# Annual Report *of the* Statistical Research Division

*Fiscal Year 2005*

Decennial Directorate Customers

| Estimation | Ethnography | Time Series |
| Usability | Edit and Imputation | Pretesting |
| Questionnaire Design | | Small Area Estimation |
| Record Linkage | SRD Expertise for Collaboration and Research | Statistical Computing |
| Sampling | Disclosure Avoidance | Modeling |
| Measurement | Seasonal Adjustment | Cognitive Testing |

Demographic Directorate Customers

Economic Directorate Customers

Field Directorate Customers

Other Internal and External Customers

USCENSUSBUREAU

*Helping You Make Informed Decisions*

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

# $S$ince August 1, 1933...

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

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

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

- Early Years (1933–1960s): *sampling (measurement of unemployment and 1940 census); probability sampling theory; nonsampling error research; computing; data capture.*

- 1960s–1980s: *self-enumeration; social and behavioral sciences (questionnaire design, measurement error, interviewer selection and training, nonresponse, etc.); undercount measurement, especially at small levels of geography; time series and seasonal adjustment.*

- 1980s–Early 1990s: *undercount measurement and adjustment; ethnography; record linkage; confidentiality and disclosure avoidance.*

- Mid 1990s–Present: *small area estimation; missing data and imputation; usability (human-computer interaction); linguistics, languages, and translations.*

---

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

U.S. Census Bureau
Statistical Research Division
Federal Building 4
Washington, DC 20233
301-763-1702

*We help the Census Bureau improve its processes and products.  For fiscal year 2005, this report is an accounting of our work and our results.*

*Statistical Research Division*

# *Highlights of What We Did...*

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

Highlights of a selected sampling of the many activities and results in which Statistical Research Division staff members made contributions during FY 2005 follow, and more details are provided within subsequent pages of this report:

- built imputation models that can be used to generate synthetic microdata and attempted to link synthetic data to the original data to ensure disclosure avoidance protection was achieved.
- developed a new and improved version of the data swapping procedure and software that can be used for the group quarters data.
- improved X-13A-S (combination of X-12-ARIMA and SEATS) through revised methods to examine the effect of outliers, new model-based diagnostics, updated signal extraction routines, improved composite adjustment, and expanded diagnostic output.
- developed and applied methodology to forecast age distribution curves for legal immigrants.
- completed work evaluating the synthetic assumption of census coverage estimation using random effects models and reported results.
- observed small average improvements in state posterior variances for most age groups in SAIPE from the examination and testing of several bivariate models for state poverty ratio using income years 2000-2002 from the Current Population Survey Annual Social and Economic Supplement and the Supplementary Surveys of the American Community Survey.
- compared 4 alternatives for controlling housing units and group quarters person estimates of demographic characteristics totals to intercensal population estimates as part of American Community Survey weighting.
- developed preliminary approaches for estimating low-valued exports and imports, and studied the underlying issues that make direct evaluation impossible.
- completed the simulation of missing count data in the decennial census.
- adapted and tested an editing strategy proposed at Statistics Sweden to develop alternative score functions which automatically identify failing records that have a significant impact on the final tabulations.
- proposed modifications to 2004 methodology to include matching between group quarters and housing unit persons and modeling of the resulting person links.
- conducted or facilitated thirty-seven pretesting activities across the decennial, demographic, and economic areas under the OMB generic clearance.
- reported research results that 91% of SIPP 2004 wave 1 respondents agreed to allow their answers to be used "as a starting point" during the follow-up interview under the new Respondent Identification Policy (RIP).
- conducted cognitive interviewing (which simulated all modes of interviewing used in the American Community Survey) that improved the "at work" question and the "looking for work" question.
- developed new wording for the Hispanic origin question, the race question, and the new modified ancestry question for the decennial census.
- completed usability tests and evaluations for 12 Census Bureau projects; completed accessibility evaluations of seven Web sites, the Linux desktop, and two data browsers.
- provided critical and significant evaluation of several statistical issues for the U.S. Postal Rate Commission.
- completed research which proposed new imputation methods for the Survey of Income and Program Participation (SIPP) wealth imputation project sponsored by the Social Security Administration.
- completed a nonresponse bias analysis (study) for the National Science Foundation's 2003 National Survey of Recent College Graduates.

# *How Did We Do...*

For a seventh year, we received feedback from our sponsors. Near the end of fiscal year 2005, our efforts on sixty-eight of our program (Decennial, Demographic, Economic, External) sponsored projects/subprojects with substantial activity and progress and sponsor feedback (Appendix A) were measured by use of a Project Performance Measurement Questionnaire (Appendix B). Responses to all sixty-eight questionnaires were obtained with the following results (The graph associated with each measure shows the performance measure over the last seven fiscal years):

*Measure 1.    Overall, Work Met Expectations*

Percent of FY2005 Program Sponsored Projects/Subprojects where sponsors reported that overall work met their expectations (agree or strongly agree) (68 out of 68) . . . . . . . . . . . . . . . . . . . . . 100%
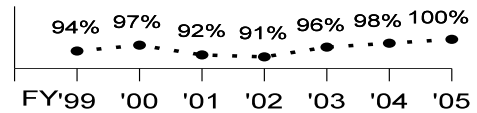
94% 97% 92% 91% 96% 98% 100%
FY'99  '00  '01  '02  '03  '04  '05

*Measure 2.    Established Major Deadlines Met*

Percent of FY2005 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met (55 out of 56 responses) . . . . . . . . . . . . . . . . . . . . 98%
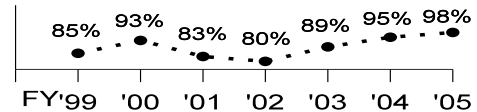
85% 93% 83% 80% 89% 95% 98%
FY'99  '00  '01  '02  '03  '04  '05

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

Percent of FY2005 Program Sponsored Projects/Subprojects reporting at least one improved method, technique developed, solution, or new insight (56 out of 67 responses) . . . . . . . . 84%

79% 91% 90% 88% 90% 82% 84%
FY'99  '00  '01  '02  '03  '04  '05
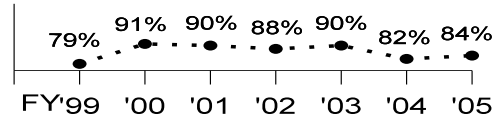
*Measure 3b.    Plans for Implementation*

Of these FY2005 Program Sponsored Projects/Subprojects reporting at least one improved method, technique developed, solution, or new insight, the percent with plans for implementation (45 out of 56 responses) . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80%
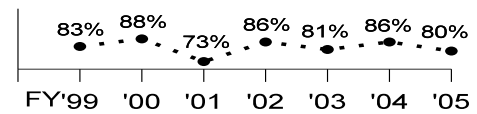
83% 88% 73% 86% 81% 86% 80%
FY'99  '00  '01  '02  '03  '04  '05

*Measure 4.    Predict Cost Efficiencies*

Number of FY2005 Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency" . . . . . . . . . . 18

10 17 11 18 9 23 18
FY '99  '00  '01  '02  '03  '04  '05
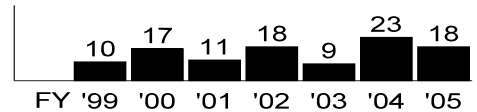
From Section 3 of this ANNUAL REPORT, we also have:

*Measure 5.    Journal Articles, Publications*

Number of peer reviewed journal publications documenting research that appeared (9) or were accepted (11) in FY2005 . 20
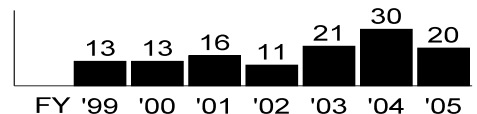
13 13 16 11 21 30 20
FY '99  '00  '01  '02  '03  '04  '05

*Measure 6.    Proceedings, Publications*

Number of proceedings publications documenting research that appeared in FY2005 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

28 23 19 34 22 30 30
FY '99  '00  '01  '02  '03  '04  '05
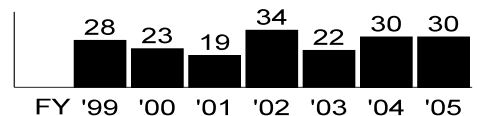
Each completed questionnaire and associated details are shared with appropriate staff to help improve our future efforts.

# TABLE OF CONTENTS

# 1.  COLLABORATION

## 1.1 - 1.2   DECENNIAL TOPICS
## (Decennial Projects 5210501 and 5210502)

### A.  Census Questionnaire Design Features

Two of these projects support the evaluation of decennial-related nonresponse and coverage survey research conducted by the Decennial Statistical Studies Division (DSSD) and the Planning, Research, and Evaluation Division (PRED).  The third project, a collaboration with the Decennial Management Division (DMD), aims to decrease mail-back response time for the Decennial Census form by embedding "deadline" messages in the mailing package.  The first two projects use behavior coding to analyze interviewer/respondent interactions to evaluate decennial follow-up survey questions in the 2004 Non-Response Follow-up (NRFU) Test and the Coverage Research Follow-up (CRFU) Survey.  Of particular interest in the NRFU is the effect of questions administration using the new hand-held computers for interviewers and respondents at the Queens, New York test site, as well as the administration/function of the race/ethnicity questions, due to a new approach that now omits a previously-used response category labeled "some other race" (Since FY 2005, "some other race" has returned as a response category). Furthermore, the NRFU evaluation entails analyzing both the English and Spanish language instruments.  Behavior coding of the CRFU, on the other hand, intends to capture problems with coverage questions in interviews conducted both in person and on the telephone at the New York and Georgia test sites. In particular, group quarters questions seem to be difficult to administer due to length, complexity, and sensitivity.  The last project, the Deadline Messaging Experiment, is an effort to determine the best types and locations for messages suggesting the mailed census form must be returned by a certain date. Cognitive interviews are the suggested methodology for determining respondents' comprehension/interpretation and reaction to messages placed in the mail materials.

NRFU Behavior Coding: During FY 2005, staff continued to carry out a time- and labor- intensive behavior coding analysis project. We continued to analyze both the quantitative and qualitative data, while writing up the findings. This research resulted in a substantial final report, provided to the PRED staff, portions of which were included in the PRED's final report of *2004 Census Test Evaluation 9: Effect of New Race and Hispanic Origin Questions*. The behavior coding study activities and findings generated three conference presentations on research related to: 1) interviewer/respondent interaction when survey data is collected on a mobile computing device; 2) interviewer/respondent interaction when person-based interviewing techniques are used to collect demographic data from multiple household members; and 3) augmenting the behavior coding method by analyzing final responses to survey questions to detect critical flaws affecting data quality.

Staff intend to carry out an evaluation of the 2006 NRFU instrument, again using the behavior coding method.

Behavior coding the CRFU: During FY 2005, the staff, in collaboration with the DSSD staff, undertook a behavior coding project from beginning to end (e.g., defining scope and research objectives, drafting training materials and conducting coder training, coder monitoring, database development, data analysis, and writing results and questionnaire recommendations). This research resulted in a substantial final report, which is currently under review. Major findings from the evaluation of the CRFU include: in general, interviewers asked questions as-worded or with slight changes that did not affect question meaning less often than is desirable (an average of 51% of the time, across 29 questions of interest); interviewers often dropped portions of the questions they may have deemed conversationally superfluous, and sometimes this included concepts from compound questions and reference periods that fluctuate across survey questions; and respondents were often able to provide acceptable answers during the first-level of exchange (80% of the time, across 25 questions).  A main effect on survey mode for appropriate interviewer behavior was found for 19 of 29 questions, though this result is likely confounded by differences that characterize the two pools of interviewers used for this research (e.g., telephone interviewers could have better performance rates than the novice field interviewers used for the CRFU due to increases in tenure, training, and direct supervision).  Many recommendations, namely those that simplified question wording and flow in order to increase the likelihood that interviewers will read the questions as-worded, were adopted.

Deadline Messaging: During FY 2005, we worked collaboratively with the Decennial Management Division staff to identify the project scope and schedule. After identifying a research window deemed too narrow for cognitive pretesting, we recommended that the deadline messages could be generated by the staffs of both divisions, then vetted through an expert review.  Language was created and forwarded to senior survey methodologists, and staff continue to be consulted further during the final stages of wording development and placement.

*Staff:* Ashley Landreth (x38457), Jenny Hunter, Theresa DeMaio, Aref Dajani, Pamela Ferrari

## B. Short Form Questionnaire Content Other Than Race and Ethnicity

This project involves participation in the 2010 Census Content Planning Group and content-related subgroups other than those focusing on race and ethnicity. It also involves consultation and testing on questionnaire content for the 2010 Census and tests leading up to it.

During FY 2005, staff participated in several planning groups and subgroups, including the 2010 Census Content Planning Group, the 2006 Wording and Content Subgroup, the NCT Implementation Group. Staff also conducted several cognitive interviewing projects. First, cognitive research was conducted on the tenure, age, and relationship questions for the 2005 National Census Test. In the tenure question, two changes were tested. First, eliminating "cash" from the "rented for cash rent" and "occupied without payment of cash rent" categories and second, adding the sentence "include home equity loans" at the end of the "owned by you or someone in your household with a mortgage or loan" category. In the age question, two changes were also tested. First, adding an instruction at the end of the question to "please report babies as age 0 when the child is less than 1 year old." Second, reversing the order of the age and month/day/year of birth questions. In the relationship question, three changes were tested. First, changing "natural born son or daughter" to "biological son or daughter." Second, adding the term "foster adult" to the "foster child" category. Third, changing the layout of the relationship categories to save space on the form. These layout changes included the inclusion or exclusion of an "other relative" write-in and wrapping the relative categories into the second column or layering them with the relative categories all above the non-relative categories. The results of the testing showed that, for the most part these changes worked well. We did not recommend any changes to the tenure or relationship question. For the age question, while the two proposed changes worked well, we found that the font and layout of the age question de-emphasized the reference date and the age of several people whose birthdays were between the reference date, and the interview date was misreported. For this reason, we recommended that the layout of the question, whichever piece comes first, revert to the layout of the Census 2000 form. This includes not bolding the words "age" and "month/day/year of birth," and deleting the parentheses around the reference date.

Second, cognitive testing was also conducted to evaluate four characteristics of the form, including: (1) Space-saving features such as layout modifications and shorter instructions; (2) Revised roster questions, the inclusion of three household level questions and one individual level question to replace residence rules instructions; (3) Revised instructions for the Hispanic-origin item and the race item; and (4) Revised instructions for Person 1. The findings showed that of the space-saving design features tested in the date of birth and age questions, the question on gender, and the question on relationship of household members to Person 1 did not have a negative effect on respondent behavior. We identified some problems with the three household-level roster questions and changes in the wording were recommended. Our findings also showed that the intent of the person-level coverage question (i.e., the question to identify where the person stayed on census day) was not clear to respondents. Therefore, we recommended that this question be replaced with a coverage question from another panel in the 2005 National Census Test. Respondents did not have difficulty with the revised instructions in the race question. Regarding the Hispanic origin question, we suggested minor formatting changes and further cognitive testing to ensure that the revised instruction is not offensive to persons of Hispanic origin. We also tested revised instructions to Person 1. Our findings indicated that modifications in fonts and some changes in wording are needed. We recommended that these changes be subjected to further cognitive testing. Finally, our findings also indicated that the "continue note" instructions at the bottom of the first page should be repeated for persons 2-6. We recommended this change, if there is room on the form.

Third, cognitive testing was conducted on the content of the Nonresponse Follow-up instrument using a paper script containing the questions. Interviews have been completed and a report is currently being prepared.

*Staff:* Terry DeMaio (x34894), Jenny Hunter, Manuel de la Puente, LaToya Barnett, Kari Grow, Maria Bruun, Lorraine Randall

## C. Development of Race and Ethnicity Questions

Staff will participate in planning and pretesting alternative versions of the race and ethnicity questions used in the Decennial Census. We will develop proposals for cognitive testing of new question formats in conjunction with decennial staff, and lead or engage in cognitive research as needed.

During FY 2005, staff completed three phases of research to develop a shortened sequence of Hispanic origin, race and modified ancestry questions for the National Content Test. In all, nine formats were cognitively tested in a total of 212 interviews. Staff analyzed and presented results and recommended question formats for use in the national test. A research report detailing these findings was finalized. Staff consulted on the development of focus group research regarding the classification of American Indian and Alaska Native tribal groups.

*Staff:* Eleanor Gerber (x34890), Melinda Crowley

## 1.3 LANGUAGE PLANNING AND DEVELOPMENT
### (Decennial Project 5210503)

**A. Language Planning and Development Research**

Staff participate in the inter-divisional 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.

The team's emphasis in the first part of FY 2005 was on the design of a language assistance panel experiment for use in testing the bilingual Decennial questionnaire in the 2006 Area Test Census. Staff provided comments on the evaluation plans for this test. Due to funding  issues, the 2006 test of the bilingual form was subsequently cancelled. The bilingual questionnaire is still being tested as a part of the 2005 National Census Test.  During the second half of FY 2005, staff held several meetings with sponsors in the Decennial Management Division to plan for cognitive testing of the Spanish portion of the bilingual questionnaire. Staff developed a draft statement of work (SOW) and distributed it to the sponsor for review and comments. We incorporated comments into the SOW, and prepared a revised  SOW  to be released to prospective contractors.  Staff reviewed contractor proposals and made a recommendation to the sponsor to assist in choosing a contractor.  RTI, International was chosen to conduct the work.  Staff organized and ran a kickoff meeting with RTI in September of 2005 and will collaborate with RTI in conducting the research for the project.

*Staff:* Patti Goerman (x31819), Yuling Pan, Leslie Brownrigg, Manuel de la Puente

## 1.4 – 1.5   DATA COLLECTION PLANNING AND DEVELOPMENT/IT DATA COLLECTION PLANNING AND DEVELOPMENT
### (Decennial Projects 5310501 and 5311501)

**A. Field Research on the Usability of Handheld Computers (HHCs)**

The division's role is to collaborate with decennial staff in planning and conducting field-based usability tests to evaluate various configuration of navigational aids (in addition to the digital maps) for use with the HHCs in upcoming Census Tests.  As a basis for making recommendations, staff members propose and implement research designed to identify user-interface configurations that best support enumerator efficiency and accuracy.  This project is 2006 Evaluation 4.

During  FY  2005,  we  collaborated  with  the Demographic Surveys Division, Field Division, the Planning, Research and Evaluation Division, and Iowa State University in an investigation of the role of spatial ability in address-canvassing performance. We secured a licensing agreement from the Educational Testing Service to use one of its standardized tests in assessing the spatial ability of address canvassers in training. We helped in developing other assessments to be administered to trainees.  At the end of the year, data had been collected in Texas and South Dakota; and data analysis was in progress.  Results will inform the development or recommendations to reduce the HHC's demands on spatial ability.  A summer intern from Iowa State University continued developing a cognitive model of the address-canvassing tasks.

*Staff*: Betty Murphy, (x34858), Bob Bakondi (DMD), Andrea Johnson (FLD), Susanne Bean, Chris Johns (PRED), Sarah Nusser and Michele Rusch (Iowa State University)

**B. Interaction Design Support for the 2006 Non-respondent Follow-up (NRFU) Instrument**

Staff is supporting the design of user interactions for the 2006 NRFU instrument on handheld computer devices through a variety of activities.

During FY 2005, we provided design ideas and identified areas for improvement by planning and facilitating a design session with users and programmers. We conducted a study comparing the performance of older and younger adults with handheld computers which provided us with a better understanding of how to make handheld computer software easier to use for all user age groups. The results of this study also suggest alternative data entry interactions that need to be further explored.  We also conducted two rounds of usability testing on early versions of the 2006 NRFU Instrument, as well as one round of expert reviews.  The results of these tests and reviews were shared with the 2006 NRFU team and some usability recommendations were incorporated in the instrument.

*Staff:* Juan Pablo Hourcade (x33690), Betty Murphy, Erica Olmsted, Manuel de la Puente, Joyce Farmer, Lorraine Randall

**C. 2005 Internet Design Team**

This project focuses on iterative prototyping and development of an Internet form for 2010.  Our division's role in preparation for the 2005 National Census Test is to provide usability and accessibility expertise in the design and evaluation of two versions of the 2005 Census Internet form, a person-based version and a topic-based version.

During FY 2005, the current round of prototyping continued with staff reviewing and critiquing various products from the contractor.  Staff prepared two conference  papers  based  on  work  with  this

interdivisional team over the last several years. Staff worked with a sub-team to prepare a usability and accessibility test plan for the 2005 prototypes. We conducted usability and accessibility testing in May 2005; we presented results and recommendations to the design team, and we produced a comprehensive final report. The Internet form was launched on time in August 2005 as an option within the 2005 National Census Test. Staff began planning another round of testing to investigate design alternatives that might enhance the usability of Internet options in the 2008 Dress Rehearsal.

*Staff:* Betty Murphy (x3458), Larry Malakhoff, Dave Coon (DMD), Suzanne Fratino, Jennifer Lins, Sarah Brady, Susan Ciochetto, Myron Smith (DSCMO), Ann Ross (POP)

### D. Support for Field Data Collection Automation (FDCA) Program

Staff is supporting the FDCA program in writing a request for proposals, evaluating proposals, and collaborating with contractors once contracts are awarded. The FDCA program is responsible for the development and deployment of mobile computing devices for Census 2010.

During FY 2005, we informed key FDCA staff on how the Usability Lab could contribute to the FDCA program. We reviewed drafts of the Request for Proposals and provided feedback to relevant staff.

*Staff:* Juan Pablo Hourcade (x33690), Betty Murphy, Erica Olmsted, Manuel de la Puente, Tommy Wright

### E. Interfaces to Census Bureau Data Designed for Children

This research aims to provide middle school children learning programming in the Squeak environment with toolkits to create information visualization tools that access Census Bureau data.

During FY 2005, we evaluated the capabilities of Squeak in the areas of data access, graphics, and geographical information systems (GIS) capabilities. We found a lot of promise in Squeak and were satisfied with its data access and graphics capabilities. However, the GIS capabilities proved disappointing and would require further development before they could be used.

*Staff:* Juan Pablo Hourcade (x33690), Kate Guttridge

### F. Listing and Mapping Instrument (LAMI) Usability

The purpose of this project is to provide iterative usability evaluations of successive releases of the LAMI software. Since the software is implemented on handheld computers (Toshiba Pocket PCs), there are many design and development issues. Making the best use of limited screen real estate is a major challenge.

During FY 2005, after completing a comprehensive report on the usability testing of LAMI 2.2, we planned and conducted usability testing of the next software release, version 3.1. We presented a briefing on test findings, including recommendations for resolving usability issues; and we prepared a final report, which included video clips of test participants experiencing difficulties with the LAMI 3.1 user interface. This completed our near-term work on LAMI usability.

*Staff:* Betty Murphy (x34858), Juan Pablo Hourcade, Lorraine Randall, Joyce Farmer, Mohammed Chaudhry, Larry Malakhoff, Andrea Johnson (FLD), Teresa Luther, Cindy Falkenstein, and Dave Earles (GEO)

### G. 2006 Census Test Telephone Questionnaire Assistance (TQA) Implementation Team

This team will develop the specifications for the web-based telephone questionnaire assistance (TQA) application in support of the 2006 Census Test.

During FY 2005, we began planning by discussing the scope of the task and roles and responsibilities for team members. Internet Questionnaire Assistance was added as part of the task. The initial scope of this project was reduced. The interactive voice response component was eliminated. The team developed and modified existing content to be read aloud in English or translated to Spanish on the fly by human agents. The team updated documents from earlier implementations of TQA for the 2006 Census Test. We collaborated to identify and include key words to assist interviewers searching for answers to callers' questions.

*Staff:* Larry Malakhoff (x33688)

### H. Making Large Databases Accessible

This project will investigate the accessibility of tables with varying levels of column headers and row stubs which are generated from querying large databases such as the Longitudinal Employer-Household Dynamics (LEHD) Quarterly Workforce Indicators and the American FactFinder. This is a collaborative effort with Westat.

During FY 2005, staff contacted the American Foundation for the Blind (AFB) to form a research partnership with the Census Bureau and Westat. We collaborated with staff from AFB to further refine a research proposal to evaluate the accessibility of statistical tables. A kick-off meeting was held with Westat in September to determine the experimental design. We agreed that the number of rows and columns would be varied to test how table layout affects accessibility and ease of use for blind users. There will be 20 blind test participants, recruited through local establishments supporting blind persons. Work will continue next quarter to test the screens created by the

contractor before test participants are brought into the SRD Usability Lab.

*Staff:* Larry Malakhoff (x33688), Sid Schneider (Westat), Elaine Gerber and Corinne Kirchner (AFB)

### 1.6 SPECIAL PLACE/GROUP QUARTERS PLANNING AND DEVELOPMENT
#### (Decennial Project 5310508)

**A. Decennial Census Group Quarters Research and Planning**

The aim of this ongoing interdivisional working group is to improve group quarters listing and enumeration in the decade leading up to the 2010 Census. Two types of forms are being developed and tested. The first is the Other Living Quarters Validation Questionnaire, which is used for initial listing and typing of group quarters. The second is the Individual Census Report, used to enumerate group quarter residents.

During FY 2005, staff worked intensively with interdivisional SP/GQ working group members to redesign the Other Living Quarters Validation Questionnaire (OLQVQ), used to list and type GQs, and the Individual Census Report (ICR), used to enumerate GQ residents. Staff redesigned and restructured the OLQVQ to 1) improve the questions, 2) make the form easier for respondents to understand and for enumerators to administer, 3) reorganize the sections to improve the flow and 4) improve data quality. Staff completed this major restructuring within five months so that the improved instrument could be used in the 2006 Census Test. In addition, staff worked with this group to revise the ICR form. Staff proposed adding a new question to this form to determine if residents were living or staying in the GQ specifically on Census Day for 2 purposes: 1) to align GQ enumeration more closely with the decennial census residence rules and 2) to provide a new approach for resolving potential HU/GQ duplications using data on the ICR during the processing stage, potentially reducing time and costs associated with phone and field follow-up. The new question was accepted by Census Managers and GQ Working Groups. Staff planned, implemented, and completed cognitive testing of the ICR in a range of group quarters. Most of staff's recommendations, including the new question, were approved for the final ICR for the 2006 Census Test. Staff also participated in early planning of OLQVQ cognitive testing for 2006.

*Staff:* Laurie Schwede (x32611), Andy Jocuns, Lorraine Randall, Anna Chan, Joanne Pascale

**B. Ethnographic Study of Hotels and Motels**

Staff proposed and initiated an ethnographic study of hotels, motels, rooming and room and boarding houses and like establishments. Objectives of this study include estimating what proportion of these types of accommodations host residents long term (for a month or more) or cyclically, determining if establishments which encourage long term stays share any identifying characteristics, and understanding the situations of individuals and families which lead them to stay long term or cyclically in hotel/motel and like establishments.

During FY 2005, a draft report entitled "People Who Live in Hotels: An Exploratory Overview of People Who Reside in Hotels, Motels, and Like Accommodations and Their Hosts' Domain," was circulated for review by key staff in six divisions. The 75-page report describes settled residents, sojourners on open-ended stays, and the "regulars" cycling through accommodations across a spectrum of price and quality-from side-of-the-road motels and urban single room occupancy hotels, through the rapidly expanding "extended stay" hotels with complete apartment units, on up into luxury five-star metropolitan hotels. The report describes the features, "deals," locations, and connections to various organizations which entice people to settle, stay long-term, or cycle through particular hotels, motels, and the like. The report provides a glossary of legal, trade, and vernacular terms important in the accommodation sector; presents basic information on hundreds of brands; focuses on brands of extended stays and selling condos in transient hotels; illustrates companies which own five or more properties or 1000 rooms; presents tables on the sector and its population from Economic Census sources; and provides illustrations of conversions (former hotels into residential single-room occupancy, condominium residences, etc.). The annotated bibliography quotes text and concepts cited and referenced in the text including both print and Internet sources.

*Staff:* Leslie Brownrigg (x34995)

### 1.7 STATISTICAL DESIGN AND ESTIMATION
#### (Decennial Project 5610502)

**A. Decennial Editing and Imputation** (See Projects 0351 and 1871 (B) general Research - Statistical Methodology)

**B. Decennial Record Linkage** (See Projects 0351 and 1871 (A), General Research - Statistical Computing Methodology)

**C. Decennial Record Linkage Support**

The purpose of this project is to apply the record linkage software to a variety of projects, primarily in the Decennial area.

See Subproject F below.

*Staff:* Ned Porter (x31798)

## D. Research on Item and Count Imputation for Implementation in Census 2010

Research and studies will be undertaken on item and count imputation for implementation in Census 2010.

During FY 2005 and in collaboration with the Decennial Management Division (DMD), Decennial Statistical Studies Division (DSSD), and Planning, Research, and Evaluation Division (PRED), staff completed the implementation of software for the simulation of the mechanisms producing missing information in Decennial Censuses. The final product of the software is a set of "truth-decks," as termed by staff in the Population Division. Truth-decks replicate the conditions leading to missing information. Staff also focused their efforts on designing test statistics for quantifying the performance of various imputation methods when running them on the truth-decks. The test statistics include a goodness-of-fit chi squared statistic based on running an imputation method on one hundred independent truth-decks, and a non-parametric Kruskal-Wallis test. These test statistics were added to other descriptive statistics proposed by DSSD.

Three broad categories of count imputation methods have been implemented and evaluated in this simulation environment: 1) methods based on the hot-deck, as implemented for Census 2000; 2) methods based on administrative record retrieval developed by PRED; and 3) methods based on spatial modeling.

Staff wrote two papers explaining the evaluation methodology and giving general results, noting that no decision has been made yet concerning the 2010 Census.

Staff implemented the Canadian Census imputation software CANCEIS in the context of Census 2000 and of the 2006 Census test. The objective of this experiment is to test and validate the methodology in CANCEIS for possible future implementation in Census Bureau edit/impute applications. The focus is on the decision logic table (DLT) technology. DLTs enable the analysts to specify and manipulate sets of edits without having to modify the code of the edit/impute program. This allows a faster turnaround when edits need to be modified and it eliminates the risk of a programming error.

*Staff*: Yves Thibaudeau (x31706), Todd Williams, Bor-Chung Chen

## E. Decennial Disclosure Avoidance

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 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.

During FY 2005, staff worked with the Disclosure Review Board, the Policy Office, Research Data Centers (RDC), and the Planning, Research and Evaluation Division to develop procedures/rules for accessing decennial and ACS data at the RDCs.

Staff developed a test data set for testing disclosure avoidance procedures on ACS Group Quarters Data. Staff developed a new and improved version of the data swapping procedure and software that can be used for the Group Quarters Data. Staff tested and documented the software.

Staff developed a contract with Duke University for work on synthetic data.

Staff wrote SAS programs for variable selection and finding unique records in the ACS Public Use Microdata Samples. As it is currently done, variable selection is not a systematic process. The code aids in finding those variables that are the most important in determining whether a record is unique. A comprehensive search leads to a computationally hard problem. Staff found that some ACS variables, such as Age, do not transform easily to variables satisfying modeling assumptions. They worked on Box-Cox transformations. They get visually appealing results, but the Shapiro-test p-values are very small. They may end up categorizing variables and using models that do not require the normality assumption. They will run both linear and multinomial models and compare. Staff wrote a Bayesian Linear Model code in R to run on the strata they constructed. They partition the ACS household data records to build multiple models for the data, as one model cannot fit the entire data. Staff prepared version 1 of the ACS synthetic data. This version includes only one synthesized variable (Age). Future versions will have several more variables that are synthesized.

Staff revised the data swapping software so that it will run correctly on special censuses.

The Disclosure Review Board (DRB) reviewed the ACS base tables. There were too many tables with too much detail. A subgroup of the DRB met with ACS staff. They moved a lot of data from the very detailed tables into profiles that have threshold requirements.

Staff will continue work on the application of synthetic data methods to the ACS.

*Staff*: Laura Zayatz (x34955), Paul Massell, Phil Steel, Sam Hawala, Jeremy Funk, Rolando Rodríguez

### F. Census Unduplication Research

The Census Unduplication Research project began with the 2004 Census Test with the goal of improving the 2010 census unduplication. Initially, the unduplication was completed by the Decennial Statistical Studies Division. The next phase involves the unduplication of the 2006 Test Census by staff, which will incorporate Group Quarters being matched to housing unit data. Ultimately, staff will provide record linkage and modeling technology which will locate more duplicates in the census. Staff began this project in May of 2004.

During FY 2005, staff reviewed Census matching output from the 2004 Census Test and matching results from the 2000 Accuracy and Coverage Evaluation to help evaluate potential modifications to matching rules for the 2006 Census Test. Staff matched the Census 2000 Group Quarter (GQ) population in Alabama, Florida, Texas, and South Dakota against the housing unit (HU) population in those states and reviewed the output from the matching. Most of the between-state matches produced by these runs appear to be of questionable quality, however, the intra-state matching did indicate duplicates in the census. Staff began exploring methods for using name frequency to help classify possible duplicates in the 2010 Decennial Census. Drafts of the three main specifications for the Duplicate Person Identification (DPI) process in the 2006 Census Test have been produced. The first specification outlines the DPI process, the second documents the computer matching blocking passes and parameters, and the third documents requirements for modeling of the person links produced by computer person matching. The last two specifications also cover the analogous procedures in the Census Coverage Measurement computer person matching.

*Staff:* Michael Ikeda (x31756), Ned Porter

## 1.8 COVERAGE MEASUREMENT PLANNING AND DEVELOPMENT
### (Decennial Projects 5610503)

### A. Coverage Measurement Research

The synthetic method for small area estimation of census coverage will be evaluated using small area random effect modeling at the local census office (LCO) level. Staff will also provide support/assistance on other coverage measurement issues.

During FY 2005, the work completed for the small area coverage project, based on generalizing the synthetic method, was summarized and presented. Due to precision problems stemming from our method of estimating design effects, we investigated estimating design effects, conditional on model covariates. Using this approach, there is some evidence of more small area variability of coverage beyond that accounted for with the synthetic approach. This work is being completed and will be included in a final report, soon. However, questions of the model-fit and the levels of geography where variability occurs (if any) remain. In addition, staff members attend the weekly meetings of the Coverage Estimation Research Group of DSSD.

*Staff:* Don Malec (x31718), Jerry Maples

### B. Accuracy of Coverage Measurement

2010 Census Coverage Measurement 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. The focus is on the design of the census coverage measurement survey and 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 in the 2006 Census Test.

During FY 2005, staff provided technical expertise and experience in the planning of coverage measurement research for the 2010 Census. Staff served on three teams formed to plan and implement census coverage measurement research for the 2010 Census in the 2006 Census Test and with data from A.C.E. Revision II and Census 2000. Staff contributed technical expertise on the methodological design and requirements for measuring and estimating coverage error components. Staff collaborated on a framework for the estimation components of error designed to be consistent with estimates of net error. Staff completed analyzing data from A.C.E. Revision II in collaboration with internal and external researchers as part of a project to develop methodology to identify causes of extreme estimates of census coverage error for some small areas.

*Staff:* Mary Mulry (x31759)

### C. Questionnaire Wording and Automation Team (QWAT)

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 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

follow-up 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.

During FY 2005, staff led the Questionnaire Wording and Automation Team. The person interview (PI) specifications were completed and the Technologies Management Office began programming the instrument. Two rounds of usability testing were conducted on draft modules of the PI instrument. Although some changes to the instrument were made, the majority of the recommendations were not accepted since programming the automated instrument required more resources than anticipated. Staff presented a Census Coverage Measurement Person Interview questionnaire overview to a subgroup of a National Academy of Sciences panel. Staff also documented recommendations for demographic questions for an interviewer-administered automated instrument. These recommendations should be considered by the Census for 2008/2010.

Staff began crafting the person follow-up (PFU) questionnaire. Initially we crafted questions for an automated instrument, and worked on a business case for automating the PFU, but the decision was made to create a paper PFU questionnaire. Staff recommended a different battery of questions be tested for collecting residency information in the 2006 PFU. The proposed questions collect the specific dates that the follow-up person stayed at addresses. Collecting the dates is different than the PI residency questions which collect where the person lived most of the time and whether the person went back and forth between one or more addresses. The new approach we termed the "dates approach" while the PI approach we termed the "cycling approach." We contracted with Westat to cognitively test this new PFU approach. The approach seemed reasonable based on Westat's interviews. Westat could not say, however, which approach would collect more accurate residency information in a less burdensome fashion. In addition to comparing the two approaches in 2006, we also proposed a split panel test for the 2006 QDERS comparing the "dates approach" to the "cycling approach." We proposed conducting observations in the field in 2006 and then debriefing interviews after the observations to gather more data to determine whether the PI and PFU interviews are collecting correct information, or whether they are missing key aspects of a person's residency. The debriefing interviews immediately after an actual

interview is a new evaluation tool that we are planning to implement in 2006.

*Staff:* Beth Nichols (x31724), Jenny Hunter, Betty Murphy, Juan-Pablo Hourcade, Erica Olmstead, Amira Abdalla

## 1.9 COVERAGE IMPROVEMENT PLANNING AND DEVELOPMENT
### (Decennial Projects 5610505)

**A. Decennial Privacy Research**

The purpose of this project is to serve on and assist the work of the Privacy Policy and Research Committee (PPRC), and to conduct research to assess public opinion on privacy-related issues, including the increased use of administrative records to assist Decennial Census enumeration.

During FY 2005, staff actively participated in the meetings and deliberations of the PPRC, which included reviews of issue papers on the following topics: informed consent/SSN reporting/data linkage; procedures for off-site access requests for ethnographic/cognitive research; data stewardship employee awareness program; access to microdata in the Census Bureau's Regional Data Centers; a proposal to allow vendors to have access to SSNs in order to locate sample cases for the National Survey of Recent College Graduates; and the proposed collection of a DNA sample (via a self-administered buccal swab) from National Epidemiologic Survey of Alcohol and Related Conditions respondents. In addition, we conducted and documented research to examine the implementation of the new "RIP" (Respondent Identification Policy) procedures in the 2004 SIPP panel. The main findings of this research are: (a) The overwhelming majority of SIPP 2004 wave 1 respondents (91%) agreed to allow their answers to be used "as a starting point" during the follow-up interview. (b) The actual impact of RIP on the use of dependent interviewing is very small. Only 2% of all wave 2 person interviews, and 1% of all wave 3 person interviews, were not conducted with the use of dependent interviewing procedures because of RIP. ©) Declining the RIP request in one wave was associated with nonresponse in the subsequent wave.

*Staff:* Jeff Moore (x34975), Anna Chan, Eleanor Gerber

**B. Development of Questionnaires for Decennial Coverage Improvement**

Staff will develop a set of related data collection instruments which will be used to resolve duplicates in the Decennial Census. The project will begin with a pretest, which staff will participate in evaluating. Staff will participate with decennial staff in revision of the

instruments for use by clerical personnel, development of training, and additional pretesting for the 2004 Test, or for other mid-decade tests, and for the 2010 Census.

During FY 2005, staff developed presentations of the residence rules and coverage questions for use in the 2005 National Content Test. Cognitive evaluations of these presentations were organized and analyzed. A question to assist in unduplication of census forms from Group Quarters was developed and tested. A contract for the further testing of residence rules presentations and the evaluation of their implementation in the 2006 Census Test was written and awarded.

*Staff:* Eleanor Gerber (x34890)

### C. Inter-divisional Decennial 2010 Working Groups on Residence Rules and Coverage Improvement

These overall inter-divisional working groups provide input to the Decennial Management Division (DMD) for planning successive operations, and tests broadly related coverage research during the decade leading up to the 2010 Census. These groups receive proposals from various subgroups on: within-household coverage, residence rules, imputation, and unduplication.

During FY 2005, staff worked weekly with DMD and the inter-divisional residence rule working group to 1) revise the decennial census residence rules, 2) develop and test wording of residence rules and coverage questions for the 2005 and 2006 census tests, 3) review iterative study plans for evaluations of residence rules and coverage in those two tests, and 4) assess whether and how we should revise our tabulation rules on which types of group quarters will allow residents to be tabulated at a "usual home elsewhere." Staff developed a new coverage question on movers for the census short form; it was approved for testing. Staff helped plan the Fall, 2005 cognitive test of four questionnaires, then analyzed the results, recommended changes, and presented them to the working group and, with DMD, to decennial census decision makers for the 2005 Census Test. Staff's new question tested well and was selected as undercoverage question version 2 for 5 of the 20 form versions in the 2005 National Census Test. Staff also wrote a statement of work, evaluated proposals, and awarded the contract for, "Cognitive Testing and Evaluation of Coverage Issues and Residence Rules for the 2006 Test, for: A) Cognitive Testing of the Mail Form, B) Observations and Interviewer Debriefing, and C) Cognitive Interviews, Field Observation and Behavior Coding to Evaluate the Coverage Follow-up Instrument. Additionally, staff attended all of the National Academy of Sciences Residence Rule panel meetings and gave a presentation at one of them. Staff did the early form

development/review a new roster approach. Staff conducted cognitive testing of this form and analyzed and presented the results.

Staff also worked with DMD and the interdivisional Coverage Improvement Working Group, reviewing, suggesting, and prioritizing research questions on coverage improvement from interdivisional working groups on residence rules, within-household coverage, duplication and imputation for both the 2005 and 2006 Census Tests. Staff's proposal to add a new question to the ICR used in GQs asking about residents' actual presence in the GQ on Census Day was accepted by Census Managers. This new question aligns GQ data more closely with census residence rules, and, in tandem with collection of the actual date of the GQ's enumeration, might improve coverage by resolving some HU/GQ duplications without follow-up interviews, possibly lowering costs. This question passed cognitive testing and is included on the final ICR for the 2006 Census Test.

In the Complex Households ethnographic project, Staff shepherded the book manuscript, *"Complex Ethnic Households in America"* through the successive stages of review and revision. Staff incorporated final revisions and sent the manuscript to AltaMira Press. The book is in press and will be released in January, 2006.

*Staff:* Laurie Schwede (x32611), Eleanor Gerber, Manuel de la Puente, Jennifer Hunter, Anna Chan, Patricia Goerman, Nita Rasmann, Tina Arbogast, and five outside authors

### 1.10 AMERICAN COMMUNITY SURVEY (ACS) (Decennial Project 5385560)

#### A. ACS Questionnaire Design Measurement

This project provides technical and research support for the development and improvement of ACS data collection instruments used in all modes of data collection available in the ACS. Staff serve on inter-divisional working groups and provide technical support in the design, and conduct questionnaire design research for the ACS.

During FY 2005, staff served on two working groups: the ACS Evaluation Team Working Group and the ACS Language Team. Both groups are chaired by staff from the Demographic Statistical Methods Division. Staff observed ACS interviews in the Tucson, AZ area with Spanish language speaking households, and prepared a draft trip report. Staff began but did not complete a proposal for the development of an ACS Spanish language form. Staff also produced a proposal for the conduct of ethnographic observations of CAPI interviews with non-English-speaking households. We

finished and circulated a trip report for the Tucson, AZ observations. Staff also conducted pretesting of the labor questions in the ACS, for the inclusion in the ACS 2006 Content Test. (See next project.)

Regarding group quarters research, staff collaborated with the sponsor on the development of debriefing questions for a small field test in February 2005. Staff also worked with the sponsor to help develop a training module for the February 2005 field test. Staff participated in this field test and collaborated with the sponsor on the conduct of the debriefing interviews. Staff developed and conducted a three-hour debriefing session with Field Representatives, debriefers, working group members, the Decennial Management Division and the American Community Survey Office on the February 2005 field test.

*Staff:* Jennifer Rothgeb (X34968), Yuling Pan, Laurie Schwede, Patti Goerman, Manuel de la Puente

**B. ACS Labor Force Questions**

The purpose of this project is to develop recommended question wording for inclusion in the ACS 2006 Content Test. Evaluation of the 2000 Decennial Census (of which the ACS questions are extracted) labor force questions indicate that responses to those questions produced lower employment estimates than the Current Population Survey. An Office of Management and Budget (OMB) Interagency Subcommittee on ACS Labor Force Questions was created and charged with producing a research plan to develop test wording for the 2006 ACS content test.

During FY 2005 and prior to conducting cognitive interview research, we evaluated the current ACS labor force questions using expert review, behavior coding of ACS-CATI interviews, and interviewer debriefings of ACS-CATI interviews to provide information to develop alternative question wording. Using two sets of revised questions, forty cognitive interviews were conducted using the three modes of interviewing used in ACS (self-administered, CATI, and CAPI.) Resulting recommendations were adopted for the 2006 ACS Content Test.

*Staff:* Jennifer Rothgeb (x34968), Joanne Pascale, Jeff Moore, Jim Esposito (Bureau of Labor Statistics), Karen Bogen (University of Massachusetts), Tom Palumbo (HHES).

**C.1 ACS Small Area Estimation Research – Tract Level Coverage and Variance Reduction using Administrative Records**

A proposed method using matched administrative records as tract level controls to reduce coverage error and variability will be evaluated.

During FY 2005, we received permission to use IRS-derived data. We are working with the Planning,

Research and Evaluation Division to obtain the data needed for the project.

*Staff:* Don Malec (x31718), Lynn Weidman, Jerry Maples, Elizabeth Huang

**C.2 ACS Small Area Estimation Research – Household and Person Level Weighting at the County Level**

There may be inconsistencies in county level estimates of occupied number of housing units using housing unit weights versus using householder's weight. The underlying causes of this inconsistency and alternative weighting methods to reduce the inconsistency will be investigated.

During FY 2005, we completed this project. ACS staff has selected raking procedures.

*Staff:* Don Malec (x31718)

**D. ACS - Missing Data and Imputation**

This project undertakes research and studies on missing data and imputation for the American Community Survey.

During FY 2005, staff completed a race imputation study based on the matched ACS sample, C2SS. The performance of race imputation is quantified for three relationships to the householder: householder; spouse; child. Based on ACS race response, census race imputation among non-Hispanics is very accurate. Race imputation for non-Hispanic children is even more accurate. Staff plans on quantifying imputation performance separately for children whose race is substituted, and children whose race is allocated. Staff plans on quantifying the performance of ACS race imputation based on census responses, provided enough sample is available.

Staff initiated a comparative analysis of the imputation of the labor force and earnings item involving the Current Population Survey and ACS. Emphasis is given first on identifying the natural advantages of each survey on the other. Additional emphasis is given on the performance potential of the imputation techniques of each survey. These two aspects serve to classify the quality of the imputation of a survey item in one of four categories: 1. Natural advantage, superior technique. 2. Superior Natural advantage, inferior techniques. 3. Natural disadvantage, superior technique. 4. Natural disadvantage, inferior technique.

The ultimate objective of the research is to develop a general imputation methodology that can be implemented consistently across surveys at the Census Bureau.

*Staff:* Yves Thibaudeau (x31706)

**E**. **ACS - Disclosure Avoidance (See highlights from Decennial Project 5610502-E)**

**F.   ACS Weighting Research**

Group quarters (GQ) weighting and estimation has been carried out only once with ACS data, for calendar year 1999 when there were 36 counties in sample. At that time, GQ stratification and sampling was done separately for each county.   For the full GQ implementation of ACS starting in 2006, a new GQ sort by type within state will be used.  Now that every county in the nation will be in sample, there is the possibility of weighting GQ persons by county or state. A simulation study is comparing options for weighting GQ persons by county or state and controlling GQ person estimates, either by themselves or together with housing unit (HU) person estimates.   A research proposal was developed for determining appropriate methodology for estimating the number of persons residing in GQ and their characteristics.

During FY 2005, staff completed a study plan and carried out the study to compare four alternatives for controlling HU and GQ person estimates of demographic characteristics totals to intercensal population estimates as part of ACS weighting. One hundred annual data samples were simulated from Census 2000 data files for 14 states and four separate sets of weightings were performed for these states. Absolute deviations of estimates of the number of GQ persons from their controls for 10 demographic characteristics were compared for states and counties. Summaries of results were presented and discussed at meetings.   Additional evaluations of estimates of population by major GQ types and of coefficients of variation for major GQ types and the 10 demographic characteristics were also completed and presented.  The results are being used to help inform the decision on how the weighting procedures should be modified when GQ persons are included in ACS along with HU persons.

*Staff:*   Lynn Weidman (x34902), Michael Ikeda, Julie Tsay

**G.   American Community Survey (ACS) Usability Study**

The purpose of this study is to discover usability problems with the American Community Survey (ACS) domain of Census.gov.  Our objective is to assess the efficiency, accuracy, and subjective satisfaction of the site, note the usability violations, and make recommendations for how to improve the site.

During FY 2005, staff conducted a usability study of the integration of the ACS data into the American FactFinder(AFF) Web site.  We observed typical users and took accuracy measures, efficiency measures, and

subjective satisfaction ratings of the site. Staff analyzed the results and presented a written report with video clip highlights to the development team with recommendations on areas to improve the website. Staff reviewed the iterative results of both rounds of testing and demonstrated in a final report the improvement of the site with the incorporation of our recommendations.

*Staff:*  Erica Olmsted (x34893)

**H.   ACS Group Quarters (GQ) Research**

The aim of this inter-divisional project is to develop, test, and revise instruments, methods, and procedures intended for use when the ACS begins to collect data for the first time on a routine basis from group quarters in January 2006.  The starting point for this questionnaire was the Other Living Quarters Validation Questionnaire (OLQVQ) used in the decennial census site test in 2004; the content is being modified to meet specific ACS objectives and needs, and the questionnaire is being automated. A small-scaled field test of the new instruments and methods is scheduled for February, 2006.

During FY 2005, staff collaborated with the sponsor and interdivisional ACS GQ working group members to plan, develop, conduct, and analyze the results of the ACS GQ field test.  At the beginning of the project, staff presented results and implications from previous relevant research, and, together with a GQ field expert, offered recommendations to the full working group on the general types of GQs to include in the field test as well as on specific GQs to sample. Staff developed debriefing questionnaires for GQ contact persons and for project staff; assisted in training colleagues to conduct debriefings and observe respondent behavior; participated in planning the test and in reviewing other project materials. During the data collection period, staff teamed up with an ACS field representative to conduct GQ contact person interviews and debriefings in four GQs, as well as participated in nightly phone conferences about the progress of data collection.   Staff wrote interview summaries and developed a list of overall issues and recommendations. Staff also planned, developed, and conducted a final project debriefing session with all project staff and summarized results.  Finally, staff participated with all team members in assessing all recommendations from this test and reviewed the draft report.

*Staff:*  Laurie Schwede (x32611)

**A.   ACS Language Research**

This project provides technical and research support for addressing language issues in ACS data collection instruments and supporting documents. Staff

serve on inter-divisional working groups and provide consultation and technical support in the design and development of language research for the ACS.

During FY2005, staff were active members of the ACS Language Team. We completed draft literature review on Census Bureau language research with a focus on Spanish language research. We developed a draft proposal for pretesting the ACS Spanish language instruments. We also helped design a cognitive pretesting project to test translations of ACS CAPI survey letters and information brochures in multiple languages. We developed a statement of work for this project and the contract was awarded to start the research in FY2006.

*Staff:* Yuling Pan (x34950), Patti Goerman, Manuel de la Puente

### 1.11 DATA INTEGRATION
### (Demographic Project 0906/7374)
#### A. Data Integration

The purpose of this research is to identify microdata records at risk of disclosure due to publicly available databases. Microdata from all Census Bureau 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.

During FY 2005, staff created documentation for the scripts that were developed that include where the files are located, how they can be used and where input files can be found or how they can be created. Staff also documented more generally the process of attempting to reidentify records. This includes complications as well as further opportunities. Staff also documented how to use the data ferret and iopus software and noted why these tools are useful in this project. Further, staff documented suggestions about next steps in record re-identification. Due to a change in staff members, previous work on the data integration project is being collected and reviewed so that the project may be continued and progress made with as little replication of previous results as possible. One project focus is working towards a standardized and automated process of record reidentification, or as much of one as possible, for which specific programming skills are necessary. Working with and automating the use of web-based search engines and databases will be an important capability and will be the focus of immediate efforts.

Staff worked on improving matching parameters for the Longitudinal Employer Household Dynamics (LEHD) project, which links Survey of Income and Program Participation (SIPP) data with Social Security Administration (SSA) earnings data. Staff analyzed results obtained from two of the LEHD matching runs.

Each run was conducted on a different implicate or version of the synthetic data. The matching was done on variables including Birth Date, Ethnicity, Education, Industry, and Occupation. The matching was done within blocks defined by the variables Sex, Black, Education, Marital Status, Sex of Spouse, Black Spouse, Education of Spouse, and Age of Spouse. Out of over 215,000 records, about 1000 records exactly matched on at least 3 variables. Staff reported on matching run time and results of preliminary matching runs to the LEHD group. Staff prepared two files that describe what has been done and how: "readme" and "summaryofresults." The file "summaryofresults" explains the results, which were produced by the two matching runs "finalrct3.out" and "finalrct4.out." Staff prepared the EM-algorithm software to run on the data, eventually to find out that the blocking strategy for the matcher resulted in blocks that are too big to obtain meaningful results from the EM program. Staff then worked on selecting other variables to block on. The original scheme for blocking included only variables that were not synthesized and therefore resulted in blocks that are too big. A future blocking strategy will include synthesized variables that remain relatively stable within groups. There are dozens of variables to choose from. Staff will work with others to select and categorize variables to use for blocking. For the same project, staff also ran the one-to-one matcher on the same data and found very few true matches (1/1000 of 1%) and a few false matches (1/10 of 1%). The one-to-one matcher is not considered the most powerful record linkage software for this purpose. It has a criterion for finding the "best" match among all possible matches. "Bigmatch," which staff used initially, is the most powerful software available for this task. It provides all possible matches. So far, the results suggest that the Disclosure Review Board's requirement for a roughly equal number of false and true matches is reasonable.

Staff worked with staff from the Demographic Surveys Division and the Demographic Statistical Methods Division to develop new disclosure avoidance procedures for the Current Population Survey. These procedures address minor problems found in our reidentification studies. Staff worked with SIPP staff to discuss potential alternative solutions for protecting life events that occur when respondents are in sample. We developed two new alternative solutions, both of which were approved of by the Disclosure Review Board.

Staff finished the latest IDFinder revision. Staff improved the performance of the Time Series Viewer using the DirectX technology (graphics programming). Staff finished their work on changing the overall system data structure to improve system performance and extensibility.

Staff will work on preparing a final report for the LEHD matching work. We will also continue working

on the re-identification experiments.

*Staff:* Sam Hawala (x34956), Laura Zayatz, Phil Steel, Paul Massell, Jeremy Funk, Sherae Daniel, Hyunmo Kang (University of Maryland)

## 1.12 SIPP METHODS PANEL
### (Demographic Project 1461)

The SIPP Methods Panel is the R&D vehicle for development of a redesigned SIPP data collection instrument for the 2004 SIPP Panel. Through a combination of expert review, user needs assessment, secondary data analysis, and laboratory research, Methods Panel (MP) staff carry out the research activities necessary to implement the recommendations of the Continuous Instrument Improvement Group (CIIG).

During FY 2005, our primary focus was conducting research to evaluate the impact of new questionnaire design features on data quality and nonresponse. Staff completed an evaluation of the impact of the new asset screening procedures on asset ownership estimates in the 2004 SIPP panel [see *Moore, J. (2005), "An Evaluation of SIPP 2004 Wave 1 Asset Ownership Reports," U.S. Census Bureau, Statistical Research Division, Research Report Series (Survey Methodology #2005-01), issued March 8, 2005*]. This research found the expected reduction in survey burden as a result of the screening procedures, but no negative impact on data quality, and some evidence of possible quality gains arising from other instrument changes. We also completed preliminary analyses comparing seam bias in wave1-wave2 of 2001 vs. 2004, to assess the impact of new dependent interviewing procedures. The preliminary results were quite positive, showing significantly reduced seam bias for 8 of 9 characteristics examined. Staff also analyzed SIPP's new Respondent Identification Policy (RIP) procedures (introduced for the first time in the 2004 panel), which require prior consent before dependent questions can be asked of someone who did not originally provide the dependent information. This analysis found that the actual impact of RIP on the ability to use dependent data was trivial – in less than 2% of wave 2 interviews, for example, did RIP restrict the use of dependent data. And we examined the effects of new questionnaire features on nonresponse for income amount items. An evaluation of the expanded use of closed-ended "range" follow-ups for asset amount nonresponse in wave 1 found that this strategy was very successful at converting "don't know" nonresponse (about 75% of those who originally responded with a "don't know" to an exact amount went on to provide a response to the "range" follow-up), and surprisingly successful with refusers (25%) as well. Additional research was also carried out comparing item nonresponse in 2001 and 2004 for assets and for

"general income" sources. Reports of the results of this work are under review, and will be issued early in FY2006.

In addition to research work, staff also participated in several working groups (comprised primarily of production staff of the Demographic Surveys Division/SIPP and subject-area experts of the Population Division and the Housing and Household Economic Statistics Division) whose task was to identify problems in preliminary SIPP 2004 panel data and to develop instrument and/or processing system solutions. Finally, staff continued work on a collaborative project with HHES staff to develop and test instrument enhancements to obtain improved wealth data in SIPP's Assets and Liabilities topical module – namely, new questions on the cash value of assets and trusts. We completed a draft set of proposed new questions which will be cognitively tested in FY2006.

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

## 1.13 SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP) RESEARCH
### (Demographic Project 1465)

### A. Longitudinal Weighting

The objective of this project is to design and conduct research required to assess the effectiveness of weighting alternatives for the SIPP longitudinal estimation.

During FY 2005, staff completed a study of weighting alternatives to compensate for the effects of longitudinal nonresponse on cross-sectional SIPP estimates. We identified several of the survey's core items for which there appears to be fairly consistent nonresponse bias when the current weighting procedure is applied. Moreover, we recommended revisions in the current longitudinal nonresponse adjustment methodology, which would incorporate the careful modeling of response propensity, more use of historical data, and the efficient selection of nonresponse weighting cells.

Staff also proceeded with the extension of the longitudinal reweighting research to include the consideration of empirical results from the 2001 panel, variance estimation for nonresponse error estimates for cross-sectional items, and an evaluation of the effects of longitudinal nonresponse on estimates for longitudinal variables.

*Staff*: Leroy Bailey (x34917), Eric Slud , Julie Tsay

### B. Quick Turnaround Pretesting of Household Surveys

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.

During FY 2005, staff conducted a second round of cognitive interviews and completed a research report documenting the findings of research to evaluate new and revised questions, specifically questions about respondents' attitudes on the friendliness, respectfulness, and professionalism of police officers during contacts at traffic stops, and questions about the use of force and the race of police officers making the traffic stops. Results showed that the question about how many officers were present assumed a constant number, while some respondents reported one officer at the beginning and additional backup officers coming later. Also, the initial version of the question about the race of the police officer obtained information about all of the officers, rather than the one(s) who made the initial contact. This information is necessary to measure the occurrence of racial profiling. Finally, while assessments of the police officers' respectfulness and professionalism were easily answered by respondents, their friendliness was seen as inappropriate to measure and difficult to evaluate when several officers were present.

Staff also conducted focus groups and cognitive interviews with stalking victims to develop a Stalking Supplement to the National Victimization Survey. One of the main issues discussed in the focus group was the range of emotions respondents might experience and how these feelings change over time. Respondents were asked to discuss issues regarding "group" stalking, how to quantify the frequency of behaviors, and safety concerns in data collection if a stalker is an intimate partner. Results of the focus group showed that for respondents who had experienced more than one stalker, it was not only possible, but easy for them to identify the person whose behavior they considered the most serious. In most cases, there was not a consistent interval at which the behaviors occurred. For example, for months the behavior would occur every day; then it would stop for a period of time and start up again later. Cognitive interviews using the questionnaire developed after the focus group have been completed and a report summarizing the research results is currently being prepared.

Staff collaborated with staff from the Demographic Surveys Division (DSD) to conduct cognitive interviews on the Survey of Fishing, Hunting, and Wildlife Associated Recreation. Activities involved reviewing materials, conducting cognitive interviews, preparing interview summaries, providing guidance on report preparation and reviewing draft reports. The interviewing and recommendations are complete on this project, and the report is nearing completion.

*Staff*: Terry DeMaio (x34894), Kristen Hughes, Lorraine Randall

### 1.14 SIPP ASSETS/LIABILITIES IMPUTATION RESEARCH/SOFTWARE DESIGN (Demographic Project 7558111)

This project will undertake research regarding imputation of assets and liabilities data items for the households sampled in the SIPP panels.

During FY 2005, staff completed the first phase of the collaborative imputation study managed by the Housing and Household Economic Statistics Division and sponsored by the Social Security Administration (SSA). The study focuses on the difficulties of the current imputation process for the assets and liabilities in the topical modules of SIPP. Among these difficulties is the dilution of the correlation between assets and liabilities through the imputation process. Staff implemented fully functional imputation methods designed to correct the problems identified by the SSA. Staff identified variables to reinstate coupling between assets and liabilities when imputing the value of real estate. Staff designed a model-based system using mortgage amount to predict real estate value. Staff also designed a predictive mean version of the model-based system. The predictive mean approach is technically the same as the hot-deck approach. Therefore, it is simpler to implement in a production environment. The methods keep the correlations between imputed assets and liabilities at a level compatible with the observed correlations. Staff took special care to ensure that the imputation methods were compatible with the processing environment of the demographic surveys at the Census Bureau. Therefore, upon approval from other interested parties, staff plan to assist the Demographic Surveys Division in implementing the method in a production environment. Staff also plan to expand on the prototype to integrate the imputation of other asset/liability variables and maintain the relevant correlations.

*Staff:* Yves Thibaudeau (x31706), Todd Williams

### 1.15 AMERICAN HOUSING SURVEY-METRO (Demographic Project 7478)

This project provides questionnaire pretesting assistance for the American Housing Survey (AHS).

During FY 2005, staff worked with the Demographic Surveys Division (DSD) and the Housing and Household Economic Statistics Division (HHES) to craft and test new and revised questions for six modules in the AHS: housing subsidies; fire safety; neighborhood quality; real estate transactions; mortgages; and initial fees associated with senior housing and services.

In Round 2 AHS Research, Housing Subsidies: Cognitive interviews were conducted with low-income housing respondents in the Washington, D.C. metropolitan area, Detroit, MI, and Chicago, IL to evaluate proposed revisions to AHS housing subsidy questions. Results showed that: 1) unfamiliar terms (e.g., voucher, housing authority, rent control, rent stabilization) and new terms (e.g., landlord) presented comprehension problems for respondents; 2) some revised questions (e.g., two screener questions used to have respondents self-identify subsidy status) performed better than the original AHS questions. Research results of Round 2 of the AHS cognitive researched were posted to the SRD website in January 2005. Staff then met with DSD, HHES and the Department of Housing and Urban Development (HUD) to discuss recommendations based on the housing subsidy research. Based these discussions, these changes were made to the housing subsidy module: question order was revised to minimize false positives and reporting burden; question wording was changed to include clearer reference to new housing programs; and the screener questions for the series were modified to further distinguish reporting requirements of subsidy recipients from other renters. In January, results of Round 2 research were posted to the SRD Studies Series website [*SS (Survey Methodology #2005-01)*].

Round 3 Research, Fire Safety, Mortgage, Neighborhood Quality, and Real Estate Transaction Modules: Cognitive interviews on the remaining AHS modules were completed in January 2005. Draft results and recommendations were distributed beginning in February through April 2005. Final reports will be issued on the SRD website.

Initial Fees, Senior Housing and Services: On DSD's request in February, staff wrote a project plan, budget and research protocol to conduct ethnographic and cognitive research for new questions on costs associated with senior housing, services and initial fees. With Georgia State University, staff arranged and conducted field visits to senior housing places in Atlanta on March 3. After briefing DSD and HHES, HUD has decided to continue its "wait and see" strategy, to develop a consensus among its own staff and data users about what should be measured in order to meet policy objectives. The Atlanta field trip established that if detailed senior housing cost questions were dropped in favor of overall cost items, analysts could use hedonic rent functions to isolate the marginal cost of additional services and housing options.

Presentation on "Interviewer Effects" at AHS Supervisors Conference: In conjunction with a new DSD interviewer-monitoring program (a.k.a. PANDA), staff gave a survey methodology talk on Interviewer Effects at the March 3 AHS Survey Supervisors Conference in Atlanta. Using PANDA data, staff demonstrated the impact of varying average workload on sample efficiency, data quality, and survey costs given different question characteristics and their associated intra-class correlation coefficients.

Analysis of "Reasons for Moving" to a Neighborhood: In another effort for HHES, staff conducted thematic analysis of respondent-reported reasons for choosing a specific neighborhood and purchasing a specific home in order to develop more exhaustive and mutually exclusive response options for those questions.

Field Pre-test of AHS for Fall 2005: The new AHS is being field tested in Pittsburgh and Tucson in fall 2005. Staff assisted DSD with OMB clearance paperwork and briefed DSD on options for the content of a respondent debriefing survey and the protocol for two Field Representative debriefings in November.

*Staff*: Eileen O'Brien (x32695), Terry DeMaio

## 1.16 NATIONAL SURVEY OF YOUTH VOLUNTEERING & CIVIC ENGAGEMENT (Demographic Project 7565084)

The purpose of this project is to evaluate questionnaire problems, conceptual difficulties, recalling related challenges and sensitivity issues with regard to service earning and community involvement activities for the Youth Volunteering Service and Civic Engineering Survey. Staff is responsible for OMB research clearance, respondent recruitment, analysis, oral briefings to internal and external sponsors, recommendations, and a final report.

During FY 2005, staff conducted cognitive interviews to pretest a new questionnaire, the Youth Volunteering, Service, and Civic Engagement Survey, administered to 13- to 18-year olds. Results of the interviews showed that there was a differential interpretation of terms for respondents of different ages in the survey population. For example, among older respondents, "partying" meant going to social events with a large number of people in attendance, while to 13-to-15-year-olds, this term referred to sleep-overs and birthday parties. As a result, this term was removed from the questionnaire. Another example was "get a bachelor's degree," which was not understood by younger respondents, although they did understand the phrase "graduate from a 4-year college." Respondents also found a question dealing with trusting different groups of people to be extremely sensitive.

*Staff:* Kristen Hughes (X38458), Melinda Crowley

## 1.17 NATIONAL SURVEY OF COLLEGE GRADUATES - COGNITIVE RESEARCH PROJECT
### (Demographic Project TBA)

The National Science Foundation (NSF) requested the Census Bureau to conduct cognitive research on the National Survey of College Graduates (NSCG) self-administered questionnaire. NSF and the Demographic Surveys Division (DSD) highlighted questionnaire items which were high priority test items. The project is a two-phase project with two rounds of cognitive interviews (20 interviews in each round). The target population is persons who received college degrees and advanced degrees in science and engineering.

During FY 2005, we tested two versions of a modified NSCG questionnaire during the Phase One testing. Our results demonstrated that several items, other than those targeted by NSF, were in need of revision. We proposed alternative design for several items for use during Phase Two testing. NSF agreed to some of our recommendations, with some modifications. We conducted Phase Two testing and identified several issues we are confident are the major problematic areas of the NSCG questionnaire. While NSF did not agree to all our recommendations, it did agree to conduct a split-panel field pretest and to do a post pretest telephone follow-up to determine if there is error with persons reporting that they have changed jobs when in fact, they have not. There are still many revisions to the NSCG which may be necessary in future years in order to improve survey quality. We informed NSF that a more comprehensive long-term research plan will be necessary to address the various issues and have time for appropriate testing.

*Staff*: Jennifer Rothgeb (x34968), Jenny Hunter, Jeff Moore, Lorraine Randall

## 1.18 NONRESPONSE BIAS ANALYSIS
### (Demographic Project 7485000)

Staff is conducting a nonresponse bias analysis for the 2003 National Survey of Recent College Graduates (RCG), sponsored by the Demographic Surveys Division (DSD) and the National Science Foundation (NSF). The findings will be used by NSF to obtain Office of Management and Budget (OMB) approval to conduct the next RCG in 2006.

During FY 2005, staff developed a statement of work to conduct a two-month preliminary nonresponse bias analysis. This statement of work was vetted through SRD and DSD and was ultimately approved by NSF. Staff replicated the 2003 response rates and calculated 2003 response rates at the nonresponse adjustment cells. In addition, staff constructed a proxy for nonresponse using late responders.

These analyses suggest that there may be benefits

to not collapsing over males and females when creating the adjustment cells. In addition, staff constructed a proxy for nonresponse using late responders. Logistic modeling was used to explore possible relationships between early vs. late responders and variables not currently used in the creation of nonresponse adjustment cells. One variable was predictive of early vs. late responders and analyses are planned to assess whether current nonresponse adjustment cells yields biases estimates for this variable. No pattern in nonresponse bias was discerned when reweighting late responders and late ineligibles as proxies for nonrespondents.

*Staff:* Aref Dajani (x31797), Jerry Maples, Adam Carle

## 1.19 RESEARCH FOR SMALL AREA INCOME AND POVERTY ESTIMATES (SAIPE)
### (Demographic Project 7165)

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

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

During FY 2005, some concerns were raised by HHES staff about large changes over time in the estimates for some states. We thus examined year-to-year changes in the poverty ratio estimates for age 5-17 for certain states (CA, CO, DE, IA, MD, MA, NV, NM, RI, SD, VA) whose estimates decreased substantially from income year (IY) 2000 to 2001 and increased substantially from IY 2001 to 2002. We noted that roughly similar changes were observed in the direct state estimates from the Current Population Survey's Annual Social and Economic Supplement (CPS ASEC). We thus attributed the large changes in the estimates to a combination of changes in the true poverty ratios and also to some random variation, given the large sampling errors in many of the direct estimates.

We observed that the District of Columbia (D.C.) is well away from other states in regard to the values of its regression variables in the CPS equation of the state poverty model for IYs 2000-2002. This raises some concerns about whether a model (including its

parameter values) appropriate for the other 50 states is also appropriate for D.C.? Fortunately, D.C. does not consistently appear to be an outlier in the model fits, ameliorating some concerns. The results suggest, however, that future research examine whether D.C. might be more appropriately treated as a county, with its estimates produced from the SAIPE county model.

We wrote a draft report documenting results of an empirical study on using ACS Supplementary Survey data for IYs 2000-2002 in SAIPE state poverty models. This was examined by using a bivariate model (for CPS and ACS data) and making various assumptions about relations between the two equations. Empirical results showed that use of the ACS data yielded only small average improvements when minimal assumptions were made, but larger average improvements when we assumed that the regression coefficients (apart from the intercept) were the same between the two equations. This improvement depends, though, on the underlying assumption being exactly true. However, we also found occasional large increases in posterior variances (from either model) corresponding to states with a large regression residual in the ACS equation. Our recommendation, therefore, was not to adopt the bivariate model at this time, but to do more research on a model that assumes regression parameters in the two equations are related but not identical. The report was reviewed by Harvard University and the University of Michigan. They agreed with the overall conclusions of the report and also made some technical comments that suggested areas for further research. We will also need to redo some of this analysis when data from the full production ACS become available.

HHES had tabulated IRS income tax data for school districts, so we evaluated several predictive models for school district population and poverty that made some use of these newly tabulated IRS data. The goal was to see if the updated information available from the IRS data would improve poverty estimates relative to the official method which uses no updated data for school districts, but instead assumes school district to county shares of poverty remain unchanged from the previous census. The evaluation was done by constructing estimates for 1999 from the current production method and from various models that made use of the IRS data for 1999, and comparing these estimates to corresponding estimates from Census 2000 (which are also for IY 1999). This evaluation showed some improvement from use of the IRS data. However, we also did an analogous evaluation by going backwards in time to construct estimates for IY 1989 and comparing these to estimates from the 1990 Census, and could not establish clear improvements in this setting from use of the IRS data. The reasons for the discrepancy between these two evaluations are, at present, unclear. They may have something to do with

difficulties in geocoding some IRS tax returns to census blocks (from which school district tabulations are constructed), a more significant problem in 1989 than 1999.

*Staff*: Elizabeth Huang (x34923), Jerry Maples, William Bell (M&S)

### B. Small Area Estimation Methodology for SAIPE

This work generally concerns the methodology of small area estimation methods applied to log-counts of child-poor and log-rates of child poverty at the level of county SAIPE estimates, based upon linear (Fay-Herriot) models, General Linear (mixed-effect logistic) models, and some variants of them.

During FY 2005, staff worked on completion and revision of two journal papers on Small Area Estimation Methodology in SAIPE, both with T. Maiti as co-author. Final revisions in response to referee comments were made and submitted on the paper, "MSE estimation in transformed Fay-Herriot model," which is now accepted and will appear in the *Journal of Royal Statistics Society*, *Series B (JRSS-B)*. A second paper, "Small-area estimation based on survey data from a left-censored Fay-Herriot model," was also completed and submitted to *JRSS-B*. This work is related to SAIPE county-level child-poverty estimation, specifically small-area estimation using Fay-Herriot models when data are left-censored, e.g., when as in SAIPE, counties with responses below a known threshold (i.e, CPS sampled count of poor children is equal to 0) are dropped from analysis.

*Staff:* Eric Slud (x34991)

### 1.20 ANALYSIS AND FORECASTING OF DEMOGRAPHIC TIME SERIES
### (Demographic Project TBA)

This project will employ time series methods to incorporate probabilistic estimates of forecast uncertainty into annual population projections.

During FY 2005, we produced final results for point forecasts of immigration. For immigration age distributions, we developed and implemented a method that combined logistic transforms, principal components analysis, cubic spline smoothing, time series modelling, and Bayesian methods; we applied this method to produce smooth 50 year forecasts. For immigration numbers, we used time series modelling with outlier adjustment and robust regression techniques to generate 50 year forecasts. We presented research results at the 2005 Federal Forecasters Conference.

*Staff:* Tucker McElroy (x33227), William Bell (M&S)

## 1.21 EDITING METHODS DEVELOPMENT
### (Economic Project 2320554)
**A. Investigation of Selective Editing Procedures for Foreign Trade Programs**

The purpose of this project is to develop selective editing strategies for the U. S. Census Bureau foreign trade programs. The Foreign Trade Division publishes monthly estimates on imports and exports for all goods traded between the United States and other countries. The division processes millions of records every month. These records are validated at every step of processing but the data are still subject to errors. Follow-up and recontact of all failing records is not possible due to the large size of the data and the amount of time available. This research proposes using selective editing. In selective editing, a score function is used for automatically identifying failing records that have a significant impact on the final tabulations.

During FY 2005, we developed and implemented three separate score functions. All functions incorporate previous cycle data to predict potential records having a large impact on the publication. We tested the score functions using representative commodities. The initial tests showed difficulties in applying this methodology due to discontinuities in the reporting patterns. We modified the code to use the median or mean value of transactions at any given month, but still could not get acceptable results. We adapted ideas proposed by Statistics Sweden to develop alternative score functions. The new scores use the unit price as the main variable to identify significant errors in either value of trade or quantity of trade. We tested the feasibility of this editing strategy on a selected number of commodities. The tests show that we can obtain negligible pseudo-bias with follow up of 10-40% of the failing records. We presented this research at the Economic Methodology Seminar.

*Staff:* María García (x31703), Yves Thibaudeau, Alison Gajcowski, Sharon Ennis (FTD)

## 1.22 DISCLOSURE AVOIDANCE METHODS
### (Economic Project 2420551)

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.

During FY 2005, staff worked with the Manufacturing and Construction Division (MCD) in reconstructing a noise addition program in SAS that can be applied to any microdata that has a magnitude variable that requires disclosure protection. This program is based on a noise protection method developed by Evans, Zayatz, and Slanta (EZS) around 1998. There is now interest in applying this method to Commodity Flow Survey (CFS) data. The Economic Statistical Methods and Programming Division (ESMPD) tailored this general SAS program (written by Slanta and Massell) to read CFS data and to create tables identical in form to those from the economic census of 2002. ESMPD worked on this to produce tables generated from noisy microdata that could be compared with the actual 2002 tables. If the noisy tables are more attractive to table users than tables that have cell suppressions, the Census Bureau and the sponsoring agency (Bureau of Transportation Statistics of the Department of Transportation) may elect to produce noisy tables in future production runs.

Staff worked closely with ESMPD in fine tuning the implementation of the EZS disclosure method to CFS test data from 2002. Various analyses were performed using SAS programs to see if the implementation works in a way that is consistent with the theory underlying the method. Testing so far indicates the implementation is working well. Two of the major percentages of interest are the percentage of sensitive cells that receive a small perturbation (e.g., less than 7%) and the percentage of non-sensitive cells that receive a large perturbation (e.g., over 12%). These percentages can be viewed as alpha and beta errors for the EZS noise procedure. The size of the coefficient of variation (CV) for a table cell value determines its data quality; cells with large CVs are published with a symbol indicating poor quality rather than a value. Analysis was begun on how CVs change when EZS noise is added. Staff are writing various notes that will comprise the basis for a paper on this work.

Staff worked on an interesting problem motivated by a specific Disclosure Review Board question raised by MCD. It has to do with how one can interpolate an annual estimate into monthly estimates using a consecutive month ratio formed from a monthly survey. The problem involves deciding how to test the sensitivity of the monthly estimates and the associated ratios.

Staff worked with the Bureau of Labor Statistics in preparing some short documents that will accompany an online version of a cell suppression program that will reside on the Confidentiality and Data Access Committee (CDAC) website. The program was developed at the Census Bureau during the early 1990s, and in 1998 the Energy Information Administration created a windows-based version that runs on a personal computer. This latter version will be put online.

Staff fielded questions on cell suppression for cells with establishments but no value of shipments.

Staff will keep current on statistical disclosure control methods for tabular data. We plan to learn more about integer programming, dynamic constraint programming, and the capabilities of the mathematical programming software, CPLEX, of which the Census Bureau currently has three licensed copies.

Staff will work on various aspects of Census Bureau software used for statistical disclosure control for tables. This includes improving capabilities and speed of the programs and their documentation.

*Staff*: Laura Zayatz (x34955), Paul Massell, Phil Steel, Sam Hawala, Jeremy Funk

## 1.23 - TIME SERIES RESEARCH
### (Economic Project 2420552)

### A. Seasonal Adjustment Support

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.

During FY 2005, staff organized and taught a course on time series modeling and seasonal adjustment to personnel from the Economic Directorate. Staff also examined Easter estimation and adjustment of the Retail Sales of Shoe Stores by comparing the current model used by the Service Sector Statistics Division with a model with an alternate Easter regressor that included a separate regressor for a post Good Friday effect, which was found not to be superior to using the standard Easter regressor. Seasonal adjustment and X-12-ARIMA support was provided to: Australian Industry Group, Capital Markets, Santender Company (Brazil), Economics from Washington, Virtua Research, The Conference Board, Fujitsu Services, Traxis Partners, Terroir Inportadere (Brazil), Oracle, Moss Pharmaceuticals (UK), Victory Capital Management, CitiCorp, Hendyplan, Banco Itau (Brazil), Infosys, SAS, General Motors, JPMorgan Chase & Co., eBay, Factset, Petros (Brazil), Refco Commodities, Credit Suisse (Brazil), Banco Galicia, Bureau of Labor Statistics, Fish and Wildlife Service, Department of Transportation, Federal Reserve Board, New Mexico Department of Labor, Nevada Department of Employment and Training, Maryland DHCD, State of Washington Economic and Revenue Forecast Council, New York City Economic Development Corporation, Australian Bureau of Statistics, Census Bureau of Columbia, ESRI (Japan), Statistics Canada, Statistics Netherlands, Statistics Norway, INSEE (France), INDEC (Argentina), Statstiks Austria, Tourism Board of Canada, Provisional Government of Quebec, UK Treasury Department, Office of National Statistics (UK), Government of Bermuda Statistics Office, National Statistics Office (South Korea), Instituto Nacional de Estadística (Spain), Statistics New Zealand, Directorate of Labour (Norway), Ontario Ministry of Finance, Bank of China, Bank of Pakistan, Bank of Spain, UC Santa Barbara, Hebrew University (Israel), Yildiz Technical University (Turkey), University of Connecticut, University of Wisconsin Whitewater Nanking University, Austral University (Argentina), Tehran University, University of Chicago, Harvard University, Nottingham University (UK), and Sungshin Women's University (Korea).

*Staff*: Brian Monsell (x31721), Tucker McElroy, David Findley (M&S)

### B. X-12-ARIMA Development and Evaluation

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) finishing a release version of the program for the general public that includes the automatic time series modeling capability of the TRAMO/SEATS seasonal adjustment program and (2) further improvements to the X-12-ARIMA user interface, output and documentation. In coordination and collaboration with the Time Series Methods Staff (TSMS), 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.

During FY 2005, development of X-12-ARIMA Version 0.3 continued, as staff revised the difference identification procedure within the automatic model identification procedure; added new diagnostics, indications of whether the sliding spans or revisions history analysis failed and other program settings to the diagnostic output at the request of the Economic Statistical Methods and Programming Division (ESMPD); changed graphics codes for the indirect adjustments at the request of ESMPD to make them more consistent and allow more seamless integration into X-12-Graph; and added new options to control aspects of the spectral plots. Linux versions of X-12-ARIMA software was created along with scripts to automate creating future versions, and a web page for distributing versions of X-12-ARIMA for Linux was posted on the Census Bureau Intranet. Staff also created updated versions of X-12-ARIMA code for DEMETRA, as well as updated versions of the version X-12-ARIMA that accesses FAME database.

*Staff:* Brian Monsell (x31721), Richard Gagnon

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

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) collaboration with ESMPD in the further evaluation of the TRAMO/SEATS model-based seasonal adjustment program; 2) the further development of a version of X-12-ARIMA called SEATS, so that X-12-ARIMA diagnostics can be used to analyze SEATS adjustments, and also so that, when appropriate, SEATS adjustments can be produced by the Economic Directorate.

During FY 2005, the program (provisionally named *X-13A-S*) that combines X-12-ARIMA with the leading fully model-based seasonal adjustment program, SEATS (from the Bank of Spain), was revised several times. Staff incorporated several enhancements, including incorporating revisions to the under/over adjustment diagnostic used in the SEATS procedure; incorporating the correct finite sample revision errors for model based seasonal adjustment; updating the signal extraction routines and testing their performance on Census Bureau series; and adding benchmarking routines developed by Statistics Canada as an option to force the totals of the seasonally adjusted series to match those of the original series. In addition, development of the regCMPNT software for estimating component time series models continued by completing testing of a new routine to invert the roots in non-invertible MA polynomials, incorporating additional regressors, and revising the namelist input to provide the same names for namelists and variables as in the input specifications of X-12-ARIMA.

Staff conducted research studies on several topics, including (1) examining the effect of outliers on Easter modeling and adjustments; (2) deriving theoretical results foundational to the use of signal extraction diagnostics for seasonal ARIMA models such as the diagnostics contained in SEATS; (3) investigating two approaches to measuring uncertainty in published seasonal adjustments, including the procedure of Pfeffermann (1994) that is under consideration at BLS; (4) developing procedures for modeling data with seasonal heteroscedasticity and applying them to Census Bureau Construction series; (5) performing real data studies of unit root prevalence in minimum AIC generalized airline models; (6) developing matrix formulas for signal extraction, which are used repeatedly in *X-13A-S*; (7) researching cycle estimation in seasonal time series, towards answering the question of how seasonal adjustment is affected by the presence of a cycle; (8) developing diagnostics for spectral peaks, relevant to judging the quality of a seasonal adjustment.

*Staff:* Tucker McElroy (x33227), Richard Gagnon, Brian Monsell, Donald Martin, Thomas Trimbur, David Findley (M&S), William Bell (M&S)

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

The purpose of this project is to develop supplementary documentation and supplementary programs for X-12-ARIMA that enable both inexperienced seasonal adjustors and experts to use the program as effectively as their backgrounds permit. This fiscal year's goals include collaborating with ESMPD staff to develop a new and improved version of the X-12-ARIMA Reference Manual in WordPerfect with an extensive index and other aids not available in the TeX version.

During FY 2005, staff continued to develop drafts of the X-12-ARIMA Reference Manual for versions 0.2.10, 0.3, and *X-13A-S*, and finished work on an index for each of these documents. Documentation for Ox modules was developed by staff for studying model based signal extraction and diagnostics. Staff also collaborated with personnel from the Economic Statistical Methods and Programming Division and Foreign Trade Division to develop several drafts of a Guideline for Seasonal Adjustment Diagnostics that was distributed to users in the Economic Directorate for comments. Staff also completed research reports on matrix formulas used in computations for model-based seasonal adjustment, and a paper for the Federal Committee on Statistical Methodology meeting on software to generate seasonal adjustment diagnostics for model-based seasonal adjustment.

Staff developed a utility in the Icon programming language to convert the old name list input of REGCMPNT to the recently revised input format, as well as revised a utility for developing user-defined Easter regressors for use with X-12-ARIMA.

*Staff:* Brian Monsell (x31721), Richard Gagnon, Tucker McElroy

## 1.24 POSTAL RATE COMMISSION/STATISTICAL CONSULTING
### (Statistical Research Division Project 8150)

The work associated with this project will entail the review of testimony, interrogatories, decisions, and other documentation relating to proceedings of the Commission in order to identify major statistical issues and provide relevant consultation. The consultation will include: 1) the briefing of the commissioners and other commission officials on the ramifications and desirable approaches to the identified statistical questions; and 2) the presentation of written summaries of the major findings from all assigned reviews.

During FY 2005, staff reviewed documentation, empirical results, and other documents related to a proposed mail forwarding procedure and discussed principal findings with Commission staff. We also participated in a technical conference on the data collection and estimation procedures for the U.S. Postal Service City Carrier Street Time Study; provided technical reviews of the survey documentation and relevant results from the study; and contributed to the framing of questions for a Presiding Officer's Information Request (POIR) relating to the current rate case before the Commission.

*Staff:* Leroy Bailey (x34917)

### 1.25 NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY/BAYESIAN STATISTICAL METHODOLOGY
### (Statistical Research Division Project 8863)

The purpose of this project is to provide technical expertise to the Statistical Engineering Division of the National Institute of Standards and Technology towards setting up Bayesian methods mainly for the project, "3D Chemical Imaging at the Nanoscale Level."

During FY 2005, we provided a short article on Bayesian tomography work at NIST for its "ITL Technical Highlights in Nanotechnology" booklet. Staff presented a half-hour talk at the ICSA 2005 Applied Statistics Symposium, June 12-15, 2005 titled: "Bayesian 3-Dimensional Image Reconstruction from 2-Dimensional Electron Microcopy Transmission Images."

*Staff:* Don Malec (x31718)

### 1.26 PROGRAM DIVISION OVERHEAD
### (Census Bureau Project 0251)

**A. Division Leadership and Support**

This staff provides leadership and support for the overall operation of the division.

*Staff:* Tommy Wright (x31702), Tina Arbogast, Hazel Beaton, Alice Bell, Maria Cantwell, Pat Cantwell, Robert Creecy, Manuel de la Puente, Judi Norvell, Barbara Palumbo, Nita Rasmann, Diana Simmons

**B. Computer Support**

The Computer Support Group provides computer support with the goal of providing a statistical computing environment that provides researchers powerful tools to develop new methods and permits them to share information easily and accurately. Hardware includes SUN servers, workstations, and PCs on a NOVELL network. Because of IT consolidation at the Census Bureau, this group was dissolved near the end of FY 2005.

*Staff:* Chad Russell (x33215), Tom Petkunas, Mohammed Chaudhry

21

# 2. RESEARCH

## *Statistical Methodology*

### A. Disclosure Avoidance

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 limitation at the stage of processing. Methods will be developed, tested, evaluated, and documented. We will also aid in the implementation of the methods.

During FY 2005, staff worked with Duke University and the Office of the Chief Economist on the microdata access system (MAS). This included a data exploration phase which was helpful in thinking about the design for the model server. They wrote up suggestions for Synectics. Staff received the second version of the Synectics screen design for the remote access system and provided comments. Synectics recruited 20 Association of Public Data Users (APDU) members to review and comment on the screen design and other aspects of the remote access system. Staff implemented the synthetic scatter plot routine that Duke University invented. We worked on the core problem which is how to give the user maximum flexibility on recodes without disclosure problems. There is also a usability problem there, in that multiple recodes affect one another.

Staff developed a survey to poll the MAS test group on usage parameters for the remote access model server. Unfortunately, we got very little response, but we worked out a plan. The prototype of the MAS is up. We can view screens and make our analysis choices. The system is writing the SAS code, and we began testing the software. See http://204.52.186.190. The MAS is still having some problems calling SAS over the firewall. The contract has been extended (no cost) to end December 31, 2005. We had a meeting about the MAS with Synectics. It subsequently sent us a revised schedule, and we have a list of priorities for remaining work. The prototype does not have much flexibility yet, but the basic function is operating. Staff had a second meeting with Synectics. We talked mostly about the implementation for dummy variables.

Staff went to the Brookings Institution to discuss the Advanced Query System, the Microdata Access System, the National Center for Health Statistics' remote access system, and other similar systems. They will be providing us with feedback on these systems.

The updated *Statistical Policy Working Paper 22* is finished.

Staff met with a group of international visitors to discuss disclosure avoidance procedures and the Disclosure Review Board (DRB) review process. They also made a presentation on disclosure avoidance for Census Bureau microdata for 20 visitors from Statistics Sweden.

Staff and others met with staff members from Congressman Waxman's office in a (successful) effort to develop a Census 2000 special tabulation that would meet the Congressman's needs and would not lead to disclosure problems.

The DRB approved a new disclosure avoidance technique involving synthetic data for the Longitudinal Employer-Household Dynamics transportation package. Staff reviewed documentation of the technique to see what needs to be removed before it can be released to the public.

Staff served on an ad hoc disclosure review board at the Social Security Administration and reviewed an earnings history and benefits file. The file was approved for release.

We plan to offer the Confidentiality and Data Access Committee/Washington Statistical Society course on disclosure avoidance again.

A staff member continued to serve as chair of the Census Bureau Disclosure Review Board.

*Staff:* Laura Zayatz (x34955), Phil Steel, Paul Massell, Sam Hawala, Jeremy Funk

### B. Disclosure Avoidance for Microdata

Staff investigates methods of microdata masking that preserves analytic properties of public-use microdata and avoid disclosure.

During FY 2005, staff investigated methods for quantifying information loss in masked microdata and additional comparison metrics that can be used in re-identification experiments. Staff reviewed literature on privacy "preserving data mining." Staff completed the report "Modeling and Quality of Masked Microdata." Staff updated the list of references on microdata confidentiality. Staff refereed five papers and reviewed a large number of papers. Staff gave the talk "Microdata Confidentiality Methods" at the Australian Bureau of Statistics in Canberra, ACT, Australia. Staff gave the invited talk "Methods of Re-identifying and Analyzing Masked Microdata" in the Computer Science Department at Cornell University.

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

**C Seasonal Adjustment (See Economic Project 2420552)**

**D. Small Area Estimation-Decennial/Demographic Applications**

A monthly meeting designed as an open forum for small area research and topics throughout the Census Bureau was established. These meetings are designed as a way to disseminate ideas on small area estimation across the group of small area researchers working on a wide variety of projects across the Census Bureau.

During FY 2005, the manuscript, "Small Area Estimation from the American Community Survey using a Hierarchical Logistic Model of Persons and Housing Units" was accepted for publication in the *Journal of Official Statistics*.

The small area discussion group is on hiatus, after all the willing members have presented their projects to the group. The plan is to reconvene infrequently, as the need arises.

*Staff:* Don Malec (x31718), Jerry Maples

**E. Nonresponse in Longitudinal Surveys**

This project requires an extensive examination of relationships between longitudinal survey nonresponse and potential explanatory variables for a variety of survey items. The research objectives are: 1) to apply the results of this investigation in the development of general analytical models which reflect potential survey errors in estimation and analysis ascribed to longitudinal nonresponse, and 2) to identify and advance a well-defined process for selecting and evaluating desirable approaches to nonresponse compensation for longitudinal surveys.

See Highlights on Demographic Project 1465-A, Longitudinal Weighting.

*Staff*: Leroy Bailey (x34917)

**F. Household Survey Design and Estimation**

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.

During FY 2005, staff drafted and revised the chapters 'Design of the Sample and Sample Selection' and 'Weighting and Estimation' for the ACS Design and Methodology document. Also see Highlights for Decennial Project/ACS 5385560.

*Staff*: Lynn Weidman (x34902), Michael Ikeda, Julie Tsay

**G. Sampling and Estimation Methodology: Economic Surveys**

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.

During FY 2005, staff assisted in planning and analyzing results from a simulation to evaluate alternative variance estimators for the Quarterly Financial Report (QFR). The variance simulation was an extension of an earlier simulation that evaluated alternative QFR point estimators. Staff reviewed drafts of "Investigation of Alternative Variance Estimators for the Quarterly Financial Report." This paper recommends replacement of the current QFR variance estimator with a delete-a-group jackknife estimator. Staff also reviewed drafts of "Investigation of Alternative Estimators for the Quarterly Financial Report." This paper recommends retention of the current QFR point estimator. It is about to be released as an SRD Research Report. Staff participation in this QFR estimation research project will be completed with the release of the Research Report. This work was done with the Company Statistics Division and the Economic Statistical Methods and Programming Division.

Another area of research is an investigation of methodology to address outlier observations that are shown (or believed) to be correct but have undue influence on estimates. In contrast, outlier observations caused by incorrect data are addressed by editing strategies currently in place. The beginning step in research is an examination of influential observations in the estimation for the Monthly Retail Trade Survey.

For the initial stage of the research, staff has identified methods in the literature for the identification and treatment of influential observations. Preliminary work with some of these methods has shown that the research team needs additional data and to examine the interrelationship with the current editing methods.

Working with staff from the Foreign Trade Division, we investigated methods to estimate low-valued exports (LVEs) and imports (LVIs). For LVEs, about 80% is never recorded or available. We are developing a procedure that combines data from Canadian exports with available data from exports to other countries to predict the LVEs that are not

available. For LVIs, we are studying the characteristics of a subset of data that are available every month, trying to project the totals upward to reflect those data categories that are not keyed or recorded in other ways.

*Staff:* Mike Ikeda (x31756), Pat Cantwell, Mary Mulry

### H. Research and Development Contracts

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 through FY2007. 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.

The multiple contracts were awarded during FY2002 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.

During FY2005, seventeen new task orders were awarded, thirty-nine modifications were awarded, and nine task orders were completed. To date, there have been fifty task orders awarded under the R&D 2007 contracts, with a monetary value of over $44.3 million (over $38.2 million of the $44.3 million obligated).

*Staff:* Ann Dimler (x34996)

## *Statistical Computing Methodology*

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

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

During FY 2005, staff sent papers, software and information to universities and other groups around the world. Staff gave the WSS Seminar "Data Quality: Automated Edit/Imputation and Record Linkage." Staff completed a version of BigMatch software containing new comparison metrics for Decennial tests and a special version with new metrics for microdata confidentiality re-identification experiments. Staff completed the SRD research report "Evaluating String Comparator Performance for Record Linkage." Staff completed a draft of new methods and software for record linkage error rate that generalizes Belin and Rubin (*JASA* 1995).

*Staff:* William Yancey (x34891), William Winkler

### B. Editing

Under this project, staff will provide advice, develop computer edit/imputation systems in support of demographic and economic projects, implement prototype production systems, and investigate edit/imputation methods.

During FY 2005, we completed research and prepared a paper on SPEER system for editing economic data. We completed a paper describing modeling in terms of data quality issues. We presented these two research papers at the UNECE work session on data editing. We wrote a draft paper on how to determine a set of edits. We tested the methodology described in this paper using Annual Capital Expenditures Survey raw data with no edit failures. We are doing more tests using artificial data generated from clean (corrected) data from the Annual Survey of Manufactures.

*Staff:* Yves Thibaudeau (x31706), Bor-Chung Chen, María García, Bill Winkler

### C. Machine Learning

Under this project, staff will investigate various methods from the artificial intelligence literature such as data mining, Bayesian networks, Hidden Markov Models, and semi-supervised learning from related computer science literature with applications to typical statistical agency problems, such as record linkage, edit/imputation, text classification for industry-and-occupation coding, microdata confidentiality, and sampling.

During FY 2005, staff gave the Washington Statistical Society Seminar "Machine Learning for Text Classification." Staff gave the talk "Machine Learning and Record Linkage" to several departments including the host Computer Science Department at the Australia National University. Staff updated reference lists for microdata confidentiality, statistical data editing, and record linkage that are available at http://www.niss.org/affiliates/totalsurveyerrorworkshop200503/tse_presentations.html.

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

## D. Developed Software Support

### D.1. General Variance Estimation Development and Support

This project will develop new methods and interfaces for general variance estimation software including VPLX, WesVar, and SUDAAN. Staff will provide support for complex applications such as the Survey of Income and Program Participation (SIPP) and the Survey of Construction (SOC), create training materials, and provide training for variance estimation software applications.

During FY 2005, staff continued to offer ongoing Hot-Line support for variance estimation software to the four program directorates at the Census Bureau. Specifically, staff provided variance estimation support to the Population Division (POP). Through the variance estimation support, POP has asked staff to provide VPLX training for its division. This will take place in FY06. Staff directed the research and methods area of Service Sector Statistics Division in how to access SAS-callable SUDAAN on the SRD UNIX. Finally, staff continued learning features in WesVar, SUDAAN, and R with the goal of benchmarking variance estimation software.

*Staff:* Aref Dajani (x31797), Mary Ann Scaggs, Ned Porter, Bob Fay (M&S)

### D.2. SRD Portal Development

This project develops a web-based application to provide user controlled knowledge management. Portals will provide access to a wide range of web-enabled applications, will bring together diverse data sources, and will provide secure access to existing information, regardless of where the information resides. The SRD research portal will ultimately become an integrated starting point for accessing all division images and applications.

During FY 2005, both a developmental portal page and a production portal page were established on a Linux machine where we are using the Oracle Export/Import procedure to transfer material from the developmental to the production page. The developmental page allows the staff to create and test content and applications before they are transferred to the SRD Production Portal (which is the Division's Intranet Web page). The Export/Import procedure is working satisfactorily and portions of the Developmental Page continue to be transferred to the Production Page.

Several members of SRD are now Page Group Administrators for their respective staffs and we continue to encourage additional members to post information on Portal. Our newest addition is the R & D Contracts (Sampling Research) Page Group. This information will also be exported/imported to the SRD Production Page.

Release of various portions of the SRD pages on a flow basis is anticipated and our aim is to release the new version of the Production Page before the end of this year.

*Staff:* Mary Ann Scaggs (x34966), Aref Dajani, Ned Porter

## E. Statistical Computation for Longitudinal Employer-Household Dynamics (LEHD)

The Longitudinal Employer-Household Dynamics project is a cooperative effort among several areas of the Census Bureau to combine Federal and State economic data with demographic data. Sources of data include the American Community Survey, IRS, and Social Security data. Using this data, researchers will now be able to perform analyses that help disentangle the effects of choices that firms make from the choices workers make.

During FY 2005, we completed a preliminary parallel version of the cg2 fixed effects modeling program using the Message Passing Interface (MPI). This version of the program is designed to run on both clusters of small computers and large shared memory computers. This program will be deployed on the 64 processor SGI supercomputer that is being acquired to support research in synthetic data, as well as on the 16 processor SGI computer to be used by the LEHD project. A parallel mixed effects version of the program will also be completed soon because it is a simple modification of the fixed effects program.

Further testing of the parallel version of the cg2 fixed effects modeling program was done, and the program was ported to the newly acquired SGI supercomputer. Assistance was given to Cornell University and LEHD in the specification, acquisition, configuration and industry/occupation testing of the SGI supercomputer, which is just about ready to be turned over to users for synthetic data research.

*Staff:* Rob Creecy (x33207)

## F. Multiple Imputation Feasibility Study

Modern high-speed computer processing makes it possible to identify, retrieve, and represent association patterns among the data available from censuses and surveys. Modern statistical agencies can and should use computing power in concert with state-of-the-art methodologies to compensate for missing data through estimation and imputation in order to improve data quality.

During FY 2005, staff began this effort.

*Staff:* Yves Thibaudeau (x31706), Bor-Chung Chen, Todd Williams

### G.  Optimizing Field Operations

This project is intended to provide the Field Division with a resource for new research in areas that will improve its processes. Over time, research topics may include modeling or forecasting. The first project will study the use of operations research techniques to improve the ability to predict survey costs in the field.

During FY 2005, the current version of simulation model focused on the National Health Interview Survey (NHIS) field operation at the national level with 2004 Q2 frequency distribution by 28 outcomes of 15,238 cases. Staff also obtained the 2004 Q2 NHIS interview length data by outcome. These interview length data are not included in the current version due to the unrealistic data involved. For example, the outcome code of 216 (no one home of Type A) should not have a maximum interview length of 646.6 minutes and mean of 12.4 minutes. Therefore, we assume that the outcome code of 201 (completed interview) has the uniform distribution of interview length with mean 46.2 minutes between 16.2 and 76.2 minutes, the outcome code of 203 (sufficient partial interview, no follow-up) with mean 31.5 minutes between 11.5 and 61.5 minutes, and the other outcomes have no interview time. The 46.2 and 31.5 were the respective medians of the outcomes of 201 and 203. The current version does not include the listing operation either since the operation is done at the beginning of each calendar year. A separate simulation model for listing operation may be needed when staff learns more about it.

Staff is requesting NHIS Contact History (CHI) data and more detailed interview length data by outcome. The CHI data are used to accurately predict the contact time and contact/no contact status, which will be simulated based on the time of a day and the day of a week starting on a Monday and lasting 17 days each month for the NHIS field operation. These structures have been built into the current version of the simulation model. The first "complete" simulation model will be ready to go as long as the requested data are available and the input data modeling is completed.

Staff met with staff from the Demographic Statistical Methods Division to learn more about the 1995-2004 NHIS sample design, Field Division to discuss the FR activities and listing operations in the field, and Demographic Surveys Division to learn more about the CHI data.

*Staff:* Bor-Chung Chen (X34857), Carol Corby

## *Survey Methodology*

### A.  Usability Research and Testing

### A.1.  Remote Usability Testing with Netraker Software on the American Fact Finder Web Site

The purpose of this study is twofold. First, we aim to experiment with the Netraker software and determine if future remote usability studies using the Netraker software will provide useful information for Census Web site development. Second, we intend to generate remote usability results of the integration of the American Community Survey (ACS) data with the American FactFinder (AFF) Web site. Staff will review results to compare and confirm with the in-person usability study results of the integration of ACS into AFF which was conducted earlier in the year.

During FY 2005, staff created the usability study with the Netraker software, and after reviewing the layout and content of the software with the design team, conducted more than 200 remote usability studies. Staff analyzed the information and wrote a report for a presentation at the Usability Professionals' Association annual meeting. Staff concluded that the remote usability testing software did not give valuable information on what the usability problems were or how to go about fixing the usability problems. The study further confirmed that the information gathered from an in-person study was more useful when working on making usable improvements to the site.

*Staff:* Erica Olmsted (x34893)

### A.2.  Usability Evaluation of Linux Desktop

Staff supported the Economic Statistical Methods and Programming Division in a pilot test of Linux desktops by helping to evaluate the usability of the systems.

During FY 2005, we designed questionnaires to evaluate pilot tester satisfaction based on the results of task analysis exercises conducted in FY2004. We conducted contextual inquiry sessions with the testers to learn about the positive and negative aspects of the Linux systems.

*Staff:* Juan Pablo Hourcade (x33690), Tia Harris (ESMPD)

### A.3.  Survey of Income and Program Participation (SIPP) Usability Evaluation

The purpose of this study is to discover usability problems with the Survey of Income and Program Participation (SIPP) domain of census.gov. Our objective is to assess the efficiency, accuracy, and subjective satisfaction of the site, note the usability violations, and make recommendations for how to improve the site.

During FY 2005, staff worked with the SIPP staff to develop user tasks. Staff observed SIPP users in their natural environment and interviewed the users to

get a better understanding of the typical user tasks of the web site.

*Staff:* Eric Raymond (x31601), Erica Olmsted

### A.4. User-Interface Standards and Guidelines

In 2003, the IT Standards and Uniform Products Program recruited an inter-divisional team for the purpose of reviewing and revising two IT standards on Web sites and applications. The major objective is to produce an up-to-date source of user-interface requirements and recommendations for developers of Web-based sites and applications. Our division's role is to serve as a team member in developing and reviewing content for the revised standard.

During FY 2005, as the culmination of a multi-year effort, the revised and integrated standard, IT 15.0.2, *Web Development Requirements and Guidelines* was released for use. Throughout this final year, we reviewed and helped revise draft material for the main text of the standard and the appendices. This project is complete, and the new standard is available on the Census intranet under IT Standards.

*Staff:* Betty Murphy (x34858)

### A.5. Redesign of the Census Bureau's Intranet

The purpose of this project is to reorganize the content of the Census Bureau's Intranet so that similar content is grouped together under functional categories. The current organization by agency structure works against new employees being able to find information they need. Our division's role is to conduct usability testing and provide recommendations on the redesign of the Intranet.

During FY 2005, after developing a plan for usability testing of a new Intranet prototype, we conducted usability testing and provided recommendations and results to the design team. We facilitated discussion sessions with two groups of participants who volunteered to review a pre-launch version of the user interface. We began preparing a presentation based on this work for the Federal Committee on Statistical Methodology (FCSM) conference in November 2005.

*Staff:* Betty Murphy (x34858), Amira Abdalla

### A.6. Desktop Applications Accessibility

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.

Usability of Linux Desktop: Staff is supporting ESMPD in a pilot test of Linux desktops by evaluating the accessibility of the systems.

During FY 2005, we determined how to configure the Linux desktop environment to function with screen-reading software. Staff worked with ESMPD IT staff to install the Gnopernicus screen reader. While the screen reader functioned properly with plain text and the GNOME windows menus, it did not function at all with office applications such as word processing, spreadsheets, and chart making software. This project is complete.

Landview 5: Staff evaluated the Landview 5 mapping application for accessibility. The application provides details about businesses in an area.

During FY 2005, staff consulted with the Geography Division (GEO) to discuss the findings of the report. The report found that Landview 5 had inaccessible content. GEO checked with the partner agencies that created Landview 5 to see if there was a way to make text accessible to the screen reader. We tested two coding methods, but neither was accessible. Staff recommended that a statement about whom to contact for help with inaccessible content be placed on the Landview 5 DVD jewel case and web site. This project is complete.

Econ Data Browser: This application permits users to look up economic reports from the 2002 Economic Census nationally, by state, or by county.

Report look up is done primarily by search trees. During FY 2005, it was found during testing that these search trees were not completely accessible. A report was drafted and provided to the sponsor detailing accessibility violations. This project is complete.

*Staff:* Larry Malakhoff (x33688), Anthony Novak, Eric Raymond

### A.7. Web Applications Accessibility

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

American FactFinder: This project focused on the accessibility of large statistical tables.

During FY 2005, staff used the newest version of the JAWS 5.1 screen reader software to determine if accessibility of geographic comparison tables improved. Staff informed the sponsor the screen reader could access text from larger tables than in earlier (before 5.0) versions.

Longitudinal Employer (Household) Dynamics (LED/LEHD) Web Site: This project evaluated the accessibility of browsing household data.

Accessibility testing was done on the old and new LED/LEHD web sites from DSD. In some cases it was determined that there was no link to download Adobe Reader if PDF documents were present on a screen. The evaluation performed on the Quarterly Workforce Indicators (QWI) query section of the LED/LEHD web

site revealed the search box and help icons were unlabeled, drop down menus did not perform as expected, and creation of detailed comparison reports was inaccessible. A report was completed and provided to the sponsor. This project is complete.

Decennial Applicant Person Payroll System (DAPPS): This project evaluated the accessibility of web forms which allow persons in the field to enter their own time and attendance information.

During FY 2005, more than 180 forms were tested. The main findings were that the spoken order did not match the visual order, many data entry fields were unlabeled, data entry labels did not match the visual presentation, and the calendar date function was inaccessible. This project is complete.

Share-A-Ride: This project evaluated both the usability and accessibility of the Share-A-Ride application.

During FY 2005, a usability test plan was created in collaboration with the Administrative and Customer Services Division (ACSD) for the Share-A-Ride car and van pooling application for employees at the Suitland Federal Center. Staff conducted usability testing on the Share-A-Ride system with 14 test subjects. These subjects performed tasks a driver or rider might do. We found that when drivers were searching for riders and riders were searching for car pools, the search results gave them information on areas other than what they were seeking. Participants had difficulty with screens that served a dual purpose, that is, registration and updating commuter information. Test subjects also had difficulty seeing information "below the fold" on scrolling screens. A preliminary report of these findings, video clip highlights, and marked up screen shots were provided to the sponsor. ACSD programmers implemented many of the changes recommended. Staff reviewed these changes and checked them for compliance with Section 508 regulations and sent a report of accessibility findings for their action. The Share-A-Ride application became available to Census personnel in July. The project is complete.

X-12 Installer Program: This application permits users to install the X-12 ARIMA program that queries databases and generates reports.

During FY 2005, accessibility testing revealed that the installer program was compliant with Section 508 laws, but the X-12 Arima program did not have a logical tabbing order. Testing also revealed a problem with the order of input parameters, which the programmer is currently correcting. A report about these findings was provided to the programmer.

SRD Portal Web Site: The SRD web site allows users to browse and search for reports and obtain information about expertise of different staffs within the division.

During FY 2005, the initial accessibility review revealed that search fields were not accessible, and repetitive groups of links, such as navigation bars, needed additional programming for screen reader users to pass by them. A report was provided to the sponsor and this project will continue until all issues are resolved.

E-Learning IT Security Awareness Training: This application provides information about IT security awareness for Census Bureau employees and contractors, and tests their knowledge with a quiz. It is mandatory, and all personnel must take the training annually.

During FY 2005, it was determined that the application did not have a logical tabbing order which made it unusable and therefore inaccessible to screen reader users. As an alternative, the application provided a 'text only' option for screen reader users which was accessible. However, about 60 people who probably were sighted used this option to complete the form, which meant that their registration needed to be entered manually. This consequence indicates that the interface was difficult to use and staff reviewed the content and provided a report to the sponsor with suggested wording changes and improvements on the interface.

During the fourth quarter, 16 test participants were recruited from Census HQ personnel through a broadcast announcement. Two versions of the instrument were tested, a CD-ROM version used by contractors, and an Internet version used by Census personnel. All sessions were audio/video taped and work began to log these tapes. This work will continue into next quarter and a report will be drafted for the sponsor.

Public Use Presentation Library: This project evaluated the accessibility of browsing for PowerPoint presentations done by Census personnel.

During FY 2005, it was determined that the application did not provide an accessible alternative to the Powerpoint presentations posted on this web site. A PowerPoint plug-in was available from Microsoft, but any embedded images, such as screen shots, were not accessible. Screen shots of title screens were not tagged with alt text, making that content inaccessible. It was recommended that captions be written for the embedded images within the PowerPoint presentations and that they be exported to rich text format. A report was provided to the sponsor and is under peer review. This project is complete.

2005 Internet Pre-Test Team: This project is in support of the Internet Census short form pre-test for the 2010 Census.

During FY 2005, staff was consulted about a time-out message for users to request more time in order to comply with Section 508 regulations. Accessibility

testing was coordinated with usability testing on the web forms to be used in the 2005 National Census Test. There were two versions of this form, person-based and topic-based. The Decennial Systems and Contracts Management Office (DSCMO) tested and reported on accessibility of the person based form. Altogether, there were four rounds of accessibility testing. Staff also observed and reported on usability testing of one blind user and one low-vision user. Testing revealed accessibility issues with the Race, sex/age/dob, and ancestry screens. These problems were reported to the sponsor and later resolved by the contractor, Z-Tech. The web form for the 2005 pre-test became live in September. This project is complete.

Harvester Survey Web Site: This project focused on the accessibility of the Harvester survey.

During FY 2005, staff met with Governments Division (GOVS) to discuss the accessibility evaluation. The web survey questionnaire was structured as a stem followed by multiple leaves. While this structure saves space on the printed page, it can cause the screen reader user to lose the context of the question because they cannot glance back easily to the stem of the question. Staff discussed these findings with programmers who would be making the changes. GOVS stated that it would need to consult with the survey sponsor, the Department of Justice, about making the web survey Section 508 compliant. Staff will follow up with GOVS to retest any modifications for compliance. This project is complete.

*Staff:* Larry Malakhoff (x33688)

### A.8. Usability Testing of American FactFinder Economic Data and Map Data

The purpose of this study is to discover usability problems with the Economic data and Mapping data within the American FactFinder Web site. Our objective is to assess the efficiency, accuracy, and subjective satisfaction of the site, note the usability violations, and make recommendations for how to improve the site.

During FY 2005, staff identified the typical users and tasks of the economic and map areas of the AFF web site. Staff created meaningful tasks and developed usability goals. Staff observed typical users using the web site and took accuracy and efficiency measures and subjective satisfaction ratings of the site. Staff analyzed the results and compiled a report for the client that included user performance measures and a list of usability problems with recommendations for solutions.

*Staff:* Erica Olmsted (x34983)

### A.9. Support for New Statistical Abstract Website

Staff is supporting the design of a new website for the *Statistical Abstract of the United States* as it moves from a giving access through a PDF version of its printed publication to a document-based portal.

During FY 2005, we consulted with *Statistical Abstract* staff on information architecture issues and on technologies that could be used to provide optimal access to the documents in the *Statistical Abstract*. We worked on contacting researchers at the University of California, Berkeley to pilot *Flamenco*, one of their technologies, for possible use in the website.

*Staff:* Juan Pablo Hourcade (x33690), Lars Johanson (ACSD), Pam Reichers (ACSD)

### A.10. Statistical Abstract Tables on www.census.gov

This project evaluated the accessibility of tables showing USA and state statistics about population, health, housing, education, welfare, law enforcement, employment, and the economy.

During FY 2005, the accessibility evaluation revealed instructions for usage of PDF and Excel documents were missing, tables lacked the units of the data value, nested row stubs were not read in the correct order, and footnotes were being read as data or regular text. A report was written and delivered to the sponsor. This project is complete.

*Staff:* Larry Malakhoff

### A.11. Census Intranet Web Site

This project evaluated the accessibility to read announcements and to access project management software tools.

During FY 2005, the accessibility evaluation revealed unexpected tabbing order, graphics tagged with redundant text, untagged graphics, no skip links to bypass the top and bottom navigation bars, and lack of a link to download Adobe Acrobat on two screens. A report was written and delivered to the sponsor. This project is complete.

*Staff:* Larry Malakhoff

### A.12. Section 508 Implementation Team

This inter-divisional team's purpose is to raise awareness of the Section 508 laws through on-line training and dissemination of information.

During FY 2005, the team began meeting and started reviewing documents from other agencies to use as policy within the Census Bureau. We agreed that the policy written for the USDA provided the best fit for the Census Bureau. We will be adapting the USDA document for our use. Additionally, we determined that two types of on-line training would be provided, one

general for all employees, and a second course for programmers. Meetings will continue bi-weekly into FY 2006.

*Staff:* Larry Malakhoff

## B. Questionnaire Pretesting

This project involves coordinating the Census Bureau's 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.

During FY 2005, forty letters describing activities conducted under the generic clearance for pretesting research were submitted to OMB. These activities involved 2,017 respondent burden hours.

Staff submitted an annual report to OMB summarizing the research that was conducted under the generic clearance between September 2003 and August 2004.

*Staff:* Terry DeMaio (x34894), Ashley Landreth

## C.1. Questionnaire Design Experimental Research Surveys 2002/2003 (QDERS)

QDERS 2002 is an omnibus survey designed to facilitate independent research related to questionnaire design issues, interviewer training, and other survey methodological issues. The QDERS 2002 was conducted from the Tucson Telephone Center in June/July 2002 using a Random Digit Dialing (RDD) sample. Three researchers conducting questionnaire design and survey methods experiments are participating.

During FY 2005, data analysis of the experiments contained in QDERS 2003 continued to be conducted and summarized in conference papers and project reports.

*Staff:* Jennifer Rothgeb (x24896), Joanne Pascale, Ashley Landreth, Terry DeMaio

## C.2. Questionnaire Design Experimental Research Survey - 2004 (QDERS)

QDERS 2004 is an omnibus survey designed to facilitate independent research related to questionnaire design issues and other survey methodology issues. The QDERS 2004 will be conducted from the Census Bureau's Telephone Center using an RDD sample. Researchers conducting questionnaire design and survey methods experiments are participating.

During FY 2005, the results of one of the experiments (testing new informed consent questions), completed by the Demographic Surveys Division was presented at the AAPOR conference in May. No other analysis has been conducted yet, due to higher priority work.

*Staff:* Jennifer Rothgeb (x34986), Joanne Pascale, Jenny Hunter, Nancy Bates (DSD)

## C.3. Questionnaire Design Experimental Research Survey 2006 (QDERS)

QDERS 2006 is an omnibus survey designed to facilitate independent research related to questionnaire design issues and other survey methodology issues. The QDERS 2006 will be conducted from one of the Census Bureau's Telephone Centers, using an RDD sample.

During FY 2005, a request for proposals was distributed to SRD, Demographic Surveys Division, Population Division, Housing and Household Economic Statistics Division, and Demographic Statistical Methods Division staff. Two experiments (one related to questionnaire design and one related to interviewer training) are planned for QDERS 2006. A contract with RTI has been developed to enable BLAISE authoring of the 2006 QDERS CATI instrument.

*Staff:* Jennifer Rothgeb (x34968)

## C.4. Evaluating Pretesting Techniques for Finding and Fixing Questionnaire Problems

The objective of this research is to determine how well laboratory question testing methods predict the types of problems that will actually be experienced in the field, and to what extent the laboratory testing contributes to improved questions. This project includes research to determine not only the relative effectiveness of different methods for detecting questionnaire problems, but will also evaluate the methods in terms of their ability to provide information on question wording repairs.

During FY 2005, we discussed potential future research in this area with external researchers, however no new research was started due to higher priority work.

*Staff:* Jennifer Rothgeb (x34986), Gordon Willis (National Cancer Institute), Barbara Forsyth (Westat)

## D. Refusal Aversion Training

As unit and item non-response rates for household surveys continue to increase, survey researchers are now focusing on the development of innovative methods to improve these rates. One area of interest is the interviewer-respondent interaction and its influence on survey participation and response quality.

During FY 2005, staff continued to provide technical support for additional surveys and evaluations of Census Response Achievement Field Training (CRAFT), disseminating research findings to the research community, and advising other organizations on their testing and adoption of CRAFT. In the coming year they will collaborate with the Demographic

Surveys Division and the Field Division in evaluating current implementations of CRAFT in CE, CPS, SIPP, NHIS and ACS: how often should it be conducted, what are the gaps in the content, considerations for different modes of data collection and so on. For example, given the unique features of telephone openings between interviewers and respondents, staff will be incorporating findings from the QDERS research to develop an expanded version of CRAFT by mode, particularly for the telephone data centers. Research will begin by studying with the hardest cases, i.e. cold-call situations in RDD sample surveys. In future years, research will advance to studying the effect on interactions with sample units with complex pre-contact histories (advanced letters, paper questionnaires, reminder cards, multiple respondents, complex households, etc.).

*Staff:* Eileen O'Brien (x32695)

### E. Research on Conversational Norms and Response Behavior

Questionnaires are designed to serve as a "script" for a conversational interaction; as such, they often violate conversational norms. When scripts ignore norms, interviewers are forced to choose between initiating an awkward and unsatisfactory interaction on the one hand, or ignoring the "read-every-question-exactly-as-worded" mantra, on the other. Constructing questionnaires that adhere to conversational norms, and training interviewers to detect and avoid violations of norms, should lead to greater satisfaction with the interview experience between both parties, and a more positive regard for the value of the survey. Greater satisfaction and a more positive assessment of the value should lead, in turn, to reduced non-response and panel attrition.

During FY 2005, staff gave two presentations on this topic: one as part of a closed panel workshop at the Centers for Disease Control and Prevention in Atlanta in June, and another at the annual Respondent Cooperation Workshop of the Council for Marketing and Opinion Research in September in Washington, D.C. Both talks related to ongoing Census Bureau research on the nature of initial interactions between interviewers and respondents and their effect on nonresponse and response quality. The first talk included a review of research and the literature on what the appropriate message and behavior might be for encouraging cooperation in initial contacts, non-response follow-up and refusal conversion in surveys on sensitive topics. CDC's chief concern is how interviewers should conduct these operations in their RDD telephone surveys on interpersonal violence with at-risk populations. The second talk related specifically to current and future Census research

investigating the nature and purpose of telephone openings in demographic surveys, considering social and conversational norms in such interactions.

Staff drafted an experiment to be applied to QDERS 2006 that will investigate the influence of paralinguistic cues and survey introductions on participation decisions and response quality. Staff is also finalizing a review of the literature on respondent-interviewer interactions, mode, and their influence on survey participation decisions.

*Staff:* Eileen O'Brien (x32695), Yuling Pan, Jeff Moore

### F. Language: Continuities in Research on the Foreign-born Population and Immigration Issues

This research effort calls for the application of varied qualitative research methods to solving problems that will improve the Census Bureau's ability to collect and provide timely, relevant, and quality data about the social and economic circumstances of the foreign-born population in the United States. It will conduct collaborative research with anthropologists, sociolinguists, and other behavioral scientists on foreign-born persons from different national origins to assess the barriers they experience to participation in the census and demographic surveys.

During FY 2005, we summarized findings from Chinese cognitive interviews using the Census 2000 English and Chinese form and issued an SRD research report. Results from cognitive interviews indicate that because of lack of the social context of survey practice in their home culture and lack of familiarity with filling out a questionnaire, recent Chinese immigrants exhibit most difficulties in form navigation and in interpretation of census key concepts. This study has also identified issues in the Chinese translation of the census form and language obstacles in the choice of writing system for Chinese translation. From this study, we helped the Decennial Management Division plan Chinese translation in both simplified Chinese and traditional Chinese characters.

*Staff:* Yuling Pan (x34950), Eleanor Gerber, Kristen Hughes, Suzanne Scollon (Georgetown University), Barbara Craig (Georgetown University)

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

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 non-English and English language data collection instruments.

During FY2005, we issued an SRD report documenting the development of the Census Bureau Translation Guideline. We continued to conduct research necessary for the development of a Census Bureau guideline for interpretation. Staff participates in monthly conference calls conducted in order to continue ongoing research, including an upcoming volume to be published by Wiley Publications in 2008. Staff completed literature review on interpretation theories and training models and completed a web search for best practices of using interpreters in survey interviews by survey organizations around the world. A division postdoctoral researcher developed a proposal to examine the extent to which cognitive interview methods are useful and applicable to Spanish monolingual respondents. Staff started recruiting Spanish-speaking respondents and conducting cognitive interviews in Spanish. Staff also developed a research proposal in collaboration with colleagues at Westat to conduct research on survey translation methods and approaches.

*Staff:* Manuel de la Puente (x34997), Yuling Pan, Patti Goerman

## H. Training for Cognitive Interviewing

Staff will train members of other divisions in the Census Bureau to carry out cognitive interviewing and provide consultation and support for project which they carry out.

During FY 2005, two training sessions were held, in which 12 cognitive interviewers were trained. These interviewers were from Decennial Management Division, National Processing Center, Population Division, Demographic Surveys Division, and the Statistical Research Division. They worked on such diverse projects as evaluating Hispanic origin, race, and modified ancestry questions for the decennial census, decennial residence rules and coverage questions, age and date of birth questions, and the National Survey of College Graduates.

*Staff:* Eleanor Gerber (x34890)

## I. Analyzing the Data from Cognitive Interviews

This independent research project focuses on the process by which cognitive interviews are conducted and analyzed by survey research organizations. The objective is to conduct an experiment to evaluate different types of interviewing and analytic procedures in terms of the substantive conclusions that are drawn, and the recommendations for questionnaire revisions that they produce. This will facilitate a comparison of the analysis procedures used by the Census Bureau

versus those used by other organizations.

During FY 2005, no work was conducted on this project.

*Staff:* Terry DeMaio (x34894), Ashley Landreth

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

As part of a postdoctoral research fellowship, staff is currently engaged in a study designed to test and identify best practices for conducting cognitive interviews with Spanish-speaking respondents. We are testing both widely accepted and new techniques and probes with both Spanish and English-speaking respondents of high and low educational levels. The research is 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.

During FY 2005, staff presented a research plan and received feedback from leading multi-cultural researchers at the 3rd International Workshop of Comparative Survey Design and Implementation in Madrid, Spain in March of 2005. Staff is currently in the process of conducting interviews and analyzing results. Staff will present preliminary results of this research at the Federal Committee on Statistical Methodology (FCSM) Research Conference in November of 2005.

*Staff:* Patti Goerman (x31819)

## K. Measurement Bias Research

Responses to census and survey questions may partly be a function of sub-population group membership rather than a direct function of respondents' true values on the questions of interest. The purpose of this research is to employ modern and classical psychometric models and theory to address this form of measurement bias.

During FY 2005, staff conducted research to examine measurement properties of items selected from the American Community Survey (ACS) across individuals living in linguistically isolated (LI) households, and households that are not LI. Results demonstrated uniformity in the internal validity of data collected across LI individuals for these items, and dispute concerns that LI substantially affects the ability of LI individuals to answer these items on the ACS in a meaningful way as compared to non-LI individuals. Similarly, staff also examined the extent to which items in a National Institute on Alcohol Abuse and Alcoholism (NIAAA) sponsored survey demonstrate equivalent functioning across the race and ethnic groups identified in the survey and across the language of interviews. Results demonstrated the presence of

differential item functioning across the identified groups on measures of alcohol abuse, alcohol dependence, and depression, and raise concerns regarding cross group comparisons using these items. Staff also collaborated with POP and HHES staff on a research project examining alternative measures of poverty. Major efforts outside the Census Bureau included consultation with National Center for Health Statistics (NCHS) staff to assess the psychometric properties of a child well being scale included on a recent NCHS survey.

*Staff:* Adam C. Carle (x31836)

### L. Q-BANK Development and Research

Q-Bank is an online interagency database management system initiated by the National Center for Health Statistics and designed to store results of pretested survey questions, particularly the results of cognitive research. Q-Bank will be used to search through previously tested questions when survey methodologists and subject-matter specialists develop new questions. To this end, crafting new survey questions can be informed by work that has already been done. Additionally, Q-Bank will allow the production of meta-data about pretesting findings (e.g., what problems are most often identified through pretesting, whether this differs by agency). It will also facilitate analysis of specific types of question characteristics that may contribute to response error. Q-Bank is operated by a steering committee formed of representatives from various federal statistical agencies (members of the steering committee represent the National Center for Health Statistics, the U.S. Census Bureau, the Bureau of Labor Statistics, the National Science Foundation, and the National Cancer Institute).

In FY 2005, staff from the Census Bureau and other member agencies made great progress in finalizing the data fields to be used in Q-Bank. Additionally we expanded the focus of the system to include not only interviewer-administered household surveys, but also self-administered and establishment surveys. This progress was documented in a series of papers presented at an invited session at the Joint Statistical Meetings in August. In June 2005, a first draft of the system was made available online.

*Staff:* Jenny Hunter (x34927), Jennifer Rothgeb

## *Research Assistance*

This staff provides research assistance, technical assistance, and secretarial support for the various research efforts.

*Staff:* Tina Arbogast, Maria Cantwell, Judi Norvell, Gloria Prout, Lorraine Randall, Nita Rasmann

# 3. PUBLICATIONS

## 3.1 JOURNAL ARTICLES, PUBLICATIONS

Aston, J.A.D. and Martin, D.E.K. (In Press). "Waiting Time Distribution of Competing Patterns in Higher Order Markovian Sequences," *Journal of Applied Probability*.

Druin, A. and Hourcade, J.P. (Eds.) (2005). Interaction Design and Children. *Communications of the ACM, 48(1)*.

Findley, D.F., McElroy, T.S., and Wills, K.C. (In Press). "Modification of SEATS' Diagnostic for Detecting Over-and Underestimation of Seasonal Adjustment Components," *Journal of Business and Economic Statistics.*

Findley, D.F. and Martin, D.E.K. (In Press). "Frequency Domain Analysis of SEATS and X-12-ARIMA Seasonal Adjustment Filters for Short and Moderate-Length Time Series," *Journal of Official Statistics.*

Hourcade, J.P., Bederson, B.B., Druin, A., and Guimbretiere, F. (2004). "Differences in Pointing Task Performance Between Preschool Children and Adults Using Mice," *ACM Transactions on Computer-Human Interaction, 11(4),* 357-386.

Hourcade, J.P., Bederson, B.B., and Druin, A. (2004). "Building KidPad: An Application for Children's Collaborative Storytelling," *Software Practice and Experience, 34,* 895-914.

Hunter, J. and Landreth, A. (In Press). "Analyzing Interviewer/Respondent Interactions while Using an MCAPI Device," *Field Methods.*

Lazar, J., Murphy, E.D., and O'Connell, T.A. (2004). "Building University-government Collaborations: A Model for Students to Experience Usability Engineering in the Federal Workplace," *Journal of Informatics Education Research 6,* 57-77.

Malec, D. (In Press). "Small Area Estimation from the American Community Survey using a Hierarchical Logistic Model of Persons and Housing Units," *Journal of Official Statistics.*

Redline, C.D., Dillman, D.A., Carley-Baxter, L., and Creecy, R.H. (2005). "Factors that Influence Reading and Comprehension of Branching Instructions in Self-administered Questionnaires," *Allgemeines Statistisches Archiv, 89,* 21-38, Physica-Verlag.

Slud, E. and Maiti, T. (In Press). "MSE estimation in transformed Fay-Herriot model," *Journal of Royal Statistical Society, Series B*.

Trimbur, T. (In Press). "Bayesian Trend Estimation," *International Journal of Forecasting.*

Trimbur, T. (2005). "Properties of Higher Order Stochastic Cycles," *Journal of Time Series Analysis.*

Hughes, K. (2005). "Respondent Debriefing," in Best, S. and Radcliff, B. (Eds), *Polling America: An Encyclopedia of Public Opinion*, Westport, CT: Greenwood Press, 683-686.

Hunter, J. (2005). "Measuring Cohabitation," in Best, S. and Radcliff, B. (Eds), *Polling America: An Encyclopedia of Public Opinion*, Westport, CT: Greenwood Press, 98-102.

Martin, D.E.K. (In Press). "Distribution of the Number of Successes in Success Runs of Length at Least in Higher-order Markovian Sequences," *Methodology and Computing in Applied Probability.*

Schneider, S, Cantor, D., Malakhoff, L., Arieira, C., Segel, P., Nguyen, K.L., and Guarino Tancreto, J. (2005). "Telephone, Internet, and Paper Data Collection Modes for the Census 2000 Short Form," *Journal of Official Statistics, Vol. 21, No. 1,* 89-101.

McElroy, T. and Sutcliffe, A. (In Press). "An Iterated Parametric Approach to Nonstationary Signal Extraction," *Computational Statistics and Data Analysis, Special Issue on Signal Extraction.*

McElroy, T. and Politis, D. (In Press). "Moment-Based Tail Index Estimation," *Journal of Statistical Planning and Inference.*

McElroy, T. and Politis, D. (In Press). "Asymptotics for Stable Marked Point Processes," *The Annals of Statistics.*

## 3.2 BOOKS/BOOK CHAPTERS

Blumberg, R.L. and Schwede, L. (In Press). "The Recent Arrivals: Latinos and Koreans," in *Complex Ethnic Households in America.* Lanham, MD: Rowman & Littlefield.

Hourcade, J.P. (In Press). "Design for Children," in G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed.). New York: Wiley.

Hourcade, J.P., Bederson, B.B, and Druin, A. (2004). "Preschool Children's Use of Mouse Buttons," *Extended Abstracts of Human Factors in Computing Systems* (CHI 2004), ACM Press: 1411-1412.

Hourcade, J.P. and Fox, J.E. (2005). "Designing Public Government Websites," *Extended Abstracts of Human Factors in Computing Systems (CHI 2005)*. ACM Press: 2039-2040.

Hunter, J. (In Press). "Not the Typical Household: Whites in Rural New York," in *Complex Ethnic Households in America*. Lanham, MD: Rowman & Littlefield.

Kim, S.Y., Preethy, G., Carle, A.C., and Srinivasan, S. (In Press). "Factorial Invariance of the Family Adaptability and Cohesion Evaluation Scale (FACES) Across Asian Americans and European Americans," in F. Columbus (Ed.), *Contemporary Parenting Issues*, Hauppauge, NY: Nova Science.

McElroy, T.S. (2005). *A to Z of Mathematicians,* New York: Facts on File.

Schwede, L., Blumberg, R.L., and Chan, A.Y. (In Press). *Complex Ethnic Households in America,* Lanham, MD: Rowman & Littlefield.

Schwede, L. (In Press). "'Who Lives Here?': Complex Households in America," in *Complex Ethnic Households in America,* Lanham, MD: Rowman & Littlefield.

Schwede, L. and Blumberg, R.L. (In Press). "The First Arrivals: Navajo and Inupiaq Eskimo," in *Complex Ethnic Households in America,* Lanham, MD: Rowman & Littlefield.

Schwede, L. and Blumberg, R.L. (In Press). "The African Americans and Whites," in *Complex Ethnic Households in America,* Lanham, MD: Rowman & Littlefield.

Winkler, W.E. (In Press), "Data Quality in Data Warehouses," in J. Wang, Ed. *Encyclopedia on Data Warehousing and Mining.*

Wright, T. and Cobb, G. (2005). "Counting and Apportionment: Foundations of America's Democracy," in *Statistics - A Guide to the Unknown (4th Edition)* (Editors: R. Peck, G. Casella, G. Cobb, R. Hoerl, D. Nolan, R. Starbuck, and H. Stern), Belmont, CA: Duxbury, 35-67.

## 3.3 PROCEEDINGS PAPERS

*Proceedings of the 2004 NBER/NSF Time Series,* Southern Methodist University, Dallas, TX, September 17-18, 2004.
- Aston, J.A., Findley, D.F., Wills, K.C., and Martin, D.E.K. (2004). "Generalizations of the Box-Jenkins' Airline Model with Frequency-Specific Seasonal Coefficients and a Generalization of Akaike's MAIC."

*2004 American Association for Public Opinion Research Meeting,* Phoenix, AZ, May 13-16, 2004.
*2004 Proceedings of the American Statistical Association.*
- Anderson, A.E., Murphy, E.D., Nichols, E.M., Sigman, R.S., and Willimack, D.K. (2004). "Designing Edits for Electronic Economic Surveys and Censuses: Issues and Guidelines," 4912-4919.
- Landreth, A. (2004). "Survey Letters: A Respondent's Perspective," 4810-4817.
- Moore, J.C. (2004). "Exploiting Computer Automation to Improve the Interview Process and Increase Survey Cooperation," 4830-4837.
- Pan, Y. (2004). "Cognitive Interviews in Languages Other than English: Methodological and Research Issues," 4859-4865.
- Pascale, J. (2004). "Medicaid and Medicare Reporting in Surveys: An Experiment on Order Effects and Program Definitions," 4976-4983.
- Schwede, L. (2004). "Household Types and Relationships in Six Race/Ethnic Groups: Conceptual and Methodological Issues," 4991-4998.

*2004 Joint Statistical Meetings (American Statistical Association),* Toronto, Canada, August 8-12, 2004.
*2004 Proceedings of the American Statistical Association.*
- Bell, W.R. and Martin, D.E.K. (2004). "Modeling Time-Varying Trading-Day Effects in Monthly Time Series," 1045-1052.
- Cantwell, P. (2004). "Discussion of Session: Wholesale and Retail Trade Surveys-Accomplishments and Challenges," 3333-3335.
- Chen, B.-C. and Winkler, W.E. (2004). "The Cutting Plane Algorithm in the Error Localization Problem," 2411-1418.
- Creecy, R. (2004). "Combining Parameter Expansion and Overrelaxation to Speed up the EM Algorithm," 2432-2437.
- Feldpausch, R.M., Hood, C.H., and Wills, K.C. (2004). "Diagnostics for Model-Based Seasonal Adjustment," 1180-1186.
- Findley, D.F., McElroy, T.S., and Wills, K.C. (2004). "Modification of SEATS' Diagnostic for Detecting Over- or Under-estimation of Seasonal Adjustment Decomposition Components," 1187-1194.
- García, M.M. (2004). "Implicit Linear Inequality Edits and Error Localization in the SPEER Edit System," 3539-3546.

- Huang, E.T. and Bell, W.R. (2004). "An Empirical Study on Using ACS Supplementary Data in SAIPE State Poverty Models," 3677-3684.
- Maples, J. and Bell, W.R. (2004). "Investigating the Use of IRS Tax Data in the SAIPE School District Poverty Estimates," 1656-1663.
- Massell, P.B. (2004). "Comparing Statistical Disclosure Control Methods for Tables: Identifying the Key Factors," 1664-1671.
- McElroy, T. (2004). "Iterative Signal Extraction," 1271-1278.
- Mulry, M.H. (2004). "Methodological Lessons from Census 2000 Coverage Error Measurement," 4066-4071.
- Slud, E.V. (2004). "Small Area Estimation Errors in SAIPE Using GLM Versus FH Models," 4402-4409.
- Steel, P.M. (2004). "Disclosure Risk Assessment for Microdata," 1766-1773.
- Stratton, J., Finamore, J., and Williams, T. (2004). "Comparison of Two Imputation Methods in the Survey of Doctorate Recipients," 4446-4453.
- Thibaudeau, Y. and Winkler, W.E. (2004). "Releasing Disclosure-Limited Count Information Through Full-Rank Minimal Sufficient Statistics," 4470-4476.
- Trimbur, T.M. (2004). "Statistical Modeling of Stochastic Level Shifts in Time Series," 1354-1361.
- Winkler, W.E. (2004). "Approximate String Comparator Search Strategies for Very Large Administrative Lists," 4595-4602.
- Yancey, W.E. (2004). "BigMatch: A Program for Large-Scale Record Linkage," 4652-4655.

*Proceedings of the 2005 Federal Forecasters Conference,* Washington, D.C., April 21, 2004.
- McElroy, T. and Bell, W. (2004) "Forecasting Age Distribution Curves."

*Proceedings of the Association for Survey Computing Conference on Mobile Computing*, April 22, 2005.
- Hunter, J. and Landreth, A. (2005). "Using Behavior-coding to Analyze Interviewer/Respondent Interactions with a Mobile Computing Device." Proceedings available online at http://www.asc.org.uk/SC/index.htm.

*Proceedings of the 2005 Human-Computer Interaction International Conference*, Las Vegas, NV, July 22-27, 2005.
- Murphy, E.D. "Reducing User Frustration in Designing Data-collection Forms for Internet Users: Can We Improve Performance and Satisfaction?"

*Proceedings of Universal Access in Human-Computer Interaction* (UAHCI), Las Vegas, Nevada, July 22-27, 2005.
- Murphy, E.D. "Steps Toward Integrating Accessibility Into Development of an Internet Option for the 2010 U. S. Census."


## 3.4   STATISTICAL RESEARCH DIVISION RESEARCH REPORTS

**RR (Statistics #2004-06)**, William E. Winkler, "Masking and Re-identification Methods for Public-uses Microdata: Overview and Research Problems," October 21, 2004.

**RR (Survey Methodology #2005-01)**, Jeffrey Moore, "An Evaluation of SIPP Wave 2004 Wave 1 Asset Ownership Reports," March 7, 2005.

**RR (Statistics #2005-01)**, Patrick Cantwell, Howard Hogan, and Kathleen M. Styles, "Imputation, Apportionment, and Statistical Methods in the U.S. Census: Issues Surrounding Utah vs. Evans," March 8, 2005.

**RR (Survey Methodology #2005-02)**, Jennifer Rothgeb, Gordon Willis, and Barbara Forsyth, "Questionnaire Pretesting Methods: Do Different Techniques and Different Organizations Produce Similar Results?" March 21, 2005.

**RR (Statistics #2005-02),** William E. Winkler, "Approximate String Comparator Search Strategies for Very Large Administrative Lists," March 21, 2005.

**RR (Survey Methodology #2005-03),** Elizabeth Nichols, Elizabeth Murphy, Amy Anderson, Diane Willimack, and Richard Sigman, "Designing Interactive Edits for U.S. Electronic Economic Surveys and Censuses: Issues and Guidelines," March 23, 2005.

**RR (Statistics #2005-03)**, Leroy Bailey, "Weighting Alternatives to Compensate for Longitudinal Nonresponse in the Survey of Income and Program Participation," May 3, 2005.

**RR (Statistics #2005-04)**, Tucker McElroy, "Matrix Formulas for Nonstationary Signal Extraction," June 9, 2005.

**RR (Statistics #2005-05)**, William Yancey, "Evaluating String Comparator Performance for Record Linkage," June 13, 2005.

**RR (Statistics #2005-06)**, Laura Zayatz, "Disclosure Avoidance Practices and Research at the U.S. Census Bureau: An Update," June 16, 2005.

**RR (Survey Methodology #2005-04),** Elizabeth Martin and Eleanor Gerber, "Results of Recent Methodological Research on the Hispanic Origin and Race Questions," August 3, 2005.

**RR (Statistics #2005-07),** Hyunmo Kang, Sam Hawala, and Laura Zayatz, "IDFinder: Data Visualization for Checking Re-identifiability in Microdata," August 8, 2005.

**RR (Survey Methodology #2005-05),** Elizabeth Martin, Manuel de la Puente, and Claudette Bennett, "The Effects of Questionnaire and Content Changes on Race Data: Results of a Replication of 1990 Race and Origin Questions," August 9, 2005.

**RR (Survey Methodology #2005-06)**, Yuling Pan and Manuel de la Puente, "Census Bureau Guidelines for the Translation of Data Collection Instruments and Supporting Materials: Documentation on How the Guideline Was Developed," August 24, 2005.

**RR (Survey Methodology #2005-07),** Pamela Campanelli, Jennifer Rothgeb, and Elizabeth Martin, "The Role of Respondent Comprehension and Interviewer Knowledge in CPS Labor Force Classification," August 25, 2005.

**RR (Survey Methodology #2005-08),** Anna Chan and Jeffrey Moore, "The Respondent Identification Policy (RIP) and its Impact on Wave 2 and Wave 3 Follow-up Interviews in the 2004 SIPP Panel," August 30, 2005.

**RR (Survey Methodology #2005-09),** Yuling Pan, Barbara Craig, and Suzanne Scollon, "Results from Chinese Cognitive Interviews on the Census 2000 Long Form: Language, Literacy, and Cultural Issues," September 6, 2005.

## 3.5  STATISTICAL RESEARCH DIVISION STUDIES

**SS (Statistics #2005-02),**  Paul Massell, "Protecting Sensitive Cells in a Cell Suppression Program Using Sliding Protection," January 1, 2005.

**SS (Survey Methodology #2005-01)**, Eileen O'Brien, "Report on Cognitive Testing of the Housing Subsidy Questions in the American Housing Survey," February 10, 2005.

**SS (Survey Methodology #2005-02),**  Laura Loomis and Jennifer Rothgeb, "Final Report on Cognitive Interview Research Results and Revisions to the Welfare Reform Benefits Question for the March 2000 Income Supplement to the CPS," March 14, 2005.

**SS (Survey Methodology #2005-03)**, Jennifer Hunter and Manuel de la Puente, "Report on the Cognitive Testing of the Space Saving Features, Roster Questions, and Revised Instructions for the 2005 National Census Test," March 22, 2005.

**SS (Survey Methodology #2005-04),** Kristen Hughes, "Final Report of Cognitive Research on the 2005 Police Public Contact Survey: A Supplement to the National Crime Victimization Survey," August 5, 2005.

## 3.6  OTHER REPORTS

Hughes, K. and Crowley, M., "Final Report of Cognitive Research on the Youth Volunteering, Service, and Civic Engagement Survey," September 13, 2005.

Hunter, J.E., "Cognitive Test of the 2006 NRFU, Round 1," October 4, 2005.

Hunter, J.E. and DeMaio, T.J., "Report on Cognitive Testing of Tenure, Age, and Relationship Questions for the 2005 National Content Test," November 22, 2005.

Malakhoff, L.A. (2005). "An Accessibility Evaluation of Linux Desktop Applications," *Human-Computer Interaction Memorandum Series #77.*

Martin, D.E.K. and Aston, J.A.D. (2005). "Distribution of the *rth* Occurrence of a Compound Pattern in Higher-order Markovian Sequences." *Technical Report Series, 2005-03, Institute of Statistical Science, Academia Sinica, Taipei.*

Mulry, M.H. (2005). "Summary of Accuracy and Coverage Evaluation for Census 2000," *Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: CM-GEN-G-04.*

Murphy, E.D., Ciochetto, S.M., Malakhoff, L.A., Coon, D., and Smith, N. (2005). "2005 Census Internet Prototype Applications: Usability and Accessibility Testing," *Human-Computer Interaction Memorandum Series #81.*

Murphy, E.D. and Johnson, A.G. (2005). "Usability Testing of the Listing and Mapping Instrument (LAMI, VERSION 3.1)," *Human-Computer Interaction Memorandum Series #79.*

Olmsted, E. (2005). "A Usability Evaluation of the Integration of the American Community Survey Data into the American Factfinder Web site," *Human-Computer Interaction Memorandum Series #78.*

# 4. TALKS AND PRESENTATIONS

*Washington Statistical Society,* Washington, D.C., October 14, 2004.
  •  Bill Winkler, "Microdata Confidentiality Methods."

*Washington Statistical Society,* Washington, D.C., October 19, 2004.
  •  Eileen O'Brien, "From Controlled Experiment to Production Environment: Refusal Aversion Training Adoption, Considerations for Future Use and Research."

*132nd Annual Meeting of the American Public Health Association,* Washington, D.C., November 10-12, 2004.
  •  Adam Carle, "Establishing the Cross-group Validity of Psychological Measurement Instruments: Measurement Bias, Depression, and Epidemiological Research."
  •  Adam Carle, "An Empirical Example Describing the Use of the Quantitative Model for Assessing Measurement Bias in Public Health Research."
  •  Adam Carle, "Empirically Evaluating the Impact of Statistically Significant Measurement Bias."

*Epidemiology and Public Health, Global Health Division,* Yale University, New Haven, CT, November 11, 2005.
  •  Tommy Wright, "Constructing Denominators for Epidemiology and First Principles of Apportionment."

*Washington Statistical Society,* Washington, D.C., November 16, 2004.
  •  Bill Winkler, Discussant, "Playing With Matches and Other Issues in Data Quality."

*Bureau of Labor Statistics*, Washington, D.C., November 22, 2004.
  •  Tucker McElroy, "Finite Sample Signal Extraction."

*Federal Committee on Statistical Policy Seminar*, Washington, D.C., December 15-16, 2004.
  •  Terry DeMaio, Nancy Bates, Diane Willimack, and Jane Ingold, "Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses."

*9th INFORMS Computing Society Conference*, Annapolis, MD, January 4-6, 2005.
  •  Paul Massell, "Viewing Mathematical Programming Models for Cell Suppression as Dynamical Systems."

*Washington Statistical Society,* Washington, D.C., January 10, 2005.
  •  Tucker McElroy, Discussant.

*Washington Statistical Society,* Washington, D.C., January 19, 2005.
  •  Bill Winkler, "Machine Learning Methods for Text Classification."

*FedCASIC Workshops,* Washington, D.C., March 1-3
  •  Juan Pablo Hourcade, "Study of Handheld Computer Stylus Skills in Older Adults."
  •  Lawrence Malakhoff, "Accessibility Testing Methods Demonstration."

*Comparative Survey Design and Implementation Conference*, Madrid, Spain, March 10-12, 2005.
  •  Patti Goerman, "An Examination of Pretesting Methods for Multicultural, Multilingual Surveys: The Use of Cognitive Interviews to Test Spanish Instruments."
  •  Yuling Pan, "The Use of Interpreters in the Conduct of Household Surveys: Guidance on How Interpreters Should be Used in Data Collection With Non-English-Speaking Respondents."
  •  Yuling Pan and Brian Kleiner, "Cross-cultural Communicative Norms and Survey Research: Interviews With Survey Researchers and Translators."

*University of Texas Department of Information Systems and Operations Management,* Arlington, TX, March 29, 2005.
  •  Mary Mulry, "Census Data for Business Use."

*Human Factors in Computing Systems Conference (CHI 2005),* Portland, OR, April 7, 2005.
  •  Juan Pablo Hourcade, "Designing Public Government Websites."

*Confidentiality Session, International Statistical Institute*, Sydney, Australia, April 9, 2005.
  • Bill Winkler, Discussant of papers by Josep Domingo-Ferrer (Rovira I Virgili de Tarragona, Spain), Ross Sparks (CSIRO, Australia), and Stephen Fienberg (Carnegie-Mellon University).
  • Bill Winkler, "Methods for Automatic Editing and Imputation."

*Exploring Culture in Health Care*, University of Virginia, Charlottesville, VA, April 10, 2005.
  • Patti Goerman, "New Hispanic Immigrants in Central Virginia."

*Australia Bureau of Statistics,* Canberra, Australia , April 13-19, 2005.
  • Bill Winkler, "Machine Learning Methods for Text Classification."
  • Bill Winkler, "Microdata Confidentiality."
  • Bill Winkler, Panel Discussion, "Statistical Data Editing: Overview and Research Problems."
  • Bill Winkler, "Statistical Data Editing: Overview and Research Problems."

*Australia National University*, Canberra, Australia, April 15, 18, 2005.
  • Bill Winkler, "Record Linkage and Machine Learning."
  • Bill Winkler, "Record Linkage."

*Australian Bureau of Statistics,* Canberra, Australia, April 19, 2005.
  • Bill Winkler, "Non-Technical Overview of Record Linkage."

*QUEST 2005*, Heerlen, Netherlands, April 21, 2005.
  • Terri DeMaio, "Standards for Developing Questionnaires and Survey Related Materials for U.S. Census Bureau Surveys and Censuses."
  • Jennifer Rothgeb, "Using Quality Assessment Findings as Input for Questionnaire Revisions and Cognitive Testing Protocol."

*Federal Forecasters Conference*, Washington, D.C., April 21, 2005.
  • Tucker McElroy, "Forecasting Age Distribution Curves."

*Applied Survey Computing Conference*, London, England, April 22, 2005.
  • Jenny Hunter and Ashley Landreth, "Using Behavior-Coding to Analyze Interviewer/Respondent Interactions with a CAPI Device."

*Statistics Department Seminar, North Carolina State University*, Raleigh, NC, April 29, 2005.
  • Tommy Wright, "Simple Proof of an Important Inequality for Probability Sampling."

*Workshop for Latent Variable Models and Survey Data for Social Science Research,* Montreal, Quebec, Canada, May 4-6, 2005.
  • Adam Carle, "Data Quality and Non-Sampling Error on the American Community Survey."

*Washington Statistical Society Seminar on Privacy,* Bureau of Labor Statistics, Washington, D.C., May 4, 2005.
  • Phil Steel, "Privacy, Confidentiality, and the Protection of Health Data: A Statistical Perspective on the HIPAA Privacy Rule," Section on Risk Assessment.

*Annual Meeting of the American Association for Public Opinion Research*, South Miami Beach, Florida, May 12-15, 2005.
  • Jenny Hunter and Ashley Landreth, "Person-Based Data Collection in Practice: An Evaluation of Interviewer/Respondent Interactions."
  • Ashley Landreth, Poster Session, "First-Level Interactions and Final Outcomes: Adding Value to the Behavior Coding Method."
  • Jeff Moore, "Use of Dependent Interviewing Procedures to Improve Data Quality in the Measurement of Change."

*UNECE Statistical Data Editing Work Session*, Ottawa, Ontario, Canada, May 16-18, 2005.
  • Elizabeth Nichols, "Designing Interactive Edits for Electronic Surveys and Censuses: Issues and Guidelines."

- María García, "Implicit Linear Inequalities Edits and Error Localization in the SPEER Edit System."
- Bill Winkler and María García, "Modeling and Analysis with Data."
- María García, Discussant, "New and Emerging Methods."

*Georgetown University 2005 Faculty Seminar on Inclusive Teaching & Learning,* Georgetown University, May 25, 2005.
- Yuling Pan, "Classroom Dynamics: International Students' Perspectives and Concerns."

*13th Annual Meeting of the Society for Prevention Research,* Washington, D.C., May 26, 2005.
- Adam Carle, "Measurement Bias Across Hispanics and Non-Hispanic Caucasians on a Standardized Measure of Alcohol Abuse: Empirically Evaluating the Impact of Observed Measurement Bias," and "Cross-Cultural Differences in the DSM-1V Construct of Alcohol Abuse: Assessing Measurement Bias Across Hispanic and Caucasian Adolescents."

*University of California,* San Diego, CA, May 30, 2005.
- Tucker McElroy, "Finite-Sample Signal Extraction."

*Stat Canada Interchange,* Ottawa, Ontario, Canada, June 2-3, 2005.
- Phil Steel, "An Outline of the Microdata Access System Under Development at the Census Bureau."

*National Center for Health Statistics Journal Club,* Hyattsville, MD, June 8, 2005.
- Yuling Pan, "Cross-Cultural Communications in Federal Surveys."

*ACM Methods of Database Conference, Workshop on Information Quality in Information Systems*, Baltimore, MD, June 10, 2005.
- Bill Winkler, "Methods and Analyses for Determining Quality."

*ICSA 2005 Applied Statistics Symposium*, Bethesda, MD, June 12-15, 2005.
- Don Malec, "Bayesian 3-Dimensional Image Reconstruction from 2-Dimensional Electron Microcopy Transmission Images."

*Conference of European Statisticians, United Nations Economic and Social Council*, Geneva, Switzerland, June 13, 2005.
- Betty Murphy, "Usability and Accessibility Testing in Support of Internet Reporting."

*Program of the Funding Opportunity Seminar in Survey and Statistical Research*, Bureau of Labor Statistics, Washington, D.C., June 13, 2005.
- Jerry J. Maples, Discussant, "Topics in Small Area Estimation."

*Second International ACM SIGMOD Workshop on Information Quality in Information Systems,* Baltimore, MD, June 17, 2005.
- Ned Porter, "Unduplication Research for the 2010 Census."

*Third Annual Workshop in Cognitive Systems: Human Cognitive Models in System Design,* Santa Fe, NM, July 6-8, 2005
- Michelle Rusch, Les Miller, Elizabeth Murphy, & Sarah Nusser, "A Cognitive Model for User Behavior in a Map-based Census Listing Application" (poster).

*14th World Congress of Applied Linguistics,* Madison, Wisconsin, July 25, 2005.
- Andrew Jocuns, "Intersubjectivity: Knowledge in Mediated Discourse."

*Joint Statistical Meetings, American Statistical Association,* Minneapolis, MN, August 7-11, 2005.
- Barbara Altman, Theresa DeMaio, and Kristen Miller, "Developing, Testing, and Recommendations for Questions on Disability for the American Community Survey."
- Paul Beatty, Gordon Willis, Jennifer Hunter, Kristen Miller, and Jennifer Rothgeb, "Design of the Q-Bank: Determining Concepts, Content, and Standards."
- Sam Hawala (Panel Member), "The Promise and Pitfalls of Data Mining."
- Leann Karl and Ashley Landreth, "Design of the Census 2004 Coverage Research Follow-up Questionnaire."

- Donald Malec and Jerry Maples, "An Evaluation of Synthetic Small-area Census Coverage Error Using a Random-effect Model."
- Jerry Maples and William Bell, "Evaluation of School District Poverty Estimates: Predictive Models Using IRS Income Tax Data."
- Paul Massell, "Comparing Ways of Using 'Protection Flow' to Protect Magnitude Data Tables from Disclosures."
- Tucker McElroy, "Statistical Properties of Signal Extraction Diagnostics."
- Brian Monsell, "Issues in Estimating Easter Regressors Using regARIMA Models with X-12-ARIMA."
- Jeffrey Moore and Chester Bowie, "Striving for Data Quality: Pat Doyle's Legacy at the U.S. Census Bureau."
- Mary Mulry, Bruce Spencer, Eric Schindler, Tom Mule, and Nganba Nguyen, "Investigation of Extreme Estimates of Census Coverage Error for Small Areas."
- Phillip Steel, "Issue in Designing a Confidentiality Preserving Model Server."
- Jennifer Hunter, Cleo Redline, Rebecca Morrison, Diane Willimack, and Roberta Sangster, "Broadening the Horizons of Q-BANK: Expanding the Framework to Encompass Self-administered and Establishment Surveys."
- Yves Thibaudeau, "Measuring Discriminatory Power of Imputation Methods in an Enumeration."
- Thomas Trimbur, "Model-based Analysis of Seasonal Heteroskedasticity in Census Bureau Construction Time Series."
- Todd Williams, "The Development of Truth Decks for the 2010 Census Count Imputation Research."
- Bill Winkler, "Modeling and Quality of Masked Microdata."
- Laura Zayatz, "Data Confidentiality Issues From a Federal Agency Perspective."
- Laura Zayatz, "Introductory Overview Lecture on Data Confidentiality."

*Annual Meeting of the American Sociological Association,* Philadelphia, PA, August 15, 2005.
- Patricia Goerman, "Two Steps Forward, One Step Back: Gendered Experiences of Work and Family Life for New Hispanic Immigrants in Central Virginia."

*Computer Science Department, Cornell University*, September 19, 2005.
- Bill Winkler, "Methods of Re-identifying and Analyzing Masked Microdata."

# 5. STATISTICAL RESEARCH DIVISION SEMINAR SERIES

Seminar Series Team: Barbara Palumbo, Laurie Schwede, Yves Thibaudeau, Julie Tsay

Caroline Jarrett, Usability Consultant, Effortmark Limited, "Why Can't I Just Take a Paper Form and Slap It on the Web?" October 6, 2004.

Paul B. Massell, SRD, U.S. Census Bureau, "An Overview of Statistical Disclosure Control for Tables," October 27, 2004.

Joseph Waksberg and Daniel B. Levine, "Census Bureau Human Capital Management Council and Statistical Research Division Seminar Series Stories by Wise Elders on Mailout/Mailback Enumeration for the Decennial Census of Population and on the Current Population Survey," November 16, 2004.

Irwin R. Katz, Educational Testing Service, "Assessment of Information and Communication Technology (ICT) Literacy: Current Development and Research Agenda," November 16, 2004.

William E. Winkler, SRD, U.S. Census Bureau, "Data Quality: Automated Edit/Imputation and Record Linkage," December 1, 2004.

Frauke Kreuter, University of Maryland, "Question Characteristics and Interviewer Effects," December 13, 2004.

Ed Merkle, Ohio State University, "The Incorporation of Multiple Imputation in Bayesian Simulation Routines," January 12, 2005.

Tucker McElroy, SRD, U.S. Census Bureau, "Finite-Sample Signal Extraction," January 12, 2005.

Helen Nissenbaum, NYU School of Law, "Privacy in Context," January 19, 2005.

Margaret E. Martin, (retired) OMB and CNSTAT, "At the Interface Between Statistics and Policy: Encountering the Unexpected," February 9, 2005.

Justin Nguyen, Iowa State University, "Australian Electrical Energy Production," February 14, 2005.

Adam Carle, SRD, U.S. Census Bureau, "Assessing Sub-population Differences in the DSM-IV Constructs of Alcohol Use and Abuse Across Non-Hispanic Caucasians, Non-Hispanic African-Americans, and Hispanics," February 15, 2005.

Lisa R. Denogean, Cornell University, "Approximations of the Density Functions of Estimators in Population Genetics," March 9, 2005.

Adam Carle, SRD, Census Bureau, "Data Quality of the American Community Survey Across Individuals Living in Linguistically Isolated and Non-Linguistically Isolated Households: A Latent Variable Model Assessment," March 17, 2005.

Harry Hochheiser, University of Maryland, "Dynamic Query Tools for Analysis and Interpretation of Time Series Data," March 24, 2005.

Joseph Powers, U.S. Census Bureau, "Using Census Data to Define Estimation Areas for the American Community Survey: A Case Study," April 6, 2005.

Cavan Capps, U.S. Census Bureau, "Interoperability in the Context of Data Dissemination: The Harvard Virtual Data Library and DataFerrett," May 3, 2005.

Roeland Beerten, Office of National Statistics, United Kingdom, "Designing New Statistical Sources in the United Kingdom," May 18, 2005.

Juan José Salazar-González, Universidad de La Laguna, Tenerife, Spain, "An Automatic Tool to Protect Statistical Tables with Controlled Rounding," May 19, 2005.

Michael Larson, Iowa State University, "Hierarchical Bayesian Record Linkage and Applications at Census," August 29, 2005.

Thomas Trimbur, SRD, U.S. Census Bureau, "Seasonal Heteroskedasticity and Trend Estimation in Monthly Time Series," September 15, 2005.

Jennifer Beck, National Opinion Research Center, "Context Effects on Metamemory Assessments," September 21, 2005.

# 6. PERSONNEL ITEMS

## 6.1 HONORS/AWARDS/SPECIAL RECOGNITION

### Customer Service Award

- **Elizabeth Ann Dimler** - for continued outstanding service to customers who make use of the R&D 2007 Contracts.

## 6.2 SIGNIFICANT SERVICE TO PROFESSION

Pat Cantwell
- Associate Editor, *Journal of Official Statistics.*
- Associate Editor, *Survey Methodology.*
- Member, Council of Sections Nominations Committee, American Statistical Association.
- Refereed papers for the *Journal of the American Statistical Association*, and the journal *Statistical Methodology.*

Manuel de la Puente
- Member, Nominations Committee, American Sociological Association.
- Board Member, Federal Executive Institute Alumni Association.

Terry DeMaio
- Refereed papers for *Public Opinion Quarterly*.
- Member, Editorial Board, *Public Opinion Quarterly*.

Jeremy Funk
- Member, Confidentiality and Data Access Committee (CDAC).

María García
- Member, UNECE Work Session on Statistical Data Editing Steering Committee.
- Session Organizer and Discussant, UNECE Work Session on Statistical Data Editing.

Sam Hawala
- Member, Confidentiality and Data Access Committee.
- Member, National Center for Education Statistics Disclosure Review Board.

Juan Pablo Hourcade
- Papers Co-Chair, Interaction Design and Children 2005 Conference, Boulder, CO, June 8-10.
- Reviewer, *Interacting with Computers*, *Human-Computer Interaction*, Human Factors in Computing Systems (CHI 2005) Conference.
- Member, National Science Foundation Advanced Learning Technologies Panel.

Jenny Hunter
- Elected Secretary, Washington-Baltimore Chapter of AAPOR.
- Member, Program Committee, D.C.-AAPOR.
- Member, Interagency Steering Committee for Q-Bank Development.

Donald Malec
- Refereed papers for *Statistics in Medicine, Journal of Official Statistics, Journal of the American Statistical Association* and *The American Statistician*.

Donald Martin
- Member, Committee on Minorities in Statistics, American Statistical Association.

Paul Massell
- • Member, Confidentiality and Data Access Committee.
- • Member, Bureau of Transportation Statistics Disclosure Review Board.
- • Member, Privacy Policy and Research Committee.
- • Refereed a paper for the *Journal of Official Statistics.*

Brian Monsell
- • Webmaster and *AMSTAT* Online Assistant Editor of the Business and Economic Statistics Section of ASA.
- • Publications Officer, ASA Business and Economic Statistics Section.

Jeff Moore
- • Refereed papers for *Public Opinion Quarterly* and *Survey Methodology.*

Mary Mulry
- • Associate Editor, *The American Statistician.*
- • Refereed a paper for the *Journal of Official Statistics*.

Beth Nichols
- • Refereed a paper for the *Journal of Official Statistics.*

Eileen O'Brien
- • Member, Survey Review Committee, American Statistical Association.
- • Refereed papers for *Journal of Official Statistics* and *Public Opinion Quarterly.*
- • Organizer, AAPOR 2004 Invited Session.
- • Designed a member survey for the Statistical Consulting Section of the ASA.
- • Reviewed the joint ASA/College Board longitudinal survey of students who have taken AP Statistics Exam.
- • Developed formal criteria ASA will use to decide whether to endorse requests for surveys of its members, committees or functions.

Yuling Pan
- • Refereed a paper for *Language in Society.*
- • Member, Advisory Board, *Journal of Politeness Research.*

Jennifer Rothgeb
- • Secretary-Treasurer, Executive Council, American Association of Public Opinion Research.
- • Refereed a manuscript for *Public Opinion Quarterly.*
- • Member, Planning Committee for QUEST 2005 International Work Group Workshop at Statistics Netherlands.
- • Member, Interagency Steering Committee for Q-Bank Development.
- • Member, Finance (Investment) Committee, American Association of Public Opinion Research.
- • Member, Committee on Committees, American Association of Public Opinion Research.
- • Session Chair, Annual Meeting of American Association of Public Opinion Research.

Laurie Schwede
- • Refereed a paper for the *Journal of Official Statistics.*
- • Organizer, American Anthropological Association Meeting Session.
- • Census Bureau representative in the Informal International Working Group on the Generations and Gender Programme, sponsored by the UNECE Population Activities Unit, in Istanbul, Turkey.

Eric Slud
- • Associate Editor, *Journal of the Royal Statistical Society, Series B (Methodological).*

Phil Steel
- • Vice-Chair, Confidentiality and Data Access Committee.
- • Member, American Statistical Association's Committee on Privacy and Confidentiality.

- Organized a Washington Statistical Society Seminar on the Research Triangle Institute's MASSC. software.

Thomas Trimbur
- Refereed a paper for the *Journal of Applied Econometrics*.

Bill Winkler
- Refereed three papers for *Data Mining and Knowledge Discovery*, and one paper for the *IEEE Transactions on Knowledge and Data Engineering.*
- Along with Phil Steel, organized Washington Statistical Society talk.
- Associate Editor, *Journal of Privacy Technology*.
- Updated widely used reference lists for microdata confidentiality, record linkage, and statistical data editing that are available at various sites including  http://www.hcp.med.harvard.edu/statistics/survey-soft/.

Tommy Wright
- Refereed a paper for *The American Statistician.*
- Associate Editor, *The American Statistician.*
- Associate Editor, *The American Journal of Mathematical and Management Sciences*
- Member, Department of Applied and Engineering Statistics Advisory Council, George Mason University.

Laura Zayatz
- Member, Confidentiality and Data Access Committee.
- Member, Advisory Board, *Journal of Privacy Technology.*
- Refereed two papers for the *Journal of Official Statistics.*
- Reviewed a book proposal on confidentiality for Springer.

## 6.3   PERSONNEL NOTES

Eric Raymond (a graduate student in Human Factors Psychology and Applied Cognition) joined our Human Factors and Usability Research Group.

Rubén Mera retired from the Census Bureau.

Hazel Beaton retired from the Census Bureau after over thirty years of federal service.

Joanne Pascale went on leave from the Census Bureau in January for one year for a temporary appointment at the National Centre for Social Research (London, England).

Kwon Soon Moon, visitor to the Census Bureau who was hosted by the Time Series Research Group, returned to the Korean National Statistical Office.

Bonnie Kegan completed a detail (Time Series) in our division and returned to the Manufacturing and Construction Division.

Rolando Rodríguez joined our Missing Data Methods Research Group.

Amira Abdalla (a graduate student in Human Factors and Applied Cognition at George Mason University) joined our Human Factors and Usability Research Group.

Julia Klein Griffiths resigned from the Census Bureau.

Melinda Crowley accepted a position at the State Department.

Diana Simmons joined our division.

Maria Cantwell joined the Demographic Directorate.

Richard Gagnon received his Ph.D. in Applied Mathematics from the University of Maryland, College Park and was converted from a student to a regular member of our Time Series Research Group.

Andy Jocuns joined the Questionnaire Pretesting for Household Surveys Group as a Postdoctoral Researcher.

Summer Visitors:
    Joint Program in Survey Methodology Junior Fellows:
      -   Theresa Berkel (junior in Psychology at Princeton University).
      -   Maria Bruun (senior in Mathematics at Carleton College).
      -   Tony Evans (junior in Mathematical Methods in the Social Sciences at Northwestern University).
      -   Danielle Sullivan (junior in Mathematics/Statistics at Winona State University).
    Karie Grow (junior at the University of Maryland, College Park).
    Kate Guttridge (graduate student in Computer Science at the University of Texas, Austin).
    Michelle Rusch (graduate student in Human-computer Interaction at Iowa State University).

Jeremy Funk joined the Disclosure Avoidance Research Group.

Adam Carle accepted a faculty position at the University of South Carolina, Aiken.

Kristen Hughes accepted a position at the Bureau of Justice Statistics.

Jennifer Rothgeb spent six weeks on detail with FEMA in Texas working on hurricane relief efforts with evacuees from Hurricanes Katrina and Rita.

| APPENDIX A | Statistical Research Division's FY 2005 Program Sponsored Projects/Subprojects<br>With Substantial Activity and Progress and Sponsor Feedback<br>(Basis for PERFORMANCE MEASURES) | | |
|---|---|---|---|
| Project # | Project/Subproject Sponsor(s) | SRD Contact | Sponsor Contact |

**DECENNIAL**

| Project # | Project/Subproject Sponsor(s) | SRD Contact | Sponsor Contact |
|---|---|---|---|
| 5210501 | Forms Development | | |
| 5210502 | Content Planning and Development | | |
| | 1. *Census Questionnaire Design Features (NonResponse Follow-up Behavior Coding* | Ashley Landreth | Courtney Reiser |
| | 2. *Census Questionnaire Design Features (Behavior Coding)* | Ashley Landreth | Elizabeth Krejsa |
| | 3. *Census Questionnaire Design Features (Deadline Messaging)* | Ashley Landreth | Kimberly Higginbotham |
| | 4. *Short Form Questionnaire Content Other than Race & Ethnicity* | Terry DeMaio | Jane Ingold |
| | 5. *Development of Race and Ethnicity Questions* | Eleanor Gerber | Kathleen Styles |
| 5210503 | 6. *Language Planning and Development* | Patti Goerman | Jane Ingold |
| 5310501 | Data Collection Planning and Development | | |
| 5311501 | IT Data Collection Planning and Development | | |
| | 7. *Field Research on the Usability of Handheld Computers* | Betty Murphy | Karen Medina |
| | 8. *Interaction Design Support for the 2006 Non-Respondent Follow-up Instrument* | Juan Pablo Hourcade | Darlene Monaco |
| | 9. *2005 Internet Design Team* | Betty Murphy | Jennifer Lins |
| | 10. *Support for Field Data Collection Automation (FDCA) Program* | Juan Pablo Hourcade | Arnold Jackson |
| | 11. *Listing and Mapping Instrument (LAMI) Usability* | Betty Murphy | David Earles |
| | 12. *2006 Census Test Telephone Questionnaire Assistance (TQA) Implementation Team* | Larry Malakhoff | Dave Coon |
| 5310508 | Special Place/Group Quarters Planning and Development | | |
| | 13. *Decennial Census Group Quarters Research and Planning* | Laurie Schwede | Annetta Clark Smith |
| | 14. *Ethnographic Study of Hotels and Motels* | Leslie Brownrigg | Rick Bitzer |
| 5610502 | Statistical Design and Estimation | | |
| | 15. *Decennial Editing and Imputation* | Yves Thibaudeau | Rick Griffin |
| | 16. *Decennial Record Linkage* | William Winkler | Maureen Lynch |
| | 17. *Research on Item and Count Imputation for Implementation in Census 2010* | Yves Thibaudeau | Donna Kostanich |
| | 18. *Decennial Disclosure Avoidance* | Laura Zayatz | Marie Pees |
| | 19. *Census Unduplication Research* | Michael Ikeda | Maureen Lynch |
| 5610503 | Coverage Measurement Planning and Development | | |
| | 20. *Coverage Measurement Research* | Don Malec | Rick Griffin |
| | 21. *Accuracy of Coverage Measurement* | Mary Mulry | Donna Kostanich |
| | 22. *Questionnaire Wording and Automation Team* | Beth Nichols | David Whitford |
| 5610505 | Coverage Improvement Planning and Development | | |
| | 23. *Decennial Privacy Research* | Jeff Moore | Shelly Martinez |
| | 24. *Development of Questionnaires for Decennial Coverage Improvement* | Eleanor Gerber | Ed Byerly |
| | 25. *Inter-divisional Decennial 2010 Working Groups on Residence Rules and Coverage Improvement* | Laurie Schwede | Frank Vitrano |
| 5385560 | American Community Survey (ACS) | | |
| | 26. *ACS Questionnaire Design Measurement* | Manuel de la Puente | Debbie Griffin |
| | 27. *ACS Labor Force Questions* | Jennifer Rothgeb | Tom Palumbo |
| | 28. *ACS Small Area Estimation Research* | Don Malec | Freddie Navarro |
| | 29. *ACS Missing Data and Imputation* | Yves Thibaudeau | Debbie Griffin |
| | 30. *ACS Disclosure Avoidance* | Laura Zayatz | Marie Pees |
| | 31. *ACS Weighting Research* | Lynn Weidman | Freddie Navarro |
| | 32. *ACS Usability Study* | Erica Olmsted | Doug Hillmer |
| | 33. *ACS Group Quarters Research* | Laurie Schwede | Wendy Hicks |
| | 34. *ACS Language Research* | Yuling Pan | Deborah Griffin |
| | 35. *Evaluation of ACS Format* | Eleanor Gerber | Wendy Hicks |

| | DEMOGRAPHIC | |
|---|---|---|
| 0906/7374 | 36. *Data Integration* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Sam Hawala . . . . . . . . . . . . . . . . Larry Cahoon |
| 1461 | 37. *SIPP Methods Panel* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Jeff Moore . . . . . . . . . . . . . . . . Kathy Creighton |
| 1465 | Survey of Income and Program Participation (SIPP) Research | |
| | 38. *Longitudinal Weighting* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Leroy Bailey . . . . . . . . . . . . . . . . Sam Sae-Ung |
| | 39. *Quick Turnaround Pretesting of Household Surveys-NCVS* | |
| | *Stalking Supplement* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Terry DeMaio . . . . . . . . . . . Marilyn Monahan |
| | 40. *Quick Turnaround Pretesting of Household Surveys-Public* | |
| | *Police Contact Survey* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Terry DeMaio . . . . . . . . . . . Marilyn Monahan |
| | 41. *Quick Turnaround Pretesting of Household Surveys-* | |
| | *Fish/Hunt Survey* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Terry DeMaio . . . . . . . . . . . . . Heather Holbert |
| 7558111 | 42. *SIPP Assets/Liabilities Imputation Research/Software Design* | Yves Thibaudeau . . . . . . . . . Thomas Palumbo |
| 7478 | 43. *American Housing Survey-Metro* . . . . . . . . . . . . . . . . . . . . . . . | Eileen O'Brien . . . . . . . . . . . . . . . . Craig Pritzl |
| 7565084 | 44. *National Survey of Youth Volunteering & Civic Engagement* | Terry DeMaio . . . . . . . . . . . . . Kenneth Kaplan |
| TBA | 45. *National Survey of College Graduates - Cognitive Research* | |
| | *Projects* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Jennifer Rothgeb . . . . . . . . . . . Dawn Nelson |
| 7485000 | 46. *Nonresponse Bias Analysis* . . . . . . . . . . . . . . . . . . . . . . . . . . . | Aref Dajani . . . . . . . . . . . . . . . . Dawn Nelson |
| 7165 | Research for Small Area Income and Poverty Estimates (SAIPE) | |
| | 47. *Research for Small Area Income and Poverty Estimates* . . . . . | Elizabeth Huang . . . . . . . . . David Waddington |
| | 48. *Small Area Estimation Methodology for SAIPE* . . . . . . . . . . . | Eric Slud . . . . . . . . . . . . . . . . . . . . . Robin Fisher |
| TBA | 49. *Analysis and Forecasting of Demographic Time Series* . . . . . . | Tucker McElroy . . . . . . . . . . . . . Fred Hollman |
| | **ECONOMIC** | |
| 2320554 | 50. *Editing Methods Development (Investigation of Selective* | |
| | *Editing Procedures for Foreign Trade Programs)* . . . . . . . . . . | Maria Garcia . . . . . . . . . . . . . . Debra Coaxum |
| 2420551 | 51. *Disclosure Avoidance Methods* . . . . . . . . . . . . . . . . . . . . . . . | Laura Zayatz . . . . . . . . . . . . . . . . . John Fowler |
| 2420552 | Time Series Research | |
| | 52. *X-12-ARIMA Development and Evaluation* . . . . . . . . . . . . . . . | Brian Monsell . . . Kathleen McDonald-Johnson |
| | 53. *Research on Seasonal Time Series - Modeling and Adjustment* | |
| | *Issues* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Tucker McElroy . Kathleen McDonald-Johnson |
| | **STATISTICAL RESEARCH DIVISION** | |
| 8150 | 54. *Postal Rate Commission/Statistical Consulting* . . . . . . . . . . . | Leroy Bailey . . . . . . . . . . . . . . . . . . . John Waller |
| 8863 | 55. *National Institute of Standards and Technology/Bayesian* | |
| | *Statistical Methodology* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Don Malec . . . . . . . . . . . . . . . . . Nell Sedransk |
| Other | Usability/Field Related | |
| | 56. *Optimizing Field Operations* . . . . . . . . . . . . . . . . . . . . . . . . . | Bor-Chung Chen . . . . . . . . . . . . . Richard Blass |
| | 57. *Usability Evaluation of Linux Desktop* . . . . . . . . . . . . . . . . . . | Juan Pablo Hourcade . . . . . . . . . . . . Tia Harris |
| | 58. *Desktop Applications Accessibility-Landview 5* . . . . . . . . . . . | Larry Malakhoff . . . . . . . . . . . . . . Paul Manka |
| | 59. *Desktop Applications Accessibility-Econ 2002 Data Browser* | Larry Malakhoff . . . . . . . . . . . Janet McGrane |
| | 60. *Desktop Applications Accessibility-Linux Desktop* . . . . . . . . | Larry Malakhoff . . . . . . . . . . . . . . . . Tia Harris |
| | 61. *Web Applications Accessibility-LEHD* . . . . . . . . . . . . . . . . . | Larry Malakhoff . . . . . . . . . . . . . Jennifer Marks |
| | 62. *Web Applications Accessibility-DAPPS* . . . . . . . . . . . . . . . . . | Larry Malakhoff . . . . . Alessandro Rebaudengo |
| | 63. *Web Applications Accessibility-Public Use Presentation* | |
| | *Library* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Larry Malakhoff . . . . . . . . . . . Andrea Sevetson |
| | 64. *Usability Testing of American FactFinder Economic Data and* | |
| | *Map Data* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | Erica Olmsted . . . . . . . . . . . . . . Marian Brady |
| | 65. *Support for New Statistical Abstract Website* . . . . . . . . . . . . . | Juan Pablo Hourcade . . . . . . . . Lars Johanson |
| | 66. *Interviewer Behavior Research* . . . . . . . . . . . . . . . . . . . . . . . | Eileen O'Brien . . . . . . . . . . . . . . . Rick Bitzer |
| | 67. *Census.gov Usability Study Main Page* . . . . . . . . . . . . . . . . . | Erica Olmsted . . . . . . . . . . . . . . Lisa Wolfisch |
| | 68. *Topic Based Usability Study on Census.gov* . . . . . . . . . . . . . . | Erica Olmsted . . . . . . . . . . . . . . . . . . Jan Ennis |

**FY 2005 PROJECT PERFORMANCE**
**MEASUREMENT QUESTIONNAIRE**
**STATISTICAL RESEARCH DIVISION**

Dear

In a continuing effort to obtain and document feedback from program area sponsors of our projects or subprojects, the Statistical Research Division will attempt for the sixth year to provide *seven measures of performance,* this time for the fiscal year 2005. For FY 2005, the *measures of performance* for our division are:

> *Measure 1. Overall, Work Met Expectations:* Percent of FY
>> 2005 Program Sponsored Projects/Subprojects where sponsors reported that work met their expectations.
>
> *Measure 2. Established Major Deadlines Met:* Percent of
>> FY 2005 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met.
>
> *Measure 3a. At Least One Improved Method, Techniques*
>> *Developed, Solution, or New Insight:* Percent of FY 2005 Program Sponsored Projects/Subprojects reporting at least one improved method, techniques developed, solution, or new insight.
>
> *Measure 3b. Plans for Implementation:* Of the FY 2005 Program
>> Sponsored Projects/Subprojects reporting at least one improved method, techniques developed, solution, or new insight, the percent with plans for implementation.
>
> *Measure 4. Predict Cost Efficiencies:* Number of FY 2005
>> Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency."
>
> *Measure 5. Journal Articles, Publications:* Number of journal
>> articles (peer review) and publications documenting research that appeared or were accepted in FY 2005.
>
> *Measure 6. Proceedings Publications:* Number of proceedings
>> publications documenting research that appeared in FY 2005.

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

To construct these seven measures for our division, we will combine the information for all of our program area sponsored projects or subprojects obtained during October 7 thru October 21, 2005 using this questionnaire. Your feedback is requested for:

> Project Number and Name: _____
> Sponsoring Division(s): _____

After all information has been provided, the SRD Contact _____ will ensure that the signatures are obtained in the order indicated on the last page of this questionnaire.

**TIMELINESS:**
**Established Major Deadlines/Schedules Met**

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

☐ Yes    ☐ No    ☐ No Established Major Deadlines

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

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

**2.**   Listed below are at most 2 of the top improved methods, techniques developed, solutions, or new insights offered or applied on this project or subproject in FY 2005 where an SRD staff member was a significant contributor. Review "a" and "b" below **(provided by SRD Contact)** and make any additions or deletions as necessary. For each, please indicate whether or not there are plans for implementation. If there are no plans for implementation, please comment.

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

☐   Yes as listed below. (See a and b.)

Plans for
Implementation?
a.   _____   Yes ☐    No ☐   _____

_____

_____
_____

b.   _____   Yes ☐    No ☐   _____

_____

_____

_____

**Comments (Sponsor Contact):**

**COST:**
**Predict Cost Efficiencies**

**3.**   Listed **(provided by SRD Contact)** below are at most two research results or products produced for this project or subproject in FY 2005 that predict cost efficiencies. Review the list, and make any additions or deletions as necessary. Add any comments.

☐   No cost efficiencies predicted.

☐   Yes as listed below. (See a and b.)

a.

_____   b.

_____

**Comments (Sponsor Contact):**

**OVERALL:**
**Expectations Met/Improving Future Communications**

**4.**   Overall, work on this project or subproject by SRD staff during FY 2005 met expectations. **(Sponsor Contact)**

☐   Strongly Agree
☐   Agree
☐   Disagree
☐   Strongly Disagree

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

*(SRD Contact will coordinate first two signatures as noted and pass to SRD Chief.)*

First   _____ -
          Sponsor Contact Signature                    Date

Second   _____
            SRD Contact Signature                       Date

*(SRD Chief will coordinate last two signatures as noted.)*

Third   _____
          Sponsor Division Chief Signature             Date

Fourth   _____
            SRD Division Chief Signature               Date

# Statistical Research Division

Assistant Division Chief
Robert Creecy
  Barbara Palumbo
  Judi Norvell
  Chad Russell

### Machine Learning & Computational Statistics Research
Bill Winkler
  Mohammed Chaudhry
  William Yancey
  VACANT
  VACANT

### Computing Applications
Aref Dajani
  Pam Ferrari
  Tom Petkunas
  Ned Porter
  Mary Ann Scaggs
  VACANT

### Missing Data Methods Research
Yves Thibaudeau
  Bor-Chung Chen
  Maria Garcia
  Rolando Rodriguez
  VACANT

Assistant Division Chief
Pat Cantwell
  Alice Bell
  Gloria Prout

### Sampling Research
Lynn Weidman
  Ann Dimler
  Mike Ikeda
  Mary Mulry
  Julie Tsay
  VACANT

### Small Area Estimation Research
Don Malec
  Elizabeth Huang
  Jerry Maples

### Statistical Estimation & Analysis Research
Leroy Bailey
  Eric Slud (U. of MD)
  VACANT
  VACANT

### Disclosure Avoidance Research
Laura Zayatz
  Jeremy Funk
  Sam Hawala
  Paul Massell
  Phil Steel
  VACANT

### Time Series Research
Brian Monsell
  Richard Gagnon
  Don Martin (Howard U)
  Tucker McElroy
  Thomas Trimbur (PostDoc)
  VACANT

Assistant Division Chief
Manuel de la Puente
  Tina Arbogast
  Diana Simmons

### Questionnaire Design & Measurement Research-1
Jeff Moore
  Anna Chan
  Beth Nichols
  Joanne Pascale (L)
  Jennifer Rothgeb

### Questionnaire Design & Measurement Research-2
Eleanor Gerber
  Eileen O'Brien
  Yuling Pan
  Laurie Schwede
  VACANT

### Questionnaire Pretesting for Household Surveys
Terry DeMaio
  Jennifer Beck (Start 12/12/2005)
  Patti Goerman (PostDoc)
  Jennifer Hunter
  Andy Jocuns (PostDoc)
  Ashley Landreth
  Lorraine Randall

### Human Factors & Usability Research
Betty Murphy
  Amira Abdalla (S)
  Leslie Brownrigg
  Sherae Daniel (S)
  Joyce Farmer
  Juan Pablo Hourcade
  Larry Malakhoff
  Erica Olmsted
  VACANT

Tommy Wright, Chief
  Kelly Taylor
  Michael Hawkins
  VACANT

(L) On Leave
(S) Student