jen Beck

**C. Development of the CARI Behavior Coding System** *Description:* This project involves consultation with staff from DSD, DSMD, DSSD, TMO, FLD, ACSO and RTI International, which is the contractor developing the behavior coding component of the CARI system. This program will enable interviews, or snippets of interviews, to be tape recorded as they are conducted and allow behavior coders to listen to sound files and see the interviewer's entries as they code the interactions between the interviewer and respondent using an automated system. The system will also output data to SAS and Excel for behavior coding analysis.

*Highlights:* Staff attended 4 joint application design sessions with staff from RTI International and other participating divisions, at which the details of the behavior coding system were discussed and decisions

were made about how the instrument would be developed. We also participated as active members of the CARI Development Team by attending meetings, commenting on documents, and responding to requests.

*Staff:* Terry DeMaio (x34894), Jenny Hunter Childs, Joanne Pascale, Jen Beck

### 1.14 RE-ENGINEERED SURVEY OF INCOME AND PROGRAM PARTICIPATION RESEARCH (Demographic Project 1465444)

#### **Re-Engineered SIPP Methodological Research**

*Description*: The re-engineered Survey of Income and Program Participation (SIPP) is scheduled to replace the current SIPP in 2013. This project conducts long-term methodological research to evaluate SIPP and to inform the design of re-engineered SIPP instruments and procedures.

Highlights: Most of the work on this project this quarter was directed toward finalizing plans and procedures for the 2010 SIPP-EHC CAPI field test ("EHC" meaning Event History Calendar), an effort of the re-engineered SIPP Research Group, which is chaired by our division. Key tasks involved procedures for recording a small sample of interviews, transcribing those recordings, which will be carried out by Tucson Telephone Center staff, finalizing details concerning interviewer training, and observing initial training sessions in various Regional Offices. The Research Group also began to look ahead to the next research activity, which will be conducted in early 2011. Key issues to be addressed in upcoming research include mover/attrition issues, and the functionality of wave 2 (and beyond) instruments and procedures.

In addition, staff presented the results of the 2008 "paper" field test at a meeting of the American Statistical Associations' Survey Research Methods Section SIPP Working Group: "A Comparison of Survey Reports Obtained via Standard Questionnaire and Event History Calendar: Initial Results from the 2008 EHC 'Paper' Test."

Staff: Jeff Moore (x34975), Anna Chan, Joanne Pascale

## 1.15 DATA INTEGRATION DIVISION (DID) SMALL AREA ESTIMATION PROJECTS (Demographic Project 7165000)

# A. Research for Small Area Income and Poverty Estimates (SAIPE)

*Description*: The purpose of this research is to develop, in collaboration with the Data Integration Division (DID) (The Small Area and Poverty Estimates Branch was

previously in Housing and Household Economic Statistics Division and is now in DID), 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. with documentation of along the methodology.

*Highlights:* Due to the observed biases in both the design based point and variance estimates of the log number of poor children for small sample sizes from the previous simulations study using ACS 2005 data, the staff is exploring several different model forms for county children poverty rates, including beta-binomial and normal mixtures with nonlinear mean functions. The models include a mixture of binomial distributions having the feature that the "sample size" n is set so that the variance of the binomial distribution matches the design-based variance (or the modeled "general variance function" variance). This may cause technical issues in estimation, which the staff is currently investigating.

The previous work on creating the model framework for the design based variance estimate of the log total number of school age children in poverty can be used for the design based variance estimates of the poverty rate with minor modifications. The bootstrap simulation computer code using the ACS 2005 data was modified to produce information about statistical properties of the variance estimate of the poverty rate.

Staff did a simulation study of Fay's replicate weight variance estimate of the child poverty rate using random samples of various sample sizes drawn from the artificial populations constructed from various Bernoulli and Poisson distributions and the ACS 2005 data from the five largest counties of Maryland. We examined the simulation results to see whether the distribution of the variance estimates of poverty rate can be approximated by a scaled chi-squared distribution and, if so, with what degrees of freedom. We obtained various measures (Kolmogorov-Smirnov statistics, degrees of freedom, coefficient of variation, relative bias) to evaluate the distribution of the variance estimates of poverty rate from the simulated samples. It shows that the measures vary by the population parameters and sample sizes. The Fay's variance estimator of poverty rate is significantly underestimated for small sample sizes. However, the degrees of freedom for the cases considered are all less than the number of the replicates 80 used in ACS Fay's replicate weight variance estimates.

*Staff*: Elizabeth Huang (x34923), Jerry Maples, William Bell (DIR)

#### **B. Small Area Health Insurance Estimates (SAHIE)**

*Description*: At the request of staff from the Data Integration Division (DID), 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 DID.

*Highlights:* Staff continued developing a new approach to estimating lower level (state level, county level) estimates using a linear transformation of the lower level parameters and drawing new parameters using the transformed conditional distribution and a random variable drawn from the higher level distribution. Two questions that were raised have been resolved. We are able to show that when there is only one linear constraint the distribution of the new lower level parameters is independent of the choice of transformation. We are also able to show that if the original parameters are nonnegative, then the new lower level parameters will be nonnegative.

Since the aim of this procedure is to use as much of the SAHIE MCMC estimation algorithm as possible, we investigated incorporating the constrained estimates back into the lower-level Markov chain versus using the lower level Markov chain to output constrained values separately at each step. By examining the distributional assumptions needed, we determined that incorporating the constrained estimates back into the model would require remodeling the lower-level model with additional constraints, throughout, and would radically change the current MCMC approach. It was decided to continue development of the other approach to use the current unconstrained lower-level model to create constrained estimates.

Staff: Don Malec (x31718), Ryan Janicki

### 1.16 EDITING METHODS DEVELOPMENT (ECONOMIC PROJECT 2370054)

# Investigation of Selective Editing Procedures for Foreign Trade Programs

*Description*: The purpose of this project is to develop selective editing strategies for the U.S. Census Bureau foreign trade statistics program. The Foreign Trade Division (FTD) processes more than 5 million transaction records every month using a parameter file called the Edit Master. In this project, we investigate the feasibility of using selective editing for identifying the most erroneous records without the use of parameters.

*Highlights:* In this project, we research methods to identify errors in our foreign trade data without the use of the edit master parameter file. Staff investigated the application of selective editing techniques to all records (not only rejects). We continued research and

development of score functions including an effect term to estimate the influence on the estimates of changes on Value and Quantity of shipments as is appropriate for these data. We revised the implementation code and testing to fine-tune program parameters. We wrote a draft of a report titled "Score functions for the U.S. Census Bureau Trade Data," for inclusion in the division's *Research Report Series*.

*Staff*: María García (x31703), Yves Thibaudeau, Rachelle Reeder (FTD)

### 1.17 DISCLOSURE AVOIDANCE METHODS (Economic Project 2470051)

*Description*: The purpose of this research is to develop disclosure avoidance methods to be used for Census Bureau publicly available economic data products. Emphasis will be placed on techniques to implement disclosure avoidance at the stage of data processing. Disclosure avoidance research will be conducted on alternative methods to cell suppression for selected economic surveys. We will also aid in the implementation of the methods.

*Highlights:* Almost all effort for the Economic Directorate was concentrated on a single major project: cell suppression modernization. A Research and Methodology (R&M) team was formed that reviewed a large number of reports and articles that have been written over the past 20 years on the subject of cell suppression. At the R&M team meetings, both theory and implementation issues were discussed. Staff served as a co-leader of the R&M team, and, in that capacity, wrote several papers that were handed out at the weekly meetings. One of the important handouts was entitled "Understanding the Network Model for Tables."

*Staff*: Laura Zayatz (x34955), Asoka Ramanayake, Jason Lucero, Paul Massell, Tapan Nayak, Bimal Sinha

## 1.18 TIME SERIES RESEARCH (Economic Project 2370052)

#### A. Seasonal Adjustment Support

*Description*: This is an amalgamation of projects whose composition varies from year to year, but always includes maintenance of the seasonal adjustment and benchmarking software used by the Economic Directorate.

*Highlights:* Seasonal adjustment and X-12-ARIMA support was provided to Bundesbank, Reserve Bank of Australia, Credo Consulting, Haver Analytics, Ford Corporation, Railway Association of Canada, SAS, Bureau of Labor Statistics, U. S. Department of Agriculture, Australian Bureau of Statistics, Office of

National Statistics (UK), Statistics Canada, Central Statistical Office of Ireland, Statistics Austria, Korean National Statistical Office, Statistics Netherlands, Statistics South Africa, University of El Manar (Tunisia). Staff prepared comments on report by Irene Brown (OSMREP) on results from an attempt to implement the Causey-Trager benchmarking method in SAS.

*Staff:* Brian Monsell (x31721), Natalya Titova, David Findley (consultant)

# **B.** Seasonal Adjustment Software Development and Evaluation

*Description*: The goal of this project is a multi-platform computer program for seasonal adjustment, trend estimation, and calendar effect estimation that goes beyond the adjustment capabilities of the Census X-11 and Statistics Canada X-11-ARIMA programs, and provides more effective diagnostics. This fiscal year's goals include: (1) developing a Windows programming interface for the X-12/X-13 seasonal adjustment software in collaboration with analysts from the Bank of Belgium; (2) finishing a version of the X-13ARIMA-SEATS program with accessible output and improved performance so that, when appropriate, SEATS adjustments can be produced by the Economic Directorate; and (3) incorporating further improvements to the X-12-ARIMA/X-13A-S user interface, output and documentation. In coordination and collaboration with the Time Series Methods Staff of the Office of Statistical Methods and Research for Economic Programs (OSMREP), the staff will provide internal and/or external training in the use of X-12-ARIMA and the associated programs, such as X-12-Graph, when appropriate.

*Highlights:* Staff repaired a defect in X-12-ARIMA to ensure that the A18 table (original series adjusted for calendar effects) had leap-year preadjustment factors removed from it, and tested the program to ensure the table was produced correctly for a number of scenarios

Staff released an updated version of X-12-ARIMA, Build 188 of Version 0.3, to the Economic Directorate for their testing and then to the general public. Staff compared adjustments from this version of the software to the last released version of X-12-ARIMA (Build 177) and found in most cases no differences in the adjustments, and small differences between compilers on Linux based machines. Staff also provided support for analysts and programmers in the Economic Directorate in their testing of the new release.

Staff prepared and taught a seasonal adjustment class using X-13ARIMA-SEATS at Eurostat.

Staff assisted in determining how length-of-month adjustment factors could be generated for the regCMPNT program.

*Staff:* Brian Monsell (x31721), Tucker McElroy, Natalya Titova, Bill Bell (DIR), David Findley (consultant)

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

Description: The main goal of this research is to discover new ways in which time series models can be used to improve seasonal and calendar effect adjustments. An important secondary goal is the development or improvement of modeling and adjustment diagnostics. This fiscal year's projects include: (1) continuing research on seasonal adjustment diagnostics; (2) studying further the effects of model based seasonal adjustment filters: (3) examining goodness of fit diagnostics for time series modeling and signal extraction; (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; (5) studying the modeling of seasonality using Bayesian methods, and determining if using such a method is feasible for short time series; (6) studying the modeling of stock holiday and trading day on Census Bureau time series; (7) examining approaches for modeling time series with heteroskedastic errors.

Highlights: During this quarter, staff (a) continued empirical studies of model-based seasonal adjustment diagnostics, (b) developed algorithms and code for estimation of seasonal long memory models, (c) studied Markov Chain Monte Carlo techniques for Bayesian estimation of seasonal time series models, (d) studied properties of parameter estimates for mis-specified models when sampling error or cyclical effects are present, (e) compared performance of X-12-ARIMA on different compilers, (f) studied performance of new distribution theory for Ljung-Box statistics and (g) studied the performance of benchmarking methods with graphics developed by Statistics Canada. Ongoing research includes: (a) examining the effects of model mis-specification on seasonal adjustment, (b) exploring seasonal adjustment for long memory models.

*Staff:* Tucker McElroy (x33227), Brian Monsell, Christopher Blakely, Ekaterina Sotiris, Natalya Titova, Bill Bell (DIR), David Findley (consultant)

#### D. Supporting Documentation and Software for X-12-ARIMA and X-13A-S

*Description*: The purpose of this project is to develop supplementary documentation and supplementary programs for X-12-ARIMA and X-13A-S 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 documentation of X-12-ARIMA, improving the documentation of X-12-ARIMA, rendering the output from X-13A-S accessible, and exploring the use of component and Java software developed at the National Bank of Belgium. *Highlights:* Staff released an updated version of the Genhol utility to the public, with expanded documentation for the utility.

Staff updated the Seasonal Adjustment papers site and the X-12-ARIMA site, uploading new versions of the Win X-12 and X-12-Graph programs developed by the Time Series Methods Staff (OSMREP) as well as X-12-ARIMA and Genhol.

*Staff:* Brian Monsell (x31721), Tucker McElroy, Natalya Titova, David Findley (consultant)

## 1.19 SURVEY OF RESEARCH AND DEVELOPMENT IN INDUSTRY, IMPUTATION AND SAMPLING RESEARCH AND SOFTWARE DESIGN (Economic Project TBA)

*Description*: This project undertakes research on the imputation of unreported mandatory items in the Survey of Research and Development in Industry. It also examines what estimators are more appropriate under alternative sampling plans; in particular, it evaluates using calibration estimators to compensate for missing data. The possibility of extending calibration to new sampling plans, such as balanced sampling, is investigated. Both traditional linear regression techniques and nonparametric regression techniques are examined.

*Highlights:* Staff implemented the Steps for the imputation method proposed by Kleing and Shao for SRDI. This was achieved by translating the method in the language of an algorithm. The algorithm processes the imputation longitudinally, starting at the current time period (wave) conditional on information on the previous wave, when it is available. If information from the wave directly preceding the current wave is not available, then this information is imputed, in the same fashion. The algorithm is as follows:

(1) Impute missing values at time t = 4 (2005)

(a) Step 1: r = 3. Impute records with response in 2004, missing in 2005. Use records with response in 2004 and 2005 to \_t imputation model.

(b) Step 2: r = 2. Impute records with response in 2003, missing in 2004 and 2005.

Use records with response in 2003 and 2004, missing in 2005 to \_t imputation model.

(c) Step 3: r = 1. Impute records with response in 2002, missing in 2003, 2004, and 2005. Use records with response in 2002 and 2003, missing in 2004 and 2005 to t imputation model.

(2) Impute missing values at time t = 3 (2004) in the same fashion.

(3) Impute missing values at time t = 2 (2003).

Staff: Yves Thibaudeau (x31706), Martin Klein, Jun Shao

## 1.20 REMOTE ACCESS - MICRODATA ANALYSIS SYSTEM (Strategic Planning and Innovation Project 0359999)

Description: Researchers and sophisticated data users' demand for Census Bureau microdata, both for general research and programmatic needs, continues to grow. Microdata allows virtually any type of analysis, and it is the desired form of data that allows modeling. Internal Census Bureau microdata files contain levels of detail. and variables, which are not available in public use files. Methods are applied to reduce detail, both by suppressing and coarsening variables in public use files, in order to protect the identity of respondents and to ensure confidentiality of responses under Title 13 of the U.S. Code. As data on individuals accumulate, and identifiable public and commercial data becomes more and more accessible, the ability to publish quality microdata while maintaining a sufficient level of ambiguity is becoming an issue.

*Highlights:* Staff continues to work with members of the Data Integration Division (DID) on the development of a new Advanced Query System (AQS) / Microdata Analysis System (MAS). Staff members have written and improved an R program to generate scatterplots of synthetic fitted values vs. synthetic standardized residual values, which are designed to mimic the patterns shown in the scatterplots of the real fitted values vs. the real standardized residual values. Staff members are currently writing a memo that documents the algorithm to generate synthetic fitted values vs. synthetic standardized residual values. Staff members have begun to explore possible synthetic data methods to create model fit diagnostic plots for binary and multinomial logistic regressions.

Staff members have found a possible solution to the differencing attack problem on the MAS. A differencing attack attempts to rebuild a confidential microdata record on the MAS through submitting a request for an m-way cross-tabulation T[] for two similar universe data sets: U(n) and U(n-1), where U(n-1) has the exact same n observations as U(n), less one unique observation. As a result, the two similar m-way tables T[U(n)] and T[U(n-1)] will be exactly the same, less one unique cell count. The difference T[U(n)] - T[U(n-1)] = T[U(1)], where U(1) = U(n) - U(n-1) is a universe that is based on only one unique observation, and T[U(1)] is an m-way table of counts that contains a count of 1 in the cell that represents the unique partial microdata record for the one observation contained in U(1), and all remaining cells equal to zero.

The original Drop q Rule was an initial confidential rule used to introduce an amount of uncertainty to the results obtained through a differencing attack. The original Drop q Rule randomly removed a fixed q observations at random from the original universe data set U(n) to yield a

subsampled universe U(n-q). If the same U(n) was selected again by the same user, or by a different user, then the exact same q observations would be removed from U(n) to yield the same U(n-q) subsampled universe as before. Therefore, if users attempted the differencing attack of T[U(n)] - T[U(n-1)], they would actually be performing T[U(n-q)] - T[U(n-1-q)], where U(n-q) and U(n-1-q) are two independently subsampled universe data sets.

Staff members have provided an evaluation of the effectiveness of the original Drop q Rule in preventing differencing attack disclosures, and found that the original Drop q Rule didn't provide as strong protection as was once believed. As a result, staff members have developed a modified Drop qv Rule. In this subsampling procedure, a Qv = qv is first randomly drawn from a discrete uniform {2,...,k} distribution, then qv observations are randomly removed from the U(n) data set to yield a new subsampled universe data set U(n-qv). Once again, if the same U(n) is selected again by the same user, or by a different user, then the exact same Qv = qv observations are removed from U(n) to yield the same U(n-qv) as before. It was found that this modified Drop qv Rule provided much stronger protection against differencing attack disclosures than the original Drop q Rule.

Staff members are in the process of writing a memo that documents how an m-way table of counts can be reconstructed on the MAS through submitting two separate queries: one for a binary or multinomial logistic regression analysis, and one for a cross tabulation for a m-1 dimensional table. The results obtained from these two separate queries can be used to reconstruct an m-way table of counts.

Staff members continue to work on a cutpoint program for the MAS. The cutpoint program is used to reduce disclosure risk by bucketing observations into bins. We require that each bin contain at least 80 observations. While programs like this have been developed before, there has not been a systematic study of the strengths and weaknesses of each binning strategy. To date, we have programmed 3 different strategies. During the summer we generated a set of cutpoints using a frequency based strategy for generating the bins. This is currently what is used in the MAS. We are replacing that with a strategy based on recursive partitioning. The code is complete and the new cutpoints will be added to the MAS in January. At this stage, it is unclear that this strategy is optimal. Therefore, we will continue to program other strategies and evaluate their strengths and weaknesses on variables with different frequency distributions.

*Future Plans:* Staff will continue to work closely with DID staff to develop the beta AQS/MAS prototype. Staff will begin thinking of possible ways to test the confidentiality rules and routines within the new beta AQS/MAS prototype for universe formation and ordinary

least squared regressions as well as binary and multinomial logistic regressions. Staff will work with members of the Human Factors & Usability Research Group and DID staff to develop, test and evaluate a user interface for the beta AQS/MAS prototype. Staff will continue to explore possible model fit diagnostic plots based on synthetic data methods for logistic regressions, and explore possible ways to develop matrix scatter plots based on synthetic residual methods.

*Staff*: Laura Zayatz (x34955), Asoka Ramanayake, Jason Lucero, Paul Massell, Marlow Lemons, Vanessa Patterson, Lisa Singh

## 1.21 PROGRAM DIVISION OVERHEAD (Census Bureau Project 0381000)

#### A. Division Leadership and Support

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

*Staff*: Tommy Wright (x31702), Tina Arbogast, Robert Creecy, Matt Gore (HRD), Michael Hawkins, Gloria Prout, Stephanie Sheffield, Kelly Taylor

### **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 this quarter, staff completed the planning phase for the blade migration. The configuration for four identical IBM LS42 blade servers, each with four quad-core processors and 128 GB of RAM and associated storage was finalized, and was approved by the Blade Architecture Review Board. CSVD is expected to build and deliver the servers for configuration and testing in early January, 2010. The current plan is to migrate from our current SGI Altix server (research1) to the new blade cluster before the end of FY2010. We have begun working with TCO to assign the necessary attributes to existing users and groups to facilitate the data migration.

The replacement of staff desktop computers that began in the fourth quarter of FY09 continues. Once it is complete, all staff desktop computers in SRD will meet or exceed the current desktop standard.

Staff: Chad Russell (x33215)

### 2.1 – 2.2 GENERAL RESEARCH AND SUPPORT TOPICS (Census Bureau Projects 0351000, 1871000)

## Statistical Methodology

#### A. Disclosure Avoidance

*Description*: The purpose of this research is to develop disclosure avoidance methods to be used for all Census Bureau publicly available data products. Emphasis will be placed on techniques to implement disclosure avoidance at the stage of processing. Methods will be developed, tested, evaluated, and documented. We will also aid in the implementation of the methods.

*Highlights:* In order to understand the effect of swapping on data quality, some simple models were carefully explored carefully using ideas and formulas from mathematical statistics. This work involved a senior staff member and an intern. The intern did the majority of the analysis and wrote computer programs in Matlab whose results were displayed graphically. This computational analysis was contained in a long class report by the intern. Staff wrote a note on "Finding the distribution of a sample formed by combining samples from two Normal distributions."

Staff attended (on Nov. 20th) the opening workshop of a new research institute at The American University in Washington, D.C. The "Info-metrics Institute" tries to promote the use of some general information theory ideas that have proven useful in econometrics, to new fields, such as finance and risk management.

The Disclosure Review Board worked with staff at the Social Security Administration (SSA) to develop an agreement on the review process for data products developed at SSA that include Title 13 data.

Staff participated in the National Academy of Sciences' Census Transportation Planning Package project panel for producing transportation data products from the American Community Survey that comply with disclosure rules.

*Staff*: Laura Zayatz (x34955), Asoka Ramanayake, Jason Lucero, Paul Massell, Shira Appelbaum, Marlow Lemons, Vanessa Patterson, Bimal Sinha, Lisa Singh, Tapan Nayak

#### **B.** Disclosure Avoidance for Microdata

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

*Highlights:* Staff reviewed literature, refereed two papers related to microdata confidentiality, and reviewed a book proposal for J. Wiley.

Staff will continue to review literature, and plans to also evaluate faster computational algorithms and perform more empirical work.

*Staff:* William Winkler (x34729), William Yancey

# C. Seasonal Adjustment (See Economic Project 2370052)

#### **D.** Household Survey Design and Estimation

*Description*: The household 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 nonresponse, proper use of population estimates as weighting controls, and the effects of imputation on variances.

Highlights: See Section 1.9, projects E and F.

*Staff*: Lynn Weidman (x34902)

# E. 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:* No significant progress this quarter.

Staff: Mary Mulry (x31759)

#### F. Research and Development Contracts

*Description*: The Research and Development Contracts are indefinite delivery, indefinite quantity task order contracts for the purpose of obtaining contractor services in highly technical areas to support research and development activities across all Census Bureau programs. The contracts provide a pool of contractors to assist the Census Bureau in conducting research on all survey and census methods and processes to improve our products and services. The prime contractors include educational institutions, university supported firms and privately owned firms that concentrate in sample survey research, methodology, and applications to create a pool of specialists/experts to tackle some of the Census Bureau's most difficult problems through research. Many of the prime contractors are teamed with one or more organizations and/or have arrangement with outside experts/consultants to broaden their ability to meet all of the potential needs of the Census Bureau. These 5-year contracts allow Census Bureau divisions and offices to obtain outside advisory and assistance services to support their research and development efforts quickly and easily.

#### R&D 2007 Contracts

Twenty-five contracts were awarded during Fiscal Year 2002 in six technical areas: 1) assessment, planning, and analysis; 2) data analysis and dissemination; 3) statistical analysis, 4) methodological research, 5) sub-population research, and 6) survey engineering. The contracts ended September 30, 2009; however, the task orders awarded prior to the end date will be allowed to continue until completion.

*Highlights:* During the first quarter of FY2010, eight task orders were modified and one was completed. To date, there have been 96 task orders awarded under the R&D2007 contracts, with a monetary value over \$129 million (over \$106 million obligated). Seventy-five task orders have been completed and one task order terminated, leaving 20 active tasks.

#### R&D 2014 Contracts

Thirty-seven contracts were awarded during Fiscal Year 2009 to thirty-one firms in five technical areas: 1) assessment, planning, and analysis; 2) data analysis and dissemination; 3) statistical analysis and evaluation, 4) methodological research, and 5) survey engineering.

*Highlights:* During the first quarter of FY2010, one task order was awarded. To date, there has been 1 task order awarded under the R&D2014 contracts, with a monetary value of \$2.6 million, (\$2.2 million obligated).

*Staff*: Ann Dimler (x34996)

#### G. Small Area Estimation

*Description*: Methods will be investigated to provide estimates for geographic areas or subpopulations when sample sizes from these domains are inadequate.

*Highlights:* Staff continued research into modeling the design components of small area models for coverage estimation and continued modeling design effects for small area models in other surveys.

*Staff:* Don Malec (x31718), Aaron Gilary, Elizabeth Huang, Partha Lahiri, Jerry Maples

### Statistical Computing Methodology

## A. Record Linkage and Analytic Uses of Administrative Lists

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

*Highlights:* Staff reviewed a considerable amount of literature on parallel and threaded programming methods. Staff reviewed literature on record linkage in the health sciences.

Staff continued investigating generalizations of the Birthday and Collision Problems.

Staff plans to work on record linkage theory and computational methods, and to connect the methods with general methods of modeling/edit/imputation.

*Staff*: William Winkler (x34729), William Yancey, Ned Porter

#### **B.1 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:* Staff presented research at the work session on Statistical Data Editing (October 5-9, 2009). This research covered new ideas on creating edit rules for a given set of data, applying the edit rules efficiently to maximize quality of the resultant database, providing strategies for monitoring the edits, and providing partial measures for assessing the quality of the edit set (M. Garcia and W. Winkler, "Determining a set of edits and quality of a database").

#### Staff: María García (x31703)

#### **B.2 Editing and Imputation**

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

*Highlights:* Staff continue to implement prototype for forecasting future and unreported labor force items in the context of longitudinal information and/or conditional on demographic variables. The current prototype is implemented in the context of the Survey of Income and Program Participation (SIPP). It is tuned for forecasting employment status and health-insurance coverage at wave 2 conditional on wave 1 information and education. The same model can be taken advantage of to forecast

labor force items, conditional on administrative information available from administrative records.

Staff: Yves Thibaudeau (x31706), Eric Slud

#### C. Developed Software Support – General Variance Estimation Development and Support

*Description*: This project will develop new methods and interfaces for general variance estimation software including VPLX, WesVar, and SUDAAN. Our staff will provide training for variance estimation software applications, and will provide support for complex applications such as the Survey of Income and Program Participation and the Survey of Construction.

Highlights: No significant progress this quarter.

Staff: Aref Dajani (x31797), Ned Porter

#### **D.** Missing Data and Imputation: Multiple Imputation Feasibility Study

*Description*: Methods for imputing missing data are closely related to methods used for synthesizing sensitive items for disclosure limitation. One method currently applied to both issues is multiple imputation. Although the two issues may be addressed separately, techniques have been developed that allow data users to analyze data in which both missing data imputation and disclosure limitation synthesis have been accomplished via multiple imputation techniques (e.g., synthetic data). This project ascertains the effectiveness of applying multiple imputation to both missing data and disclosure limitation in the American Community Survey (ACS) group quarters data. Statistical models are used to generate several synthetic data sets for use within the multipleimputation framework.

*Highlights:* Staff altered ACS group quarters disclosure avoidance program to operate via a configuration script, incorporated DISCRETE editing program to ensure edit consistency of synthetic data, and implemented an improved flagging procedure for identifying at-risk records.

Staff: Rolando Rodríguez (x31816), Ben Klemens, Yves Thibaudeau

#### E. Modeling, Analysis, and Quality of Data

*Description*: Our staff investigates methods of the quality of microdata primarily via modeling methods and new software techniques that accurately describe one or two of the analytic properties of the microdata.

*Highlights:* Staff reviewed a Ph.D. dissertation and several journal articles on imputation. One staff person provided software, documentation, and advice related to generalized edit/imputation software to other staff within our division. Staff reviewed literature on Statistical Matching.

Staff plans to extend a general data analysis framework so that it includes applications for general modeling, edit/imputation and non-trivial record linkage.

*Staff*: William Winkler (x34729), Rob Creecy, William Yancey, María García

## Social & Behavioral Sciences Survey Methodology

#### A. Usability Research and Testing

#### A.1. Web Applications Accessibility

*Description*: This project focuses on the accessibility of Internet and Intranet applications by blind and low vision users in accordance with the *Section 508* regulations.

<u>Census 2010 Web Site</u> (Systems Support Division) *Description:* This Web site provides information via text, audio, and video to respondents about the 2010 Census.

*Highlights:* Staff tested many different aspects of the Census 2010 Web site for conformance with *Section 508*. Video and audio players were tested for keyboard access to controls. The main feature of the site, the marquee, was found to be inaccessible because there was no keyboard access to the video histories. The interactive 2010 Census form was made accessible by providing a plain text version. SSD has resolved these issues and work will continue into next quarter in preparation for the 2010 Census.

*Staff:* Larry Malakhoff (x33688), Lisa Wolfisch (SSD), Carollyn Hammersmith (SSD)

2010 Title 13 Awareness Training E-Learning Application (Systems Support Division) *Description:* This application permits Census Employees to satisfy the annual requirement for Title 13.

*Highlights:* Testing revealed what appeared to be text and graphical links were all button labels. Issues with button labels were reported because users expect items appearing as links to function as links. All text and graphical buttons were unlabeled, making pop-up windows inaccessible. In the course of accessibility testing, two main usability issues were detected. 1) There was no visual indication a link had been visited (clicked) by turning from blue to purple. 2) Response option selection was overly complex. Conventional radio button groups as response options were replaced by a layout where users were directed to select a letter for an item instead of selecting the item directly. These issues are being addressed by SSD and testing will resume next quarter.

*Staff:* Larry Malakhoff (x33688), Lisa Lawler (SSD)

Support for X-12 ARIMA Documentation and Software (Statistical Research Division) *Highlights:* No significant progress this quarter.

*Staff*: Larry Malakhoff (x33688), Brian Monsell

## <u>CPSI-PAL High-Level Expert Reviews</u> (Systems Support Division)

*Description*: This project is a follow-up of the review of the high-level concept document of the CPSI-PAL web site performed in 2009. This Web site will allow users to research roles and responsibilities for usage in process improvement.

*Highlights:* Staff performed two expert reviews in preparation for the planned release in January 2010. Staff found in the first review Census jargon was used on the welcome screen, potentially confusing users, acronyms were not clearly defined, and alignment of text varied between pages. The second review confirmed recommended changes were made. Staff found the usefulness of the search widget to be problematic. First, search results were not meaningful and the position of the search box made users believe pressing the "Go" button next to the search box would also work for the "Process," "Role," or "topic" searches. Staff is anticipating performing usability testing next quarter.

*Staff*: Larry Malakhoff (x33688), Temika Holland, Jennifer Romano, Jennifer Chen

<u>NotifyMe</u> (Economic Planning and Coordination Division)

*Description*: NotifyMe allows persons to select manufacturing reports and be notified when they become available.

*Highlights:* Staff met with the Economic Planning and Coordination Division to verify if changes to include instructions to use two extended list-boxes were programmed properly. It was found the instructions were programmed incorrectly. This work will continue next quarter.

*Staff*: Larry Malakhoff (x33688)

Data Tables (Systems Support Division)

*Description*: Staff reviews various data tables for accessibility and provides recommendations if the table is not coded properly.

Highlights: No significant progress this quarter.

Staff: Larry Malakhoff (x33688), Laura Yax (SSD)

<u>Censtats & USA Counties</u> (Administrative Customer Services Division)

*Description:* These Web sites permit users to get national, state, and county statistics.

*Highlights:* Staff provided feedback to a programmer about lack of labels for combo boxes used to select parameters to generate tables. Tables generated by these applications used numeric codes as a description of the remaining items in the row instead of an actual text description. Further, all tables had buttons in them to show different views of the data that were all labeled with the same text. This project is on hold until the ACSD programmer makes corrections.

Staff: Larry Malakhoff (x33688), Tina Egan (ACSD)

#### Economic Directorate Document Management System (EDMS) (ESMPD)

*Description:* This application informs users about document approval status and project updates.

*Highlights:* The EDMS Web site had virtual "drawers" that could be opened to show links. Staff determined the drawers could be opened by keyboard commands but found none of the links were accessible to the screen-reader software. Testing will resume when corrections are made.

Staff: Larry Malakhoff (x33688), Vinh Le (ESMPD)

#### A.2. Desktop Applications Accessibility

*Description*: This project focuses on accessibility of desktop applications by blind and low vision users in accordance with the *Section 508* regulations. Desktop applications are either downloaded or sent to the respondent on disk.

#### Census In the Schools (CIS) Documents (CLMSO)

*Description:* Teachers instructing children in all grades would use these lesson plans and exercises to learn about what information the Census Bureau collects and how it is used.

*Highlights:* Staff reviewed 39 documents (PDFs) for conformance with *Section 508*. This project will continue into next quarter.

Staff: Larry Malakhoff (x33688)

#### A.3. Census.gov Template Development

*Description*: The purpose of this study is to develop a set of templates with a consistent and usable look and feel for the Census.gov website. The template is intended to be used by both the demographic and economic domains of Census.gov. Some of the techniques to develop the template include card sorting, low-fidelity prototype testing, and usability testing.

*Highlights:* No significant progress this quarter.

Staff: Erica Olmsted-Hawala (x34893)

#### A.4. AFF Usability Study: Iterations 1 and 2– Conceptual Design and Low-Fidelity Prototype Testing

Description: The U. S. Census Bureau releases much of the nation's economic and demographic data on the American FactFinder (AFF) Web site. In conjunction with the massive data release anticipated at the conclusion of the 2010 Census, AFF is currently undergoing a major redesign, which is scheduled to launch sometime in 2011 or 2012. The Data Access and Dissemination Systems Office (DADSO) has asked the Census Bureau's usability lab to participate in the redesign effort. This effort encompasses the full spectrum of user-centered design activities, from iterative lowfidelity paper prototype testing to high-fidelity testing with a working prototype. The purpose of the testing is to identify usability issues. Recommendations made to resolve the issues are intended to improve the usability of the Web site for all users.

*Highlights:* Staff ran four additional expert users through the latest iteration prototype of the AFF Web site. Staff incorporated their results into the quick report and began work on compiling the results of two rounds of testing into a final report for the division's *Research Report Series*.

*Staff*: Erica Olmsted-Hawala (x34893), Jennifer Romano, Jennifer Chen

#### A.5. AFF Usability Study: Baseline Testing

*Description*: American FactFinder (AFF) is a free, public online tool that allows users to find, customize and download data on the population and economy of the United States. The AFF Web site is undergoing a thorough redesign under the sponsorship of the Data Access and Dissemination Systems Office (DADSO). In order to evaluate whether the re-designed effort is successful, the Usability Lab proposed and conducted a baseline usability study to measure user performance and satisfaction with the current site. Ultimately the results will be used to compare user performance and satisfaction with the same measures taken on the final release of the new AFF Web site, expected sometime in 2011 or 2012.

*Highlights:* Staff ran four additional expert users through the baseline study and began to add the results into the final baseline report.

*Staff*: Erica Olmsted-Hawala (x34893), Jennifer Romano, Jennifer Chen

# A.6. Spatial Ability Research with Iowa State University

*Description*: The purpose of this research is to continue the Census Bureau's investigations of the role of spatial ability in mediating the success of field personnel in performing computer-based tasks. Highlights: No significant progress this quarter.

Staff: Kathleen Ashenfelter (x34922)

#### A.7. Usability Evaluation of the Business & Industry Web Site

*Description*: Our division was asked to provide usability guidance and testing support to the Economic Current Surveys Web Site Redesign Team, chaired by Mark Wallace (Chief, Sector Services Statistics Division). This effort entailed attending regular team meetings and commenting on proposed design elements; providing input to occasional planning meetings with the SSSD Chief; and implementing a wide range of usability evaluation methods, from expert review to high-fidelity prototype testing.

*Highlights:* Staff met with the team and provided expert advice on screen shots of some of the emerging designs of the Business and Industry pages. Staff submitted the final report of the Business and Industry round 1 usability study to the division's *Research Report Series*.

Staff: Erica Olmsted-Hawala (x34893), Jennifer Chen

#### A.8. Baseline Usability Testing of the American Community Survey Web site

*Description*: Our division was asked to provide a baseline measure of the current American Community Survey (ACS) Web site. Within a year, the development team of the ACS Web site plans to make some major re-design changes to the interface of the Web site. In order to evaluate whether the re-designed effort is successful, the Usability Lab proposed a baseline usability study to measure user performance and satisfaction with the current site. Ultimately the results will be used to compare user performance and satisfaction with the same measures taken on the final release of the new American Community Survey Web site.

*Highlights:* Staff ran expert users through the baseline study, identified usability problems, wrote a quick report which detailed the usability problems and presented the results to the client. Results, as in the first study with novice users, show that the main tabs across the top of the Web site convey little about what is found under each section. The average task success rate across all participants was 59%. The average amount of time it took for participants to complete their tasks was five minutes and seven seconds. These are below the goals set for the site.

Staff: Erica Olmsted-Hawala (x34893), Jennifer Chen, Temika Holland

#### **B.** Questionnaire Pretesting

*Description*: This project involves coordinating the Census Bureaus generic clearance for questionnaire pretesting research. Pretesting activities in all areas of the

Census Bureau may use the clearance if they meet the eligibility criteria.

*Highlights:* During this quarter, two letters were sent to OMB for approval under the generic clearance for questionnaire pretesting research: cognitive interviews for the Survey of New Foreign Direct Investment in the United States (BE-13) and cognitive interviews for the Enterprise Statistics Program questions for the 2012 Report of Organizations.

Staff also completed and sent to OMB the annual report summarizing projects completed under the clearance during the last year.

*Staff*: Terry DeMaio (x34894)

#### C. Questionnaire Design Experimental Research Survey 2006 (QDERS)

*Description*: QDERS 2006 is an omnibus survey designed to facilitate independent research related to questionnaire design issues and other survey methodology issues. The QDERS 2006 was conducted from the Hagerstown Telephone Center. The focus of the 2006 QDERS is a questionnaire design experiment examining different ways to determine a person's place of residency on Census day.

*Highlights:* Staff continued working on a paper comparing this Random Digit Dialing study to a field pretest.

*Staff*: Jennifer Hunter Childs (x34927), Beth Nichols, Rolando Rodríguez, Aref Dajani

#### D. Language: Interdisciplinary Research on Language and Sociolinguistic Issues Relevant to Survey Methodology

*Description*: There is a need for both qualitative and quantitative interdisciplinary research on how to best develop and successfully use non-English language collection instruments and other survey materials. Interdisciplinary research is also needed to determine the quality of the data that respondents with little or no knowledge of English provide the Census Bureau using both English and non-English language data collection instruments.

*Highlights:* Our staff worked collaboratively with researchers in academia and survey research organizations on cross-cultural issues in survey interviews and translation methods. Specifically, we studied the following problems: 1) cross-cultural communication norms and survey interviews, 2) the use of interpreters in survey interviews, 3) language and cultural effects on conducting cognitive interviews in non-English languages, 4) methods to encourage survey participation from speakers of languages other than English, and 5) creation of best practices for the

management of non-English language cognitive testing research.

In collaboration with researchers at the National Cancer Institute, we started a research project comparing English and Chinese cognitive interviews. This research effort aims at identifying methodological issues, including protocol development, probing techniques, and optimal number of interviews for cognitive testing in multiple languages. During the first quarter of FY2010, we continued interviewing English and Chinese respondents, and fine tuned the linguistic coding scheme to code key linguistic features associated with the probing questions and responses.

Staff members continued to be actively involved with an international group of researchers, who are members of the Comparative Survey Design and Implementation (CSDI) group, and with the Interagency Roundtable on Languages (IRL) to work on the development of the Census Bureau language assessment tool. Staff conducted literature review on language proficiency testing and compiled a list of experts for organizing a panel of experts to help the Census Bureau to work on language proficiency tests. We were also actively writing up research results during this quarter to be published in various journals and conference proceedings.

In collaboration with university researchers, we continued to work on two book projects on discourse analysis and politeness communication.

*Staff*: Yuling Pan (x34950), Patricia Goerman, Jennifer Hunter Childs, Anna Chan, Virginia Wake Yelei, George Higbie, Matt Clifton

#### E. Training for Cognitive Interviewing

*Description*: Our staff will train members of other divisions in the Census Bureau to carry out cognitive interviewing and provide consultation and support for projects.

Highlights: No significant progress this quarter.

*Staff*: Jennifer Hunter Childs (x34927), Yuling Pan, Patricia Goerman, Terry DeMaio

#### F. Research on Cognitive Testing of Non-English Language Survey Instruments

*Description*: The staff is currently engaged in a study designed to test and identify best practices for conducting cognitive interviews with Spanish-speaking respondents. We have tested both widely accepted and new techniques and probes (e.g., "What does the term foster child mean to you in this question?") with Spanish-speaking respondents of high and low educational levels. The research was based on a segment of the CAPI version of the American Community Survey. Future applications of this research should extend to cognitive interview techniques for use with respondents who speak additional non-English languages.

*Highlights:* Staff continued ad hoc review of Spanish and Chinese 2010 Census promotion materials.

*Staff*: Patricia Goerman (x31819)

### G. Interviewer-Respondent Interactions

Description: Survey nonresponse rates have been increasing, leading to concerns about the accuracy of (demographic) sample survey estimates. For example, from 1990 to 2004, initial contact nonresponse rates have approximately doubled for selected household sample surveys including the Current Population Survey (CPS) (from 5.7% to 10.1%). While mailout/mailback is a relatively inexpensive data collection methodology, decreases in mailback rates to censuses and sample surveys mean increased use of methodologies that bring respondents into direct contact with Census Bureau interviewers (e.g., field representatives) using CATI (computer assisted telephone interviewing) or CAPI (computer assisted personal interviewing). CAPI can include face-to-face or telephone contact. Unsuccessful interviewer-respondent interactions can lead to increased costs due to the need for additional follow-up, and can also decrease data quality.

*Highlights:* Database was checked and finalized with data from a sample of over 500 Current Population Survey interviewers regarding their beliefs about what is effective in gaining cooperation from sample households. Initial plans for analyses began centered around finding associations among various interviewer beliefs.

*Staff*: Tommy Wright (x31702), Tom Petkunas

### H. Q-Bank: A Database of Pretested Questions

Description: Q-Bank was developed through an interagency committee, led by the National Center for Health Statistics (NCHS), of which the Census Bureau is a member. The objective of Q-Bank is to have an online interagency database of pretested survey questions and research results. The database is maintained at NCHS and is guided and used by other participating Federal statistical agencies, including the Census Bureau. Q-Bank serves many purposes. When survey questions and questionnaires are being developed, O-Bank can be used by survey methodologists and subject matter experts to search through previously tested questions. Q-Bank provides a forum to catalog pretesting reports in a manner that is easy to search by content or subject matter. Q-Bank also will allow us to produce meta-data about our pretesting findings. And, finally, Q-Bank will be an additional resource for analysts to interpret survey data. Q-Bank has just reached the production phase and is currently being populated with cognitive test reports, which is a necessary step before it becomes available to a broader audience.

*Highlights:* Staff participated in the Questionnaire Evaluation Methods workshop at NCHS, with leaders in the field of questionnaire evaluation. Staff continued development of the Q-Bank database and began investigating eTraining to use the Q-Bank system.

*Staff*: Jennifer Hunter Childs (x34927), Jennifer Beck, Yuling Pan, Patricia Goerman

### I. Health Insurance Measurement

*Description*: The U.S. health care system is a patchwork of public and private programs and plans, thus there are no definitive centralized records on the number of individuals without insurance. Researchers must rely on surveys for this estimate, and the Current Population Survey (CPS) is the most widely-cited source for this statistic. It is not without its critics, however, and recent official reports have included caveats regarding the data quality. The purpose of this research is to identify particular features of the CPS questionnaire that are associated with measurement error, and to explore alternative designs to reduce that error.

*Highlights:* Preparations for the 2010 field test continued. Tasks included testing each module of the automated instrument, developing sampling specifications for both the Random Digit Dialing sample and the Medicare enrollee sample, developing training materials, writing data processing and out specifications, developing sampling and operations procedures. Staff also developed plans for several evaluations methods, including taperecording interviews for later behavior coding, and conducting a focus group with interviewers to develop methods for obtaining question-specific feedback on the questionnaire. In conjunction with the Policy Office, staff also finalized the wording for an experiment on three different versions of a question requesting permission to link survey data with administrative records.

*Staff*: Joanne Pascale (x34920)

#### J. Emerging Social Trends on Household Structure and Living Situations, Race/Ethnicity, and Linkages to Enumeration Methods and Coverage

Description: In 2006, the National Academies of Science (NAS) Panel on Residence Rules recommended that the Census Bureau establish a trends office with an ongoing research program on social trends, enumeration methods, and coverage. This program would include monitoring emerging social trends and their impact on the accuracy of basic residence information and census coverage. It would also include developing, conducting, and synthesizing new research to suggest changes in enumeration methods and improve census coverage. Specifically recommended ongoing research topics include: "research on changing factors influencing people's attachments to locations where they are counted," "living situations," "large households," "sources of omissions in the census, as well as

duplications," and "questionnaire strategies" (NRC 2006: 175-178).

Highlights: No significant progress this quarter.

Staff: Laurie Schwede (x32611)

### K. Using Vignettes to Explore Survey Concepts

Description: Vignettes are a common tool for survey pretesting. Vignettes depict hypothetical situations and allow us to evaluate concepts without actually having to recruit people in those situations. Vignettes are also useful when evaluating survey topics that may be highly sensitive. This research will identify and explore how teens classify their contacts with online strangers and the degree to which they are aware of the danger in such interactions. The study will be a mixed-design qualitative and quantitative study. Participants will classify vignettes depicting online contacts with strangers and online contacts with non-strangers as either being appropriate and harmless or inappropriate and dangerous. Participants will also answer open-ended questions about why they feel these contacts are or are not dangerous. The results of the vignette classification task and the open-ended questions will help to identify how teens conceptualize their online relationships and reveal potential online vulnerabilities.

*Highlights:* Staff completed a preliminary content analysis on previously collected data.

Staff: Jennifer Beck (x31736), Terry DeMaio

#### L. Retrieval Effects on Judgments about Knowledge

*Description*: Surveys are a common way to collect information on a variety of topics. It is easy to assume that if people understand the intended meaning of and know the answer to a survey question, they should have relatively little problem providing an accurate answer. However, research on human memory and knowledge assessment casts significant doubt on this assumption. Context, in the form of both situational variables and individual differences, can have a significant effect on how accurately people answer questions.

In an attempt to investigate the effects of these variables on evaluations of knowledge, we have developed a set of experiments that will investigate the effects of retrieval context on how people evaluate their knowledge of general, factual information. This research will be jointly conducted with researchers at SUNY Stony Brook.

*Highlights:* Staff collected a large portion of the data and completed a preliminary analysis. The results suggest that retrieval difficulty affects judgments about knowledge only when people find it easy to remember past memory failures. In this situation, people tend to have a more pessimistic view of their own knowledge, despite being more accurate about what they do or do not know.

Staff: Jennifer Beck (x31736)

## **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:* Tina Arbogast, Matt Clifton, Matt Gore (HRD), George Higbie, Temika Holland, Gloria Prout, Lorraine Randall, Kelly Taylor

## **3. PUBLICATIONS**

#### **3.1 JOURNAL ARTICLES, PUBLICATIONS**

- Holan, S. and McElroy, T. (In press). "Tail Exponent Estimation via Broadband Log Density-Quantile Regression." Journal of Statistical Planning and Inference.
- Klein, M., Sinha, B., and Subramaniam, R. (In press). "Statistical Inferences from Formaldehyde DNA-protein Cross-link Data: Improving Methods for Characterization of Uncertainty." *Journal of Biopharmaceutical Statistics*.
- McElroy, T. and Findley, D. (In press). "Discerning Between Models Through Multi-Step Ahead Forecasting Errors." *Journal of Statistical Planning and Inference*.
- Pan, Y. (In press). "Epilogue for Special Issue on Institutional Politeness in (South) East Asia." Journal of Asian Pacific Communication, 20.2.
- Roberts, C. G., Holan, S. H., and Monsell, B. C. (In press). "Comparison of X-12-ARIMA Trading Day and Holiday Regressors with Country Specific Regressors." *Journal of Official Statistics*.

#### **3.2 BOOKS/BOOK CHAPTERS**

#### **3.3 PROCEEDINGS PAPERS**

2009 American Association for Public Opinion Research Meeting, Hollywood, FL, May 14-18, 2009. 2009 Proceedings of the American Statistical Association.

- Chan, A. and Pan, Y. (2009). "The Use of Cognitive Interviewing to Explore the Effectiveness of Advance Supplemental Materials among Five Language Groups," 5822-5836.
- Goerman, P. and Clifton, M. (2009). "Vignettes in Cross-Cultural Cognitive Testing: Adaptation for Spanish-Speaking Respondents of Lower Educational Levels," 5794-5806.
- Pan, Y. and Landreth, A. (2009). "Conveying Translated Informed Consent Concepts: Effects of Language and Culture on Interpretation of Legally Required Messages," 5902-5916.
- Park, H., Sha, M., and Pan, Y. (2009). "Cognitive Testing as a Method of Pre-Testing Questionnaires in High and Low Context Cultures," 5676-5690.
- Schwede, L. and Sorokin, A. (2009). "To Link or Not to Link?: Exploring Approaches to Maintaining American Community Survey Response Rates During Census 2010," 5844-5858.
- Sha, M. and Pan, Y. (2009). "The Use of Vignettes in Evaluating Multilingual Questionnaires," 6122-6133.

2009 Joint Statistical Meetings (American Statistical Association), Washington, D.C., August 1-6, 2009. 2009 Proceedings of the American Statistical Association.

- Childs, J., Norris, D., and Jurgenson, N. (2009). "Pretesting 2010 Census Questionnaires for People with Atypical Living Situations," 1880-1891.
- Findley, D. (2009). "Stock Series Holiday Regressors Generated by Flow Series Holiday Regressors," 477-486.
- Huang, E. and Bell, W. (2009). "A Simulation Study of the Distribution of Fay's Successive Difference Replication Variance Estimator," 5294-5309.
- Ikeda, M. (2009). "Developing Guidelines Based on CVs for when One-Year Estimates Can Be Used Instead of Three-Year Estimates in the American Community Survey," 1670-1677.
- Lucero, J., Singh, L., and Zayatz, L. (2009). "The Current State of the Microdata Analysis System at the Census Bureau," 3762-3772.
- Maples, J., Bell, W., and Huang, E. (2009). "Small-Area Variance Modeling with Application to County Poverty Estimates from the American Community Survey," 5056-5067.
- Massell, P. (2009). "An Overview of Uncertainty Creation to Protect Statistical Data," 1550-1558.
- McDonald-Johnson, K., Findley, D., and Cepietz, E. (2009). "Investigating Quarterly Trading Day Effects," 3305-3317.
- McElroy, T. and Holan, S. (2009). "The Detection of Cycles in Raw and Seasonally Adjusted Data," 702-707.

- Monsell, B. (2009). "Update on the Development of X-13ARIMA-SEATS," 1907-1920.
- Mule, V., Malec, D., Imel, L., Nguyen, N., and Moldoff, M. (2009). "Missing Data Methods for the CCM Component Error Estimation," 3265-3278.
- Mulry, M. and Adams, T. (2009). "Overview of Evaluations of the 2010 Census Coverage Measurement Program," 3117-3128.
- Mulry, M. and Oliver, B. (2009). "A Simulation Study of Treatments of Influential Values in the Monthly Retail Trade Survey," 2979-2993.
- Mulry, M. and Olson, T. (2009). "Analyses for Partnerships Based on the Census Barriers, Attitudes, and Motivator Survey," 5727-5741.
- Schwede, L., Sorokin, A., and Wake, V. (2009). "You Really Have to Puzzle This Out': Checking Residence and Coverage Duplications on a Census 2010 Questionnaire," 4504-4518.
- Slud, E. and Thibaudeau, Y. (2009). "Simultaneous Calibration and Nonresponse Adjustment," 2263-2272.
- Titova, N. and Monsell, B. (2009). "Detecting Stock Calendar Effects in U.S. Census Bureau Inventory Series," 2108-2122.
- Weidman, L. and Ashenfelter, K. (2009). "Results of a Survey on Choice of Sampling Error Display in American Community Survey Data Products," 1137-1150.
- Winkler, W. (2009). "General Discrete-data Modeling Methods for Producing Synthetic Data with Reduced Reidentification Risk that Preserve Analytic Properties," 2568-2576.

2009 Proceedings of the Federal Conference on Statistical Methodology, Washington, D.C., November 2-4, 2009.

- Bates, N. and Pan, Y. (2009). "Motivating Non-English-Speaking Populations for Census and Survey Participation."
- McElroy, T. and Holan, S. (2009). "Using Spectral Peaks to Detect Seasonality."
- Pascale, J. (2009). "Findings from a Pretest of a New Approach to Measuring Heath Insurance in the Current Population Survey."

Proceedings of the Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Bilbao, Spain, December 2-4, 2009.

• Zayatz, L., Lucero, J., Massell, P., and Ramanayake, A. (2009). "Disclosure Avoidance for Census 2010 and American Community Survey Five-year Tabular Data Products."

#### 3.4 STATISTICAL RESEARCH DIVISION RESEARCH REPORTS

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

**RR** (Statistics #2009-07): Christopher G. Roberts, Scott H. Holan, and Brian Monsell, "Comparison of X-12-ARIMA Trading Day and Holiday Regressors With Country Specific Regressors," October 8, 2009.

**RR** (Statistics #2009-08): William E. Winkler, "Sample Allocation and Stratification," November 16, 2009.

**RR** (Statistics #2009-09): Jason Lucero, Lisa Singh, and Laura Zayatz, "Recent Work on the Microdata Analysis System at the Census Bureau," November 16, 2009.

**RR** (Statistics #2009-10): Laura Zayatz, Jason Lucero, Paul Massell, and Asoka Ramanayake, "Disclosure Avoidance for Census 2010 and American Community Survey Five-year Tabular Data Products," November 23, 2009.

**RR** (Survey Methodology #2009-07): Joanne Pascale, "Findings from a Pretest of a New Approach to Measuring Health Insurance in the Current Population Survey," November 16, 2009.

**RR** (Survey Methodology #2009-08): Leticia Fernández, Eleanor Gerber, Matt Clifton, George Higbie, and Mikelyn Meyers, "Cognitive Pretesting of 2010 Alternative Questionnaire Experiment (AQE) Race and Hispanic Origin Treatment Panels," November 24, 2009.

**RR** (Survey Methodology #2009-09): Jeffrey Moore, Jason Fields, Gary Benedetto, Martha Stinson, Anna Chan, and Jerry Maples, "The 2008 Survey of Income and Program Participation Event History Calendar Field Test: Study Design and Initial Results," December 30, 2009.

#### 3.5 STATISTICAL RESEARCH DIVISION STUDIES

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

**SS** (Survey Methodology #2009-16): Yuling Pan and Ashley Landreth, "Conveying Translated Informed Consent Concepts: Effects of Language and Culture on Interpretation of Legally Required Messages," October 15, 2009.

**SS** (Survey Methodology #2009-17): Jennifer Hunter Childs, Leticia Fernández, Matt Clifton, and Mikelyn Meyers, "CFU CPEX Experimental Question Cognitive Testing: Undercount, Overcount and Duplicate Experimental Question Sequences," November 16, 2009.

**3.6 OTHER REPORTS** 

## 4. TALKS AND PRESENTATIONS

UNECE Work Session on Statistical Data Editing, Neuchâtel, Switzerland, October 5-7, 2009.

• Maria Garcia and William Winkler, "Determining a Set of Edits and Quality of a Database."

The Washington Statistical Society, Washington, D.C., October 8, 2009.

• Yuling Pan, "The Sociolinguistics of Survey Translation." (Invited lecture.)

Eurostat Training Session, Luxembourg, Luxembourg, October 20-22, 2009.

• Tucker McElroy, "X-13 ARIMA-SEATS."

Statistics Canada Symposium 2009 on Longitudinal Surveys: From Design to Analysis, Gatineau, Canada, Oct. 27-29, 2009.

• Eric Slud and Leroy Bailey, "Evaluation and Selection of Models for Attrition Nonresponse Adjustment." (Invited talk.)

2009 Conference for the Federal Committee on Statistical Methodology, Washington, D.C., November 2-4, 2009.

- Nancy Bates and Yuling Pan, "Motivating Non-English-Speaking Populations for Census and Survey Participation."
- Tucker McElroy, "Using Spectral Peaks to Detect Seasonality."
- Joanne Pascale, "Findings from a Pretest of a New Approach to Measuring Health Insurance in the Current Population Survey."

Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality, Bilbao, Spain, December 2-4, 2009.

• Laura Zayatz, "Disclosure Avoidance for Census 2010 and American Community Survey Five-year Tabular Data Products."

## 5. STATISTICAL RESEARCH DIVISION SEMINAR SERIES

Seminar Series Team: Aref Dajani, Richard Griffin (DSSD), Paul Massell, Laurie Schwede, Katherine Thompson (ADEP)

Partha Lahiri, (ASA/NSF/Census Research Fellow) University of Maryland, College Park, "Hierarchical Modeling and Related Inferential Issues In Small Area Estimation," October 21, 2009.

Jaejik Kim, University of Georgia, "Clustering Techniques for Histogram-valued Data," October 27, 2009.

Erica Olmsted-Hawala, SRD, U.S. Census Bureau, "Think-Aloud Protocols: A Comparison of Three Think-Aloud Protocols for Use in Testing Data Dissemination Web Sites for Usability," November 9, 2009.

Jennifer Romano, SRD, U.S. Census Bureau, "Benefits of Iterative Usability Testing on a Web Site," November 9, 2009.

Larry R. May and Rahul Tikekar, Office of Research, Internal Revenue Service, "Graph Query: A Pattern Matching Tool to Detect Structures on Linked Data," November 12, 2009.

Anton Korinek, University of Maryland, and Johan Mistiaen, World Bank, "Survey Nonresponse and the Distribution of Income," December 7, 2009.

Victoria Leaver, Australian Bureau of Statistics, "Statistical Disclosure Control: Methods and Directions in the ABS," December 14, 2009.

## **6. PERSONNEL ITEMS**

## 6.1 HONORS/AWARDS/SPECIAL RECOGNITION

### 6.2 SIGNIFICANT SERVICE TO PROFESSION

#### Kathleen Ashenfelter

- Associate Editor, Frontiers in Quantitative Psychology and Measurement.
- Representative, Capital Area Social Psychological Association (CASPA) to the American Association for the Advancement of Science's (AAAS) Science and Human Rights Coalition.
- Reviewed papers for the Computer Human Interaction (CHI) conference.

### Terry DeMaio

• Refereed papers for Journal of Official Statistics and Public Opinion Quarterly.

#### Jason Lucero

• Member, Confidentiality and Data Access Committee (CDAC).

#### Paul Massell

- Member, Confidentiality and Data Access Committee (CDAC).
- Member, Bureau of Transportation Statistics Disclosure Review Board.

#### Tucker McElroy

• Refereed papers for *Econometric Theory, Statistica Sinica, Journal of Multivariate Analysis,* and *Journal of Nonparametric Statistics.* 

#### Brian Monsell

• Webmaster, Business and Economic Statistics Section, American Statistical Association.

#### Jeff Moore

• Refereed a paper for the *Journal of the Royal Statistical Society*.

#### Mary H. Mulry

- Past-Chair, Survey Research Methods Section, American Statistical Association.
- Associate Editor, Journal of Official Statistics.

#### Tapan Nayak

- Member, Committee on Privacy and Confidentiality, American Statistical Association.
- Member, Editorial Board, *Communications in Statistics*.

### Yuling Pan

- Member, Editorial Board, Journal of Chinese Language and Discourse.
- Member, Advisory Board, Journal of Politeness Research.
- Member, Editorial Board, Advances in Pragmatics and Discourse Analysis Book Series, Cambridge Scholar Publishing.
- Member, Multilingual Interest Group, American Association for Public Opinion Research.
- Co-organizer, Session on Cross-Cultural Studies, 2010 Joint Statistical Meetings.

Asoka Ramanayake

- Member, Confidentiality and Data Access Committee (CDAC).
- Member, NAS CTPP Project Panel for Producing Transportation Data Products from the American Community Survey that Comply with Disclosure Rules.
- Refereed papers for the Sri Lankan Journal of Applied Statistics.

#### Laurie Schwede

• Reviewed abstracts for the 2010 AAPOR meeting.

#### Bill Winkler

- Refereed papers for Survey Methodology and the Journal of Official Statistics.
- Reviewed a book proposal for John Wiley Publishers.
- Associate Editor, *Journal of Privacy Technology*.
- Associate Editor, Journal of Privacy and Confidentiality.
- Associate Editor, Transactions on Data Privacy.
- Member, Committee on Voter Registration Databases, National Academies of Science.
- Member, Program Committee, Statistical Data Protection 2010 in Corfu, Greece.
- Member, Program Committee, QDB 2010 at the 2010 Very Large Database Conference in Singapore.

#### Tommy Wright

- Associate Editor, *The American Statistician*.
- Member, Fellows Committee, American Statistical Association.
- Member, Morris Hansen Lecture Committee, Washington Statistical Society.
- Member, Advisory Board, Mathematics and Statistics Department, Georgetown University.

#### Laura Zayatz

- Member, Confidentiality and Data Access Committee (CDAC).
- Member, NAS CTPP Project Panel for Producing Transportation Data Products from the American Community Survey that Comply with Disclosure Rules.
- Advisor, Disclosure Review Board, Social Security Administration.
- Member, Advisory Board, Journal of Privacy Technology.
- Member, Committee on Privacy and Confidentiality, American Statistical Association.
- Member, UK Census Design and Methodology Advisory Committee.
- Refereed Report on Proposed Disclosure Avoidance Techniques to be used by the Office of National Statistics, United Kingdom, for Census 2011.
- Organizer, Privacy in Statistical Databases 2010.
- Member, Advisory Board, Journal of Empirical Research on Human Research Ethics.

#### **6.3 PERSONNEL NOTES**

Pam Ferrari retired from the Census Bureau after 31 years of federal service.

Thomas Mathew (member of the statistics faculty and Presidential Research Professor at the University of Maryland-Baltimore County) accepted a Schedule A Appointment in our Computing Applications Group.

Chandra Erdman joined our Missing Data Methods Research Group.

Victor Quach (graduate student in human factors psychology at The Catholic University of America) joined our Human Factors and Usability Research Group as an intern.

Jon Krosnick (Frederic O. Glover Professor in Humanities and Social Sciences as well as Professor of Communication and of Political Science at Stanford University) accepted a Schedule A Appointment in our Language and Measurement Research Group.

Matt Gore joined the Human Resources Division and is on a temporary assignment to our division through the Mixed-Tour Program.