Annotated Bibliography of
BLS Statistical Papers
1999-2003

Compiled by the
Office of Survey Methods Research
U.S. Bureau of Labor Statistics
2 Massachusetts Ave. N.E.
Washington DC 20212

April 2003

Note: The Annotated Bibliography is now available on the internet. In addition to the abstracts, many of the papers are also available. The location is:

http://www.bls.gov/ore/orebib.pdf

(Authors who were BLS employees at the time of the writing of each paper are shown in boldface type.)

Allard, Mary, Mikkelson, Gordon and **Unger, Linda** (2001), "Researching a Score Function to Prioritize Business Survey Edit Failures at BLS," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Covered Employment and Wages Program, commonly referred to as the ES-202 program, is a cooperative program between the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor and the State Employment Security Agencies. The ES-202 program produces a comprehensive tabulation of employment and wage information for workers covered by State unemployment insurance laws and for Federal workers covered by the Unemployment Compensation for Federal Employees program. Standardized State processing systems include over 150 microdata edits and ten macrodata edits. Recent time-use studies have shown that State staff spend approximately one-fourth of their time reviewing, researching, correcting, and explaining suspect data. Macro level data are also edited and reviewed at the BLS after the States have submitted the data quarterly. BLS and State agencies have recently undertaken an initiative to develop a score function that could prioritize this work so that suspicious microdata that may have the most important impact on published macro level data may be more easily identified and reviewed. We empirically tested several score functions and present the results in this paper.

Balazik, Matthew, Crankshaw, Mark and **Mueller, Charlotte**, (2001) "Ensuring Data Quality in The Job Openings and Labor Turnover Survey*," Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Bureau of Labor Statistics (BLS) is producing a new survey collecting job openings and labor turnover data. The Job Openings and Labor Turnover Survey (JOLTS) collects data and produces estimates for job openings, hires, and separations. This survey is expected to provide valuable statistics to help evaluate the U.S. economy. The JOLTS data will help develop a complete picture of the economy when used in conjunction with the monthly employment level and the unemployment rate. In an effort to improve data quality and the resulting estimates, BLS has been editing and reviewing data starting with March 2000, the survey's first month of collected data. As with any new program, JOLTS has had to start from scratch to determine data edits and microdata review and how to handle all the output from these procedures. The edits and parameters were determined based on the nature of the JOLTS data, procedures of other surveys, and those of a JOLTS pilot study. Methodology for outlier detection and management was selected and implemented. Microdata review is conducted each month to identify problem records and correct data errors before final estimates are run.

Blaha, Jeffrey, Johnson-Herring, Sylvia, Krieger, Sharon and **Ferguson, Martine**, (2003), Some Practical Considerations in Choosing a Survey's Sample Design,"

Proceedings of the Section on Survey Research Methods, American Statistical Association.

The Consumer Expenditure Survey is a nationwide survey of households conducted by the U.S. Bureau of Labor Statistics and U.S. Bureau of Census to find out how Americans spend their money. The current sample design consists of selecting noncontiguous housing units for participation in the survey. This method minimizes variance but can result in high survey costs, particularly in relation to the amount of travel required by data collectors. In this paper we summarize research conducted to determine the feasibility of sampling clusters of contiguous housing units so as to reduce survey costs but still keep variances small.

Bobbitt, Patrick, (2003), A Graphical Method for Assessing the Results of Fuzzy Clustering," *Proceedings of the Section on Statistics Graphics ,* American Statistical Association.

When attempting to ascertain the clustering of groups in the absence of other information, the standard technique has been to use "typical" clustering methods such as K-Means clustering to assess the closeness; then on the basis of an arbitrary criterion of "close," assign elements that are close to the same grouping. There are some applications in which this method could be seen as taking data reduction too far because it reduces all of the information of closeness into a series of indicator variables, and gives no information to the end user as to possible "leanings" into "neighboring" groups. To address this issue, a method of fuzzy clustering was developed that assigned to each data element a probability of inclusion into an arbitrary number of groupings (effectively a weighting that is applied to the indicator variables). The output from this method, unfortunately, leads to more data that may be difficult to quickly interpret. We have developed a method whereby the probabilities of inclusion are mapped to a set of primary colors whose mixture will allow the end-user a quick summary of the information produced by the fuzzy clustering algorithm.

Bosley, John, Fisher, Sylvia, Mockovak, William, Goldenberg, Karen and Tucker, Clyde, (2003), "A Qualitative Study of Nonresponse Factors Affecting BLS Establishment Surveys: Results," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

This paper will provide a summary of the information gained from a qualitative study examining factors associated with establishment survey nonresponse in four BLS surveys. The study focused on documenting the nature and possible causes of nonresponse in these surveys. The findings stem from extensive qualitative research conducted with BLS program and field staff, as well as establishments that respond, drop-out, fail to respond, or intermittently report their BLS survey data. Overall results obtained across the four BLS establishment surveys under study will be reported. The paper includes a set of recommendations that, if implemented, could reduce BLS establishment survey nonresponse.

Bosley, John, Dashen, Monica, and **Fox, Jean** (1999), "Effects on List Length and Recall Accuracy of Order of Asking Follow-up Questions About Lists of Items Recalled in Surveys," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Surveys often require respondents to recall a number of items belonging to a class or category, and also to recall significant facts or attributes about each recalled item. For example, the survey may ask for recall of purchases within a category such as "furniture" during some reference period, and to also recall for each item retrieved from memory some descriptive details (item attributes), such as price, date of purchase, model number, and the like. This purpose of the study was to examine the effects of two different ways of presenting a complex free recall task on the amount and accuracy of information retrieved from memory. The recall task focused on retrieving multiple instances of item within named categories, along with descriptive item details. The study findings are intended to apply to recall of purchased items, and so an experimental procedure simulating actual purchases was carried out on all study participants. Their simulated purchase and pertinent details about each one were recorded to serve as a basis for assessing accuracy of recall. After an interval during which they worked on a distracter task, participants were presented with a recall task. In one condition, they were asked to recall several items within a category before being asked to supply detail about any of the items. In the second condition, they supplied details about each item as soon as it was recalled. Outcome measures included the number of items per category recalled, accuracy of recall of both items and item detail (attributes), ratings of perceived task difficulty, and frequency of verbal and non-verbal behaviors indicative of fatigue, boredom and other signs of motivational or affective reactions to the task, or of cognitive burden. Study findings will be presented and discussed from the perspective of what cognitive and motivational factors account for any differences in the effectiveness of the two methods for eliciting information of this nature.

Bournazian, Jacob, Eden, Donna, Greenberg, Brian and **Cohen, Steve**, (2002), "Access to Confidential Statistical Agency Data," *Proceedings of the Section on Government Statistics,* American Statistical Association.

Federal statistical agencies face increasing complexity in our information and technological environment for developing disclosure avoidance and restrict access procedures to preserve the confidentiality of survey data and protect individual privacy rights. Critical topics for understanding the issues relating to preserving data confidentiality and expanding public access to statistical data that will be discussed include: Overview: legal issues relating to an agency's responsibility and obligation to protect the confidentiality of the survey data--Privacy Act, Freedom of Information Act, and State Laws; Statistical Disclosure Limitation Techniques (SDL) for preserving the confidentiality of published data in microdata files and tabular data; the CDAC checklist on the potential disclosure risks of proposed data releases; and Restricted Access Procedures.

Bradley, Ralph (2001), "A Simple Binary Model for the Self Identification of Sexual Orientation," *Proceedings of the Section on Health Policy Statistics,* American Statistical Association.

Because of sample selection bias, historical estimates on the non heterosexual fraction of the population may be biased. Additionally, comparisons of education and income between heterosexuals and non heterosexuals could be biased. This study develops a microeconomic model of self-selection to explain the cause of this potential bias. It then uses the micro model to develop the structural equations for the econometric estimation of the total non-heterosexual fraction of the population. The econometric method that is used to account for this sample selection bias is similar to Heckman (1979) except that the variable of interest is discrete in this study. The National Health and Social Life Survey estimates that 2.6% of all men self identify as non-heterosexual. Yet, using the same data, I get an estimate that 8.2% of all men are non heterosexual.

Bradley, Ralph, (2000); "Finite Sample Effects in the Construction of Price Indexes," *Proceedings of the Section on Government Statistics,* American Statistical Association.
No abstract available.

Butani, Shail J., and **McElroy, Michael** (1999), " *Managing Various Customer Needs for* Occupational Employment Statistics and Wage Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Originally, Occupational Employment Statistics (OES) Survey was designed to measure occupational employment by industry at the state level. For example, number of electrical and electronic engineers in research and testing services industry (Standard Industrial Classification Code 873). In the early 1990's, about 15 states asked guidance from U.S. Bureau of Labor Statistics to collect and produce wage rates for occupations. In response to the increasing demand for occupational wage rates, the OES sample expansion was implemented in all states plus District of Columbia with the 1996 survey. In this paper, we describe the major changes made to the survey design in order to meet the needs of new sponsors, states, and other data users. The implementation of the new design took place in a federal/state environment under an extremely tight time schedule.

Chiu, Christina, Robertson, Kenneth, and **Tou, Albert** (2000), "An Evaluation of the Occupational Staffing Pattern and Wage Rates of State Government Workers," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Occupational Employment Statistics (OES) survey, conducted by the U.S. Bureau of Labor Statistics, is an annual survey of 400,000 nonagricultural business establishments. The survey is designed to collect information about occupational employment and wages. Estimates of occupational employment and

wages are produced using three years of data. In order to reduce the burden to respondents most establishments are only once during a three year time period. State government establishments, however, are included in the sample every year. Some of the survey's data collectors have conjectured that the occupational staffing patterns of state government workers are stable from year to year, and that the year to year changes in the occupational wage rates are explained by the rate of change in wages developed by the Bureau's Employment Cost Index survey. If both of these conjectures are true, then we do not need to collect these data each year. A variation of the chi-square statistic and the Bonferroni multiple comparison test will be used to evaluate the year to year change in occupational staffing patterns and wage rates, respectively, of state government workers. The evaluation will be based on 1996, 1997, and 1998 OES state government data.

Cho, Eungchun, **Cho, Moon Jung**, and **Eltinge, John**, (2003), "Variance of a Sample Variance in a Finite-Population Sampling," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

A simple, direct derivation of the randomization variance of the sample variance and related formulae are presented. Examples of the special case in which the populations consist of finite arithmetic sequences are given. A link to the elaborate algebra of "polykays" developed by Tukey to describe relationships between the moments of a finite population and associated moments of simple random samples is also given.

Cho, Moon Jung, (2003), "Inferential Methods to Identify Possible Interviewer Fraud Using Leading Digit Preference Patterns and Design Effect Matrices," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Interviewer fraud can damage the data quality severely. How can we detect it? We use the leading digits to detect the curbstoning in this paper. The effect of the sampling design, such as stratification and clustering, on standard Pearson chi-squared test statistics for goodness of fit is investigated. Rao and Scott (1981) suggested that a simple correction to a chi-squared test statistic which requires only the knowledge of variance estimates for individual cells in the goodness-of-fit problem would be satisfactory. This paper extends the Rao-Scott methods and considers inference for a large number of proportion vectors and optimum allocation of resources (re-interview time). The eigenvalues of design effect matrix are used to obtain related diagnostics regarding the efficiency of a given complex design. For cases with heterogeneous eigenvalues, the eigenvalues and eigenvectors of design effect matrix are used to identify specific linear functions of proportion to which a test is sensitive.

Cho, Moon Jung and Eltinge, John, (2001), "Diagnostics for Evaluation of Superpopulation Models for Variance Estimation Under Systematic Sampling," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Under systematic sampling with multiple random starts, one may use variance estimators based, respectively, on (1) a relatively simple design based approach; or (2) specific superpopulation models. Variance estimators derived from (1) generally will be approximately design unbiased, but may be somewhat unstable if the number of random starts is small or moderate. In addition, the performance of estimators based on (2) will depend on the extent to which the underlying finite population is consistent with the assumed superpopulation model. This paper considers diagnostics for the comparison of estimators from (1) and (2), with special emphasis on (a) exploratory analysis of the underlying finite population; (b) variance estimator bias; (c) variance estimator stability; and (d) coverage rates and widths of associated confidence intervals. Some of the proposed methods are applied to sample data from the U.S. Bureau of Labor Statistics.

Cho, Moon Jung, Eltinge, John, Gershunskaya, Julie and **Huff, Larry**, (2002), "Evaluation of Generalized Variance Function Estimators for the U.S. Current Employment Survey," *Proceedings of the Section Survey Research Methods,* American Statistical Association.

In applied work with generalized variance function models for sample survey data, one generally seeks to develop and validate a model that is relatively parsimonious and that produces variance estimators that are approximately unbiased and relatively stable. This development and validation work often begins with regression of initial variance estimators (computed through standard design-based methods) on one or more candidate explanatory variables. Evaluation of initial modeling results is often complicated by correlation among the initial variance estimators. This paper considers ways in which to address this problem, with principal emphasis on three issues: 1.) approximate conditional or unconditional independence of subsets of initial variance estimators; 2.) use of 1.) and additional conditions to evaluate the properties of the estimators of the coefficients of a generalized variance function model; and 3.) evaluation of the stability of the resulting variance estimators. Some of the proposed diagnostics are applied to data from the U.S. Current Employment Survey.

Clark, Kelly and Phillips, Mary Anne (2002), "A Comparison of Job Openings Surveys: Concepts and Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

Responding to the need for statistics that measure the demand for labor as a complement to measures of labor supply such as the unemployment rate, surveys in both the United States and Europe have been developed recently to measure the

number of job openings. In the United States, the Bureau of Labor Statistics (BLS) has developed an establishment survey designed to produce new, nationally representative data on labor demand by measuring both the number and rate of job openings in the United States. This paper begins by briefly describing the new Job Openings and Labor Turnover Survey (JOLTS) and its development based on lessons learned from previous BLS surveys and pilot studies. Simultaneously, the European Union is currently embarking on an effort to measure job openings in member countries. This paper also describes that recent effort and how it builds on the experience of several individual countries that already collect job openings data. The overall comparison focuses on definitions, reference periods, and sample design. Complement to measures of labor supply such as the unemployment rate, surveys in both the United States and Europe have been developed recently to measure the number of job openings. In the United States, the Bureau of Labor Statistics (BLS) has developed an establishment survey designed to produce new, nationally representative data on labor demand by measuring both the number and rate of job openings in the United States. This paper begins by briefly describing the new Job Openings and Labor Turnover Survey (JOLTS) and its development based on lessons learned from previous BLS surveys and pilot studies. Simultaneously, the European Union is currently embarking on an effort to measure job openings in member countries. This paper also describes that recent effort and how it builds on the experience of several individual countries that already collect job openings data. The overall comparison focuses on definitions, reference periods, and sample design.

Clark, Kelly, Phillips, Mary Anne, and Stephens, Brady (2001), "The Job Openings and Labor Turnover Survey: Preparing to Publish Data for a New Survey" *Proceedings of the Section Government Statistics,* American Statistical Association.

Bureau of Labor Statistics (BLS) has developed an establishment survey that produces new, nationally representative data on labor demand and turnover in the United States. Currently, the Job Openings and Labor Turnover Survey (JOLTS) is collecting monthly information on the number of job openings, hires, and separations for selected establishments. This paper briefly describes the JOLTS program and outlines the process leading up to the release of its new data series. The paper is divided into three main sections, the first being a brief background on the development of the survey. The second section will describe how, as part of the survey development, the JOLTS staff researched prior and existing data series that estimate job openings and labor turnover-type information. Each month, JOLTS estimates are tracked against each of these data series to determine if there are recognizable trends and if those trends are as predicted, taking into consideration the differences between JOLTS and each of the other surveys. Part three of the paper will discuss publication issues, including the format, style, availability, and frequency of the official press release.

Clements, Joseph, (2000), "Protecting Data in Two-Way Statistical Tables Using Network Flow Methodology," *Proceedings of the Section on Government Statistics,* American Statistical Association.

Cell suppression is a standard technique used by many federal statistical agencies to protect statistical tables. If a cell in a table reveals too much information about individual respondents' values, then it is suppressed when the table is published. To protect these cells (i.e., primary cells) from being resolved using other information in the table, additional cells, called complementary cells, must also be suppressed before the table is published. For even moderately sized tables, identifying an optimal complementary cell suppression pattern, that is, one that protects all of the primary cells and minimizes the amount of information lost, is a complex numerical problem. Because of this, heuristic techniques have been developed that attempt to minimize the cost function by protecting one primary cell at a time, iterating until all of the primary cells are protected. For two-way tables, heuristic techniques based on network flow formulations are preferred because they provide good results and they can be solved using numerically efficient algorithms such as the network simplex algorithm. In this paper, we examine some of the network flow formulations that have been proposed to solve the complementary cell suppression problem in two-way tables.

Cohen, Stephen, Wright, Tommy, Vangel, Mark, and Wacholder, Sholom, (2000), "Postdoctoral Research Programs in the Federal Statistical System," *Proceedings of the Section on Survey Research Methods,* American Statistical Association. .

Federal agencies use a variety of means to ensure exchange of expertise between the academic environment and the federal statistical system. For example, agencies encourage established researchers to apply for sabbaticals by funding efforts such as the ASA/NSF fellowship programs. The Postdoctoral Research Programs in the federal agencies are designed to attract recent PhD graduates. In either program, researchers gain access to data that is not easily accessible outside the agencies and have the opportunity to do research which enhances both their career development and the federal sector. Program participants work on a variety of applied scientific/policy/data collection problems. The goal is to attract talented people to consider the federal sector for a career and to build ties between the federal statistical community and researchers who ultimately plan careers in the academic sector. Panel participants will discuss the potential benefits of developing these programs, benefits of the program to the participants, different ways of administering postdoctoral research programs, and implications of their effect on the federal statistical system.

Zarate, Alvan, Greenberg, Brian, Bournazian, **Cohen, Stephen** and Eden, Donna (2001), "Privacy, Confidentiality and the Protection of Health Data - A Statistical Perspective," *Proceedings of the Section on Government Statistics,* American Statistical Association.

Topics critical to understanding issues involved in data confidentiality and data access will be discussed. They include: Overview; Legal Issues Relating to the Confidentiality of Health Data - State law, Privacy Act, Freedom of Information Act, HIPAA regulations; Informed Consent in surveys and other research; Statistical Disclosure Limitation Techniques (SDL) for Microdata; Statistical Disclosure Limitation Techniques for Tables; The CDAC Checklist on the Disclosure Potential of Proposed Data Releases; Restricted Access Procedures.

Rips, Lance, **Conrad, Frederick** and **Fricker, Scott,** (2000), "Unraveling the Seam Effect," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Panel surveys sometimes ask respondents for data from several different intervals within a longer reference period. For example, the Survey of Income and Program Participation interviews its respondents every four months, but it asks about each month separately within the four-month interval. Findings from such surveys often show larger changes from one month to the next when the data come from two different interviews than from the same interview. We have studied this seam effect experimentally in a setting that allows us to control the information that respondents should report. The results of the experiments are consistent with a theory in which the seam difference is due to two factors: (a) respondents' forgetting information within the response interval, and (b) their bias in reporting when they can no longer remember correct answers.

Conrad, Frederick, Blair, Johnny and Tracy, Elena (1999). "Verbal Reports are Data! A Theoretical Approach to Cognitive Interviews." *Proceedings of the Federal Committee on Statistical Methodology Research Conference.*

The use of verbal reports to pretest questionnaires (cognitive interviews) is the most tangible outcome to date of the dialogue between cognitive psychology and survey methodology. Cognitive interviews are a standard survey pretesting tool yet they rarely exploit the theory and body of knowledge about verbal report methods. For example it is believed that if people are not aware of a thought process, they cannot verbalize it. In addition, there is evidence that verbalizing certain processes can affect the process being reported. When these are overlooked, it threatens the validity of verbal reports. We have developed a two part technique for collecting and analyzing verbal reports in cognitive interviews that takes this into account. One part concerns collecting verbal reports – the administration of cognitive interviews. The second part concerns the analysis and interpretation the verbal reports. In the data collection part, cognitive interviewers use certain generic probes when a verbalization indicates the respondent is aware, of but has not reported, useful information. In the interpretation and analysis part, coders assign segments of the verbal reports to a problem taxonomy in which a set of problem classes can be occur throughout the stages of the response process. We advocate using an approach like this to increase the validity and objectivity of cognitive interview data. Preliminary data suggests that it is promising.

Conrad, Frederick, Brown, Norman and **Dashen, Monica** (1999). **"**Estimating the Frequency of Events from Unnatural Categories." *Proceedings of the Section on Survey Research Methods,* American Statistical Association. .

"How often do you do light or moderate activities for at least 10 minutes that cause only light sweating or a slight to moderate increase in breathing or heart rate?" This is a hard question because it asks about an event category ("light or moderate activities …") which seems to be at odds with the way most respondents think about events. We argue that the danger in asking respondents about the frequency of such categories – we call them unnatural categories – is that relevant episodes may not come to mind, leading to underreporting. We explored this in two studies. In both, participants studied a list of words and were then asked to estimate the number that were members of either unnatural or natural categories. In the first experiment, the unnatural category group was asked how many of the words they had studied contained particular *properties* (e.g. shiny, smelly, round); the group tested on more natural categories was asked how many words were members of *taxonomic categories* (e.g. furniture, mammals, fruit). Both groups underestimated actual frequency but it was far more extreme for those answering about properties. In the second experiment, both groups were tested on properties but one of these groups, the *instance + property* group, studied the properties along with the individual words (e.g. milk-white) to see if by promoting encoding of the properties we could render them more natural. Response-time patterns indicated that this was the case: *instance-only* estimates took three times as long as *instance + property* estimates and involved far more underestimation of actual frequency. We recommend decomposing unnatural categories into their natural parts and asking separate questions about each.

Conrad, Frederick G. and **Schober, Michael F.** (1999) "Conversational Interviewing and Data Quality." *Proceedings of the Federal Committee on Statistical Methodology Research Conference.*

Standardized interviewing is widely practiced because it promises to reduce interviewer related error and because it is cheap. Yet the technique cannot guarantee uniform understanding of questions and, thus, may reduce data comparability. Conversational interviewing may standardize the meaning of questions by allowing interviewers to clarify survey concepts, but it cannot guarantee uniform interviewer behavior. We discuss four experiments (three in the laboratory where respondents answer on the basis of fictional scenarios allowing us to directly assess accuracy) and one in the field (in which respondents are interviewed twice so that response change and explanation allow us to indirectly assess accuracy). We have found that conversational interviewing improves response accuracy when respondents' circumstances are atypical (e.g. Should a lamp purchase be counted as a furniture purchase?) but requires additional time to clarify concepts. The more ways in which clarification can be provided the more accurate are responses and the longer interviews last. Respondents are not always willing to seek clarification when they need it, though

we were able to overcome this in one experiment by reducing the social and cognitive demands of asking for clarification. Instead of interacting with an interviewer respondents interacted with a computer and clicked a mouse on highlighted text to obtain clarification; when respondents were told that clarification was essential for accurate responding they frequently obtained it. We conclude that researchers have the option of keeping interviews short and risking some misunderstanding or investing more time to be sure questions are understood as intended.

Schober, Michael F., **Conrad, Frederick G.,** and **Fricker, Scott S.** (2000). **"**When and How Should Survey Interviewers Clarify Question Meaning?" *Proceedings of the Section on Survey Research Methods,.* American Statistical Association.

Conversational interviewing, in which interviewers and respondents work together to make sure questions are understood as intended, can help respondents answer more accurately than strictly standardized interviewing. It takes longer and can be costly. Here we measured response accuracy and interview length for three kinds of partially conversational interviewing, which resemble current practice in some organizations. Census Bureau interviewers telephoned paid laboratory respondents, who answered factual questions from ongoing government surveys on the basis of fictional scenarios. Interviewers either (1) read scripted definitions of question concepts when respondents explicitly asked for clarification; (2) used their own words to present official definitions of question concepts when respondents explicitly asked for clarification; (3) presented scripted definitions whenever they deemed it necessary, even if respondents hadn't explicitly requested clarification. For all three partially conversational techniques responses were reliably more accurate than for strictly standardized interviews, and interviews took reliably longer; compared to more fully conversational interviews (in which interviewers could present definitions in their own words whenever they thought this might help) responses were less accurate and interviews shorter. Results suggest that response accuracy improves whenever respondents get help, whether or not they ask for it explicitly; but if it is left only up to respondents to ask for help, they often won't.

Crankshaw, Mark and **Stamas, George,** (2000), "Sample Design in the Job Opening and Labor Turnover Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Job Opening and Labor Turnover survey (JOLTS) is a new program being fielded by the Bureau of Labor Statistics to compile measures of job openings and employee turnover for nonfarm businesses using a sample of 16,000 establishments. The job opening, employee turnover (hires and separations), and employment data are collected on a monthly basis. Sample members will rotate out of the sample after participating for 18 months. Estimates will be produced for broad industry groups and Census region. In this paper we will discuss JOLTS sample design, sample coordination with other BLS surveys, JOLTS estimators,

and issues arising from sample selection using the Standard Industrial Classification (SIC) system as a strata identifier and estimating for North American Industrial Classification System (NAICS) sector. The JOLTS sample was coordinated with other BLS establishment surveys using permanent random numbers to minimize overlaps between surveys. Sample was selected by industry divisions defined on SIC because the conversion from SIC to NAICS was not yet completed, however, the NAICS conversion will be complete before we begin publishing an experimental series.

Dashen, Monica, (2000), "Improving Purchase Recollection," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Retrospective reports about personal events are often collected by a survey long after the event in question. It is well known that people forget personal events over time, which can affect the data quality. Yet little is known about how to improve recollection. Accordingly, two studies explored how well survey respondents were able to remember frequently purchased items and examined ways to improve purchase recollection. In both studies, participants recorded their own purchases for two weeks and then returned, either immediately, one week, two weeks or four weeks later for a recall test of their own purchases. Study 1 was designed to measure the effects of time and salience on memory for purchases. The results indicate that memory deteriorated over time, as indicated by low match rates and high intrusion (or false report) rates. In Study 2, some participants were given memory enhancers to improve their purchase recollection. The match rate data suggest that participants whose memories had been enhanced reported more purchases correctly than did those participants whose memories had not been enhanced. The results of both studies underscore the importance of understanding memory-based response error in surveys, as well as devising and implementing techniques to improve memory.

McConnell, Shelia, **Dillender, Rebecca, Logothetti, Ted** and **Strifas, Sharon** (2001), Correcting the Calendar-related Fluctuations in Average Weekly Hours and Average Hourly Earning Series ," *Proceedings of the Section on Government Statistics*, American Statistical Association.

In 1998, The BLS published findings on the calendar-related fluctuations in the average weekly hours and the average hourly earnings series from the Current Employment Statistics (CES) program. The research revealed that the survey responses conflicted with the conversion process used to convert these reports into weekly equivalents. BLS has since implemented a REGARIMA-based smoothing technique that has eliminated a significant amount of the non-economic volatility in the CES hours and earnings series. While application of the REGARIMA models improves the measurement of the seasonally adjusted over-the-month changes, it does not correct the underlying microdata response errors. To determine if the underlying microdata response errors can be corrected, BLS tested a modified collection instrument that allowed payroll and hours information

for salaried and hourly paid workers to be reported separately. Two methods to convert the hours and payroll data into weekly equivalents also were tested. This paper provides a discussion of the effectiveness of the modified collection instrument and the two data conversion methods in correcting the calendar-related distortions.

Dippo, Cathryn S., Gillman, Daniel W. (1999), The Role of Metadata in Statistics," *Paper presented at UN/ECE Work Session on Statistical Metadata . Metadata.*

The term is used by many people speaking about many different things. Just what does the term mean with respect to official statistics? While the dictionary definition "data about data" is concise and accurate, it lacks the specifics and context needed to communicate meaning. So, a few years ago, we developed the following definition:" Statistical Metadata is descriptive information or documentation about statistical data, i.e. microdata, macrodata, or other metadata. Statistical Metadata facilitates sharing, querying, and understanding of statistical data over the lifetime of the data." This definition is also fairly concise and accurate; moreover, it provides some context. But is it sufficient to convey meaning to a diverse set of users such that their comprehension of the term is equivalent? Probably not.. Thus, our goal in this paper is to indicate the breadth of meaning associated with the term metadata in the context of official statistics and the agencies that produce them.

Gregg,Valerie and **Dippo, Cathryn S.** (1999), *"*FedStats: Partnering to Create the National Statistical Information Infrastructure of the 21st Century*," Proceedings of the Section on Government Statistics,* American Statistical Association.

Two years ago, the Office of Management and Budget's Interagency Council for Statistical Policy opened the FedStats WWW site (http://www.fedstats.gov)--a "One-Stop Shop for Federal Statistics". The FedStats Interagency Task Force continues to design and develop new approaches to make it easy for users to find Federal statistical information, without having to know which of the 70+ agencies holds the data and information users seek. The Task Force keeps FedStats a dynamic, award winning website by working together to develop new features. FedStats lessons learned contribute to defining longer-term computer science and information technologies research requirements for realizing the Federal statistical community's vision for a National Statistical Information Infrastructure. To achieve this vision, FedStats agencies are undertaking research partnerships by collaborating with academic computer scientists, social scientists, statisticians; and, our local and state partners. Research partnerships seek to bring the latest technologies to bear on data integration, knowledge discovery, data validation and visualization, data retrieval, storage and archiving, computational intensive statistical modeling, on-line collaborative problem solving environments, and universal access.

Dixon, John, (2001), "Using 'Gross Flows' to Evaluate the Impact of Nonresponse on Federal Household Survey Estimates*," Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Previous studies have found that nonresponse varies for different demographic characteristics of households. The impact of nonresponse on estimates from Federal household surveys (e.g.; The Current Population Survey) will be examined by comparing estimates from one period depending on response or nonresponse from an adjacent period. This technique is a form of "gross flow", which has been used to study potential bias in survey estimates in other studies. The differences in estimates will be examined relative to household characteristics which have been found to differ in nonresponse. This will provide a measure of the degree of bias due to nonresponse as well as a measure of differential bias for subgroups of the populations. Some of the household characteristics of interest include ownership status, rural/urban environment, and demographic characteristics of the household.

Dixon, John, (2002), "Item Nonresponse and Unit Nonresponse in Household Surveys," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Previous studies have examined item nonresponse as a predictor of household nonresponse. The usefulness of item nonresponse and unit nonresponse in examining potential nonresponse bias in surveys is explored. Models of item nonresponse and household nonresponse will be related to survey estimates. The theory that households vary on a continuum of cooperation will be examined through a scale from their unwillingness to respond to the survey at all to their unwillingness to respond to certain items.

Dixon, John, (2000), "The Relationship Between Household Moving, Nonresponse, and the Unemployment Rate in the Current Population Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

In the Current Population Survey, a household survey from which labor force estimates are produced, selected housing units remain in sample during a 16-month period. The households are interviewed during the first 4 and last 4 months of this period. During this time, the household occupying a sample housing unit may change. Matching households between months allows an analysis of the relationship between whether a household moves and estimates of the employment rate. Many households move during the 16 months they are in sample. Since change in employment may be related to the household's decision to move, the estimates of employment status may be affected. "Inmovers" don't completely make up for the number of "outmovers" so their relative effect may not be offset. The differences in response rates can also affect estimates. The current study examines the nature of this relationship through an analysis of the characteristics of movers and the resulting effect on labor force estimates.

Dorfman, Alan and Chambers, Raymond, (2003), "Estimating Finite Population Totals Using Transformations, *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

It can happen, especially in economic surveys, that we are interested in estimating the population mean or total of a variable Y, based on a sample, when a linear model makes sense, not for Y itself, but for a transformation (strictly monotonic function) of Y. We mainly focus on the important case where the transformation is logarithmic. Currently available methods based on the lognormal distribution are reviewed, and two new methods introduced, one based on the idea of "smearing" (Duan 1983.), which do not require the lognormal assumption. Theoretical biases and variances are given, with suggestions for sample design and variance estimation, and a practical measure suggested to reduce sensitivity to deviant points. Estimators are evaluated and compared in an extensive empirical study on four economic populations.

Dorfman, Alan H., Chambers, Raymond and Wang, Suojin (2002), "Are Survey Weights Necessary? The Maximum Likelihood Approach to Sample Survey Inference," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In survey sample inference, there are two fundamental positions that can be taken with respect to randomized designs and the use of design weights (inverses of selection probabilities): (1) both are necessary; (2) neither is necessary. Indeed, neither is necessary, and there are occasions when insistence on their use undermines inference. There are other occasions when, the analyst being at a remove from the sampling process, the selection probabilities are helpful information, which it makes sense to incorporate into the method of inference. Strict maximum likelihood inference (as distinguished from the pseudo-likelihood or weighted distribution likelihood approaches) can suitably incorporate the sample weights. The theory of this is not simple, but the practice usually is. We illustrate these points.

Dorfman, Alan H. (1999), "The Stochastic Approach to Price Indices," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

More than most statistical concepts, price indices seem arbitrary and elusive. Three rather different approaches have been developed for choosing among alternative price index formulae: (1) the test approach, (2) the economic approach, and (3) the "stochastic" approach through statistical models. The stochastic approach was investigated and championed in the last century and beginning of this, most notably by Jevons and Edgeworth, but after criticism by Keynes, and with the rise of the economic approach, the stochastic approach fell by the wayside. It is perhaps not an accident that no article on price indices has appeared in JASA for many decades. To calculate price indexes, data on "the same item" (in most cases, actually a collection of items narrowly defined) is collected across

time periods. Is it feasible to use statistical modeling of such "quasi-longitudinal" data for characterizing price indexes? A state space model of price data is suggested, yielding a consumer price index defined in terms of the parameters of the model. Feasibility of the new index is shown using scanner data for canned tuna. The stochastic approach, brought up to date, brings new insight to the question of price indices.

Dorfman, Alan H. (1999), "Issues in the Analysis of Complex Surveys"," *Proceedings Book2 Topic 67 (Bulletin of the International Statistical Institute)*.

Data from a survey of a population can be used for two goals: (1) description *of* the population (i.e. estimation of such numerical characteristics as the total values of a given variable); (2) inference *from* the population to the process which generated it. Deming (1950, Chap 7) refers, on the one hand, to descriptive (or *enumerative*) studies, and, on the other, to *analytic* studies, which are aimed at "the causes of patterns and variations…". According to Skinner, Holt, &Smith 1989 [henceforth SHS], p.1, "… the analytic use takes us beyond the summary measures embodied in description and into the causal explanation of the processes that underlie the descriptive measures." Thus, analysis is concerned with uncovering the causal relations between things, and the relations between corresponding variables, not only in the population sampled, but in like populations. A derivative, but not entirely coincident, definition of the analysis of surveys, understands the population to be characterized by a statistical model, and the goal to be the estimation of the parameters of the model (SHS, p.5). This paper explores several questions that arise from the tension between these two notions of survey analysis.

Dorfman, Alan H., (2000) "Non- Parametric Regression for Estimating Totals in Finite Populations," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Nonparametric regression provides computationally intensive estimation of unknown finite population quantities. Such estimation can be more efficient than inference tied to design-probabilities (in design-based inference) or to parametric regression models (in model-based inference), and, furthermore, gives deeper insight into traditional methods such as poststratification.. A nonparametric regression based estimator for finite population totals is introduced. Its utility is supported by determination of its asymptotic bias and variance and by simulation studies on real and artificial data.

Dorfman,Alan, Leaver,Sylvia; and **Lent, Janice** (1999). "Some Observations on Price Index Estimators," Statistical Policy Working Paper 29 - Part 2 of 5, pages 56-65.

Price indexes can be divided into two broad classes, the *superlative* indexes, and the non-superlative indexes. Superlative indexes in theory approximations a true cost-of-living index (which it is impossible to directly calculate), and tend in

practice to be very close to each other, so that anyone of them can be taken as representing the class. Non-superlative indexes (which are the ones actually used in practice, because of the timeliness with which their sample based estimates can be produced) deviate in theory from the cost of living index, and in practice from the superlative indexes. Particular forms of-non-superlative indexes tend to lie above the superlative indexes, others below. The (positive or negative) gap between a non-superlative index and the cost of living index [has been characterized as] is its *substitution bias* or *substitution effect.* Implicit is the notion that the indexes are *population indexes*, representing the totality of transactions of a given sort in a given economy. It seems natural to estimate the magnitude of the substitution effect of a given index by retrospectively measuring the distance between it and a superlative index. However, such measurements are necessarily made on *sample estimates* of the corresponding indexes. We show that the relationships among sample-based indexes, and between them and population indexes, are not necessarily a straight forward reflection of the relation between population indexes. Thus estimates of the direction and magnitude of the substitution effect are more imprecise than has previously been supposed.

Wang, Suojin, **Dorfman, Alan H.,** and Chambers, Raymond (1999)"*M*aximum Likelihood Under Informative Sampling*," Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Informative sampling occurs when the probability of inclusion in sample depends on the value of the survey variable. Most sample survey inference methods depend on the method of sampling being uninformative. A general methodology for likelihood based survey inference, based on application of the Missing Information Principle (MIP; Orchard and Woodbury, 1972), which allows for informative sampling is set out in Breckling, et al (1994). Unfortunately, this approach does not specifically model the distribution of the sample data, and so it is not obvious that maximum likelihood based on it leads to exactly the same inference as a more direct approach which builds upon this sampling distribution. Consequently it is of interest to demonstrate the equivalence of the ``MIP-based'' and more direct approaches to maximum likelihood for sample data obtained via an informative sampling scheme. This paper does this in the context of a particular method of sampling, called array sampling. Extensions to the theory are also explored.

Short, Kathleen, **Doyle, Patricia,** Hernandez, Donald, Naifeh, Mary, Johnson, David, **Garner, Thesia** and Rozaklis, Patricia (1999), " Measuring Poverty: Questions, Approaches,and Findings," *Proceedings of the Section on Social Statistics,* American Statistical Association.

Session description: This presentation will summarize new findings described in the first report presenting alternative poverty measures. Released in the spring of 1999, that report will present an array of alternative poverty measures based primarily on the recommendations of the National Academy of Sciences. As a

result of the release of that report, and subsequent discussion and comment, additional research will be underway. Led by an interagency committee formed for this purpose, the Census Bureau and the Bureau of Labor Statistics will continue to refine the proposed measures. Authors will discuss relevant issues, report on work undertaken to address those issues, and outline plans for additional research and data collection for the longer term.

Echols, Carrae, Robertson, Kenneth, and **Tou, Albert** (1999), *"*Developing Estimators for Use with Multiple Years of Grouped Data in the Occupational Employment Statistics Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Occupational Employment Statistics (OES) survey is an employment and wage survey of nonfarm business establishments conducted by the Bureau of Labor Statistics. Occupational employment and wage rate data are collected in a continuous 3-year cycle. The employment of each occupation is reported by coding employees into one of 11 contiguous, non-overlapping wage intervals. Note that the exact wage of individual workers is not reported. In this paper we evaluate alternative estimators used to calculate the mean and median wage rates of each occupation based on multi-year data sets. Exact wage data collected by the Bureau's Compensation and Working Conditions (CWC) surveys are used to evaluate these estimators by grouping the CWC data into wage intervals. Afterwards, alternative mean and median wage rate estimators are used to estimate wage rates for each occupation based on the grouped data. "True" mean and median wage rate values are calculated for each occupation by using the exact wage data reported by the CWC surveys. The wage rate estimates produced by each estimator are then compared to these corresponding "true" wage values.

Eltinge, John (2003), "Variance Estimation and Inference for Stratified Sample Designs with Nontrivial Sample Fractions and Nonresponse," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In establishment surveys, sample designs often include stratification by the size of the establishment, with large establishments included in strata that have nontrivial sample fractions. For these designs, development and evaluation of appropriate variance estimators generally involve trade-offs among several factors, including the following: (1) the intended use of the variance estimator, e.g., for formal inference or for construction of weights in a weighted-least-squares procedure; (2) the extent, if any, to which one must account for nonresponse effects; (3) the relative magnitudes of the effects of standard variance approximation methods, e.g., stratum collapse. (4) the simplicity and ease of implementation within a given agency production environment.

This paper explores issues (1)-(4), with principal emphasis on diagnostics to identify acute problems with variance estimator bias or instability. The paper also

adapts sparse-effect models from the experimental design literature to evaluate some properties of the proposed diagnostics.

Sukasih, Amang and **Eltinge, John** (2001), "A Goodness-of-Fit Test for Response Probability Models in the Analysis of Complex Survey Data," *Proceedings of the Section on Survey Research Methods*

Nonresponse adjustment may be carried out by weighting observations according to the inverse of estimated response probabilities. Estimation of these probabilities always depends to some degree on underlying models, which may use varying degrees of parametric structure. For example, one may use probability estimates based on a relatively simple logistic regression model; or probability estimates computed through weighting cells. This paper discusses a test statistic to compare the probability estimates computed through these two methods. Limiting distributions of the test statistic are first derived under basic assumption of independent and identically distributed observations; and then are extended to observations obtained through a complex sample design. This test is applied to data from the Third National Health and Nutrition Examination Survey (NHANES III).

Yansaneh, Ibrahim, and **Eltinge, John** (2001), "Methods for Comparison of Design Effect Components Across Surveys," *Proceedings of the Section on Survey Research Methods.*

In the study of social, economic and health conditions in developing countries, survey work often uses similar sample designs and questionnaires in several countries. Consequently, design effect components estimated from previous surveys may be useful in the refinement of sample designs for subsequent surveys in other countries. The potential merits and limitations of this idea depend heavily on the extent to which specific design effect components are homogeneous across countries. This paper discusses some methods for testing these homogeneity properties. In some cases, similar methods can be used to test for homogeneity of design effect components across years or subdomains within a single country. The proposed methods are applied to data from the Demographic and Health Surveys (DHS).

Eltinge, John, (2001), "Accounting for Design and Superpopulation Components of Variability in Approximations for Design Effects, Generalized Variance Functions and Related Quantities," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In the analysis of complex sample survey data, we frequently approximate the variances of point estimators through generalized variance functions, average design effects and related quantities. This paper reviews some methods for evaluation of the adequacy of these approximations, with special emphasis on three issues. First, appropriate measures of adequacy can depend on whether the

variance approximation is intended for design of a new procedure, or for more direct inferential use, e.g., construction of confidence intervals from previously collected data. Second, variance function estimators are affected by several sources of uncertainty, e.g., sampling variability, superpopulation variability and model lack of fit. Third, applications in this area frequently are based on the assumption that the variance functions have a considerable amount of common structure across time, subpopulations or survey items. Some of the evaluation methods under consideration are applied to Bureau of Labor Statistics data.

Yansaneh, Ibriham, and **Eltinge, John,** (2001), "Design Effect and Cost Issues for Surveys in Developing Countries," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Operating characteristics of surveys, such as design effects, cost, and response rates, are frequently used for planning purposes by international agencies that support the design and implementation of surveys in developing and transition countries. In this paper we describe a framework for comparing operating characteristics of surveys across different types of surveys and different countries or regions. We will extend the work of Verma and Le (1996) on design effects by considering other surveys and the decomposition of design effects into more detailed components associated with stratification, clustering, and unequal-probability sampling. The decomposition of the other operating characteristics into relevant components will also be discussed.

Park, Inho, and **Eltinge, John,** (2001), "The Effect of Cluster Sampling on the Covariance and Correlation Matrices of Sample Distribution Functions," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

This paper uses a superpopulation to examine the effect of cluster sampling on the covariance and correlation matrix of sample distribution functions. A simplified two-stage sampling design leads to a covariance matrix approximation which is an extension of the well-known formula (e.g.,Cochran, p. 242) for the univariate sample mean. The resulting estimation method is then be applied to a multistage stratified sampling design in an ad hoc way to provide a model-based estimator of the covariance matrix of sample distribution functions and the approximate distribution of sample quantiles. The proposed methods are applied to medical examination data from NHANES III.

Eltinge, John, (2000), "Implications of Model Validation Criteria for the Performance of Small Domain Estimation Methods," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In work with sample survey data, small domains are subpopulations from which one has collected relatively few observations. Due to sample size limitations, customary design-based estimation and inference methods may perform relatively poorly when they are applied directly to data from these small domains.

Consequently, the sample survey literature frequently uses model-based approaches to develop small domain estimation methods, and to evaluate the operating characteristics of these methods. This paper examines the extent to which some customary model validation criteria can be linked with the operating characteristics of the resulting small domain estimators. Special attention is devoted to cases involving moderate lack of fit.

Eltinge, John (1999), "Evaluation and Reduction of Cluster-Level Identification Risk for Public-Use Survey Microdata Files," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Design-based analyses of cluster sample data generally require the use of primary sample unit (PSU) labels, or of closely related replicate weights. However, for cases in which PSUs are identical to counties or groups of counties, inclusion of PSU labels in public-use data can increase the risk that some PSUs will be identified. This is problematic because statistical agency policies frequently prohibit the public release of information that would allow the identification of small or medium sized PSUs. In addition, identification of a given primary unit can increase the risk of identification of constituent sample elements, e.g., persons or establishments. This paper discusses methods for the evaluation and reduction of PSU-level identification risk. Principal emphasis is placed on methods that modify the original PSU label structure, but still permit the computation of approximately design unbiased variance estimators. The proposed methods are applied to data from the U.S. National Health Interview Survey (NHIS).

Heo, Sunyeong and **Eltinge, John** (1999), "The Analysis of Categorical Data from a Complex Sample Survey: Chi-squared Tests for Homogeneity Subject to Misclassification Error*," Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In the analysis of categorical data, if misclassification errors exist, then estimated cell probabilities may biased and standard Pearson chi-squared tests may have inflated true type I error rates. This paper considers methods to evaluate the power of chi-squared tests for homogeneity with complex survey data subject to misclassification errors. Three cases are considered: adjustment with a known misclassification matrix; adjustment with an estimated misclassification matrix; and no misclassification adjustment. The proposed methods are applied to the data from the Dual Frame National Health Interview Survey (NHIS)/Random-Digit-Dialing (RDD) Methodology and Field Test Project.

Lee, Sangrae and **Eltinge, John** (1999), *"*Diagnostics for the Stability of an Estimated Misspecification Effect Matrix *," Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Analyses of sample survey data frequently use quadratic-form test statistics. Applications include comparisions of subpopulation means or proportions,

simultaneous confidence bounds for a parameter vector, or goodness-of-fit chi-squared tests. One approach is to base the quadraticform statistics on the inverse of an associated design-based covariance matrix estimator. This Wald-type approach works well for many applications, but in some cases, it can involve an unstable covariance matrix estimator and thus can inflate type I error rates. An alternative is to use a covariance matrix estimator closely related to first-order Rao-Scott adjustments. This paper develops diagnostics to compare the relative stability of the Wald and first-order Rao-Scott statistics. The diagnostics are based on the eigenvalues of an estimated misspecification effect matrix. Simulation methods are used to develop reference distributions and critical values for these diagnostics. The proposed methods are applied to data from the U.S. Third National Health and Nutrition Examination Survey (NHANES III).

Park, Inho and **Eltinge, John** (1999), "Fitting Complex Survey Data to the Tail of a Parametric Distribution," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

For simple random samples, one sometimes estimates the mean and Standard deviation of a normal population from regression of observed quantiles on the corresponding standard normal quantiles. This paper considers an extension of this idea to data collected through a stratified multistage sample survey. Principal attention is devoted to fitting a parametric model to the tail of an underlying superpopulation distribution. Direct application of ordinary least squares and generalized least squares methods lead to point estimators of the superpopulation parameters; associated variance estimators; and related goodness-of-fit test statistics. Separate consideration is given to methods based on constrained misspecification effect matrices. The proposed methods are applied to medical examination data from the U.S. Third National Health and Nutrition Examination Survey (NHANES III).

Eltinge, John, (2003), "Adoption and Diffusion of Statistical Methods in Survey Sampling with Applications to Small Domain Estimation and Nonresponse Adjustment," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Parsons,Van and **Eltinge, John** (1999) "Stratum Partition, Collapse and Mixing in Construction of Balanced Repeated Replication Variance Estimators," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The National Health Interview Survey (NHIS) uses a stratified multistage sample design, with stratification carried out separately within each state. For some analyses, it is desirable to approximate the true sample design with a simple two-primary-units-per-stratum design. For example, this simplified design allows the use of standard replicate half-sample variance estimators. This paper discusses the use of three methods to construct this two-per-stratum approximation within self-

representing strata. First, stratum partition assigns most or all of the sample units in a stratum to a pair of pseudo-primary sample units. Second, stratum collapse is used to group together selected singleton primary units, as well as residual units from the above mentioned stratum partition. Third, stratum mixing allows one to reduce the effective number of strata without exacerbating customary bias problems induced by stratum collapse. Some state-level analysis work limits the number of replicates, and thus the number of effective strata, due to the computational burden associated with the use of a large number of replicates. This paper closes with a discussion of tradeoffs encountered in use of the above mentioned three methods.

Ernst, Lawrence, (2003), Sample Expansion for Probability Proportional to Size without Replacement Sampling," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

A drawback of probability proportional to size (PPS) without replacement, fixed size sampling, in comparison with equal probability sampling, is the difficulty of expanding the sample with the expanded sample remaining PPS without replacement. In fact, this author is unaware how to perform such a sample expansion for most PPS sampling schemes, such as systematic PPS sampling. However, Tille (1996) has developed a step by step PPS sampling scheme, for which at each step one unit is eliminated until the desired sample size is reached. It is shown in the current paper that sample expansion is very straightforward if Tille's method had been used to select the original sample, provided a record had been kept of the order in which units had been eliminated in selecting the original sample. Even if no such record had been kept, we demonstrate that it is still possible to expand the sample. We also discuss a modification of Tille's method that allows it attain some of the advantages of ordered systematic PPS sampling. Finally, we discuss applications to the National Compensation Survey conducted by the Bureau of Labor Statistics.

Ernst, Lawrence and **Guciardo, Christopher**, (2002), "Allocating Sample to Strata Proportional to Aggregate Measure of Size with Both Upper and Lower Bounds on the Number of Units in Each Stratum," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

For stratified PPS sampling, it is common to allocate the sample among strata proportional to their aggregate measures of size. In some cases, this allocation yields some strata whose sample sizes exceed their frame sizes, in which case the sample sizes for these strata are reduced to their frame sizes, and the remaining sample units are allocated among the remaining strata proportional to their aggregate measures of size. The process is repeated as necessary. An analogous iterative procedure may be used if there are lower bounds on the number of sample units in each stratum and these constraints are violated, but the sample size does not exceed the frame size for any strata. However, as explained in the

paper, when the proportional allocation produces both strata with sample sizes above upper bounds and strata with sizes below lower bounds, an algorithm that sets the sample sizes for all violating strata to their respective bounds does not lead to an optimal allocation. An algorithm is presented that does yield an optimal allocation. An application to the sample selection for the National Compensation Survey is discussed.

Ernst, Lawrence and **Paben, Steven,** (2000), "Maximizing and Minimizing Overlap When Selecting Any Number of Units per Stratum Simultaneously for Two Designs with Different Stratifications," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

A number of procedures have been developed, beginning with the work of Keyfitz, for maximizing or minimizing the overlap of sampling units for two stratified designs. Certain overlap procedures have been developed for use when the two samples may be selected simultaneously. They generally produce a better overlap than procedures developed for sequential selection applications or are computationally more efficient. One of these procedures (Ernst 1996) is applicable when the stratifications for the two designs may be different, but is restricted to one unit per stratum designs. A second procedure (Ernst 1998) has no restrictions on the number of sample units per stratum, but requires that the designs have identical stratifications. In the present paper we present a simultaneous overlap procedure without the restrictions in the two previous papers; that is, there are no restrictions on the number of sample units per stratum, nor is there a requirement that the two designs have identical stratifications. This procedure, like the two previous procedures, requires the solution of a sequence of transportation problems and produces an optimal overlap. Potential applications are discussed.

Ernst, Lawrence R. (2001), "The History and Mathematics of Apportionment of the U.S. House of Representatives," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

We will discuss the history of methods used to apportion the House, from George Washington's very first veto in 1792 to the U.S. Supreme Court case on this subject in 1992. We will also discuss the various mathematical properties--such as the Alabama paradox--of key apportionment methods, including the four methods that have been actually used to apportion the House for at least one census.

Ernst, Lawrence R. (2001), "Retrospective Assignment of Permanent Random Numbers for Ohlsson's Exponential Sampling Overlap Maximization Procedure for Designs with More than One Sample Unit per Stratum," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Until recently, all of the procedures developed for maximizing the overlap of sample units for two or more stratified designs with the samples selected

sequentially, destroyed the independence of sampling from stratum to stratum for all but the first sample selected, unless the stratifications were identical or variable sample sizes were allowed. This independence is needed to guarantee the validity of the usual variance estimation procedures. Ohlsson (1996, 1999, 2000) has developed a simple overlap procedure, applicable to a wide variety of designs, that preserves this independence. This procedure, which he calls exponential sampling, uses transformed permanent random numbers (PRNs) to select each sample. For situations where exponential sampling has not been used to select the first sample, Ohlsson (1996) has developed a method, for one sample unit per stratum designs only, for retrospectively assigning the PRNs after the initial sample has been drawn and then selecting subsequent samples using these PRNs and exponential sampling. In this paper we generalize this result of Ohlsson by demonstrating how to retrospectively assign PRNs for designs of more than one unit per stratum.

Ernst, Lawrence R. (1999), "The Maximization and Minimization of Sample Overlap Problems: A Half Century of Results," *Bulletin of the International Statistical Institute*, Proceedings Tome LVII, Book 2, 293-296.

Many procedures have been developed in the last half century, beginning with Keyfitz's (1951) pioneering work, to maximize or minimize the expected number of units retained in sample when a new sample is selected with selection probabilities that are different than those used to select the initial sample.. In this paper we discuss the properties of more than a dozen overlap procedures. For example, certain procedures are usable only for one sample unit per stratum designs, while other procedures can be used for designs for which there are a large number of sample units per stratum. Some procedures require identical stratifications for the designs being overlapped, while others do not. Some procedures do not work properly if used in two successive redesigns. Certain procedures use linear programming to produce a better overlap at the cost of additional computational complexity. Some recently developed overlap procedures, with mostly desirable properties, can be used only when the samples for the designs being overlapped are selected simultaneously.

Esposito, James L., (1999) "Evaluating the Displaced Worker/Job-Tenure Supplement to the CPS: An Illustration of Multimethod Quality Assessment Research," Paper Presented at the Conference of the Federal Committee on Statistical Methodology.

Displaced workers are persons who have lost or left jobs involuntarily as a result of unfavorable economic events/conditions (e.g., plant closings, insufficient work, downsizing). The main objective of the paper is to illustrate the utility of using multiple methods to evaluate survey questionnaires. Three methods were used: (1) interaction coding (via behavior coding); (2) interviewer debriefing (using a focus group format and rating forms); and (3) respondent debriefing (using follow-up probe questions). An effort is made to demonstrate how the qualitative and

quantitative information provided by these techniques contributed to our "understanding" of how data quality may have been affected by various sources of measurement error.

Evans, Thomas, (2003), "Gains from Concurrent Seasonal Adjustment of Household Employment and Unemployment Data," *Proceedings of the Section on Business and Economics Statistics*, American Statistical Association.

In 2004 the U.S. Bureau of Labor Statistics plans to adopt concurrent seasonal adjustment for national household employment and unemployment data from the Current Population Survey. Gains from concurrent over the present method with projected factors will be presented for 21 series, including the national unemployment rate and series used to derive it. Revision diagnostics for both levels and month-to-month changes are analyzed. Alternative revisions policies with concurrent will be examined, including no revision, revising last month only, and revising all values.

Faberman, Jason, (2002), "Longitudinal Analysis of Economic Impacts: A Case Study of the Rust Belt," *Proceedings of the Section on Business and Economics Statistics*, American Statistical Association.

The "Rust Belt" region of the U.S., located mostly in the Central and Northeastern portions of the country, has for decades been characterized by a strong manufacturing presence. Consequently, the employment dynamics of the region paralleled the stagnant growth associated with this industry. During the 1990's, an economic boom resounded across all sectors of the economy. Within Michigan, Ohio, and Pennsylvania, local labor markets flourished, but the extent to which each grew varied widely.

This paper uses the variation in economic outcomes across metropolitan areas over the 1992-2000 period to study the cross-sectional relations between employment growth, job flows, and establishment characteristics. The study finds that the metropolitan areas with high growth had the high rates of job creation and job destruction, as well as establishments that were larger and younger, on average. Differences in industry composition had a minimal role in these findings.

The BLS Longitudinal Database is constructed of linked establishment microdata derived as a by-product from the ES-202 Program's Unemployment Insurance (UI) system. It is unique in that it covers establishments in all industries at a quarterly frequency. It is essential to a study such as the one above.

Ferguson, Gwyn, Cohen, Stephen H. and Rosen, Richard, (2003), "Survey Response Measurement at the Bureau of Labor Statistics," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Historically, each survey at the Bureau of Labor Statistics has computed and analyzed its own response rates, but there was no systematic comparison of the rates across surveys. In the 1980s, a framework was developed for computing similar response rates across all surveys. Over the last several years, each of the Bureau's surveys has revised their response rate definitions and formulas to conform to the Bureau-wide framework. Using the response rates computed using these definitions and formulas, we have begun analyzing response rates across similar surveys. This paper will present the agency-wide framework and definitions, describe the current status of the agency-wide analysis, and discuss planned enhancements to this effort.

Figueroa, Eric, Davis, Jeanette, Tan, Lucilla, To, Nhien, Reyes-Morales, Sally E., (2003), "Is a User-Friendly Diary More Effective? Findings from a Field Test," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Consumer Expenditure Survey (CE) is a nationwide survey of households conducted by the U.S. Bureau of Labor Statistics and U.S. Bureau of Census to find out how Americans spend their money. The Diary survey, one of two survey instruments in the CE, is intended to collect weekly expenditures of frequently purchased items such as food and beverages, tobacco, personal care products, and nonprescription drugs and supplies. To improve the "user-friendliness" of the existing Diary, feedback was used from survey respondents, field interviewers and program staff to design several alternative diary instruments. Based on the findings from focus groups on these alternatives, CE management selected one of them (the Redesigned Diary) for field testing. The Redesigned Diary field test took place during the last four months of 2002. The primary objective of the field test was to compare data quality and response rates obtained using the Redesigned Diary with those obtained using the current Diary. This paper reports findings, as well as lessons learned, from the field test.

Fisher, Sylvia, Fox, Jean, Rho, Christine, Sangster, Roberta and **Tucker, Clyde** (2003), "A Qualitative Approach to the Study of BLS Establishment Survey Nonresponse," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In response to a perception that nonresponse in establishment surveys is increasing, a team at the U.S. Bureau of Labor Statistics (BLS) examined the scope and causes of nonresponse in four BLS surveys: the Current Employment Statistics Program, the National Compensation Survey, the International Price Program, and the Producer Price Index. This paper describes the qualitative methodology used to collect data, namely, by means of personal interviews and focus groups with national and regional BLS program managers and staff members. Interview and focus group protocols administered to the different staff groups were designed to generate comparable information across surveys. In

addition, the team conducted telephone interviews with a small number of survey respondents, nonrespondents, drop-outs, and intermittent responders to help understand the causes of nonresponse more thoroughly. The paper also describes the implications and limitations associated with qualitative methods for this research.

Fisher, Sylvia K. (2001), "A Clinic-Based Needs Assessment Study of Women Who Partner with Women: The Relationship Between Sexual Orientation/Gender Identity, Health-Seeking Behaviors and Perceived Quality of Care Issues," *Proceedings of the Section on Committee on Gay and Lesbian Concerns in Statistics*, American Statistical Association.

We conducted a needs assessment survey with approximately 700 lesbian, bisexual, and transgendered (LBT) women in the greater metropolitan Washington DC area under the auspices of the Lesbian Services Program (LSP) at Whitman-Walker Clinic to help describe the population served by the LSP catchment area, and to gain a better understanding of the populations; health concerns and needs, care-seeking practices and perceptions of risk and access to care. We collected respondent information about: significant physical and mental health concerns; recent health care and disease prevention practices; decision to come out; to health care providers and perceived effect of this decision on quality of health care; SO/GI related concerns regarding confidentiality issues and selection of health care providers; use of alternative health care providers; perceptions of LSP in the greater LBT community; and preferred types of programming to meet individualized respondent needs. Respondents have significant concerns about the effect of their SO/GI and its impact on their health care or failed to seek health care, partly in response to concerns about institutionalized homophobia from the community.

Fisher, Sylvia K. (2000), "Improving the Quality of Data Reporting in Business Surveys: Discussant Comments," *Proceedings of the International Conference on Establishment Surveys II.*

Discussant comments focus on papers devoted to improving data reporting in business surveys. Willimack, Sudman et al. posit a model delineating what transpires in large business establishments at each stage of the response process. Eldridge et al. Describe cognitive interview procedures conducted by the Office for National Statistics, United Kingdom, to improve business surveys. Babyak et al. map out procedures implemented by Statistic Canada's to pretest their establishment questionnaires. Both papers prescribe useful strategies to ameliorate the effects of barriers to conducting cognitive interviews in establishments. All three papers demonstrate the efficacy of cognitive methods in improving establishment data reporting.

Fisher, Sylvia. K., Ramirez, Carl, Stanley McCarthy, Jaki, and Shimizu, Iris**,** (2000),

"Examining Standardization of Response Rate Measures in Establishment Surveys " *Proceedings of the Council of Professional and Federal Statistics.*

The Interagency Group on Establishment Nonresponse (IGEN) has studied issues associated with unit nonresponse in Federal establishment surveys and published proposals for research in this area. Previous work has documented the types of response rates calculated across a number of establishment surveys conducted by U.S. statistical agencies. In this paper: 1) we review the similarities and differences among those rates; 2) discuss the extent to which there is standardization in the calculation and publication of such rates and the reasons why this is so; and, 3) explore the possibilities, advantages, and disadvantages of fostering greater coordination and standardization across IGEN-member agencies in this regard.

Kominski, Robert, **Frazis, Harley,** Kaye, Kelleen, and McArthur, Edith, (2000), "How do we know what we've got? Looking at Measures of Educational Attainment across the Federal Government," *Proceedings of the Section on Social Statistics,* American Statistical Association.

The Interagency Committee on Measures of Educational Attainment is charged with proposing a standard set of attainment categories. The committee has reviewed the various measures used by federal agencies in their data collections. It has arrived at a consensus on a set of attainment levels for general purpose uses and is proposing further research in an area that is of great interest across the federal agencies. This discussion will review the committee's findings, describe the range of policy needs for data and the proposed future research.

Fricker, Scott and **Dashen, Monica**, (2001), "How Do People Interpret Open-ended Categorical Questions?" Paper Presented at the American Association for Public Opinion Research Conference.

Two studies investigated how people interpret open-ended categorical questions. The observed findings of both studies show that people do in fact misinterpret category titles and that they do so in systematic ways. The results of Study 1 indicate that people were most likely to give a false positive when they interpret a category as including items that serve a particular goal. People were more likely to give correct (desired) responses when they thought in terms of varieties of items literally belonging to a category. Study 2 confirmed these findings. The use of supplemental instructions is recommended as a means to improve data quality in these questions.

Fricker, Scott and Lisa K. Schwartz. (Aug 2001) "Reporting Absences from Home: Results of Cognitive Testing of the American Time Use Survey's Missed Days Summary Question."

The American Time Use Survey (ATUS) measures how people spend their time, what they spend their time doing, and with whom they spend it. The ATUS interview consists of several sections including an update of key CPS variables, a 24-hour time diary, and several summary questions that clarify and/or augment diary reports. The primary objective of this study was to evaluate respondent accuracy in recalling the occurrence and the duration of the trips away from home. The question was designed to help measure systematic bias that might occur in ATUS because the survey only includes "yesterday" interviews, thus missing activities done on trips away from home lasting more than 2 days. Researchers examined the effect of the recall period on reporting accuracy and identified meaningful trip purpose categories. Based on the testing, several recommendations were made including lead-ins to these "missed day" summary questions and the use of a backwards-recall strategy to collect information about multi-trip months. A recommendation was also made that the ATUS restrict the recall period to two months or less to facilitate accurate recall.

Harris-Kojetin, Brian A., and **Fricker, Scott** (1999), " The Influence of Environmental Characteristics on Survey Cooperation: A Comparison of Metropolitan Areas," Paper Presented at 28th Session International Conference on Survey Nonresponse– Portland, OR. .

A request for survey participation takes place within a broad context – a social and economic environment that can vary over time, across societies, or even across different geographic areas within a society. Over these diverse contexts, there may be differing norms and expectations for interacting with strangers and for complying with requests for help. In the context of solicitation for survey participation, the end result may be varying response rates across areas of a country. House and Wolf (1978), Groves and Couper (1998), and others, have demonstrated such variations, particularly between urban and rural areas. In this paper we begin by reviewing the literature from social psychology, sociology, and survey methodology that identifies likely ecological correlates of survey cooperation. Based on these findings, we operationalize indicators of the demographic, social, and economic environment, and build composite indices to reflect the putative social psychological attributes of metropolitan areas in the United States. We then examine the extent to which these indicators are related to differing levels of survey cooperation across areas. As a necessary first step in developing a fuller nonresponse model, we begin by looking at response rate data from just two large-scale, national surveys- one public and one private- the Current Population Survey (CPS) and the Arbitron Radio Diary survey. First, we examine the pattern of cooperation rates for the two surveys to see whether different surveys conducted by different organizations experience similar relative cooperation rates in specific metropolitan areas. Secondly, we construct regression models utilizing response rates from both surveys as dependent variables, and the environmental indicators of the metropolitan areas as predictors. Finally, we discuss the implications of these findings for theories of survey cooperation and for improving data collection procedures.

Garner, Theisa, (2003), "Economic Well-Being Based on Consumer Expenditures and Personal Assessments of Minimum Income and Minimum Spending," *Proceedings of the Section on Social Statistics,* American Statistical Association.

Minimum income and spending data from the SIPP are used to ascertain levels of living using expenditure data from the CE. Thresholds are compared to the out-of-pocket spending by consumers for all items. This approach is new, as earlier studies compared income only to such thresholds. An additional analysis is conducted using the minimum spending thresholds in which actual out-of-pocket spending on food, shelter, clothing and other basic items is considered. Coefficients from SIPP regression-intersection estimations are applied to the characteristics of the CE sample to produce minimum income and minimum spending thresholds. Thresholds are produced for different consumer unit types as well as for different regions of the country, and for owners and renters. Personal assessment thresholds and results are compared to those based on official poverty thresholds and thresholds based on National Academy of Sciences Panel recommendations. Out-of-pocket expenditures serve as the basis for the resource measure rather than income.

Gershunskaya, Julie, Eltinge, John and **Huff, Larry**, (2002), "Use of Auxiliary Information to Evaluate a Synthetic Estimator," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Bureau of Labor Statistics has considerable interest in estimation of total monthly employment for small domains defined by the intersection of metropolitan statistical area and major industrial division, based on data from the Current Employment Survey (CES). One of several possible elementary estimators is a synthetic estimator based on state-level changes in employment within a major industrial division. It is important to evaluate empirically the magnitude of the bias of this estimator, relative to the magnitude of the standard error of this estimator, and relative to the magnitudes of the biases and standard errors of other candidate-elementary small-domain estimators. This paper studies the extent to which this type of evaluation may be enhanced through the use of auxiliary data from the Covered Employment and Wages (ES-202) Program, a nominal census of employment that provides data several months after production of CES estimates. Principal attention is devoted to evaluation of components of mean-squared error attributable, respectively, to: 1.) lack of fit in the implicit synthetic model; 2.) sampling error in the CES data; and 3.) nonsampling error in the CES data.

Gillman, Daniel, (2003), "Standards and Metadata in a Statistical Agency," *Proceedings of the Section on Social Statistics,* American Statistical Association.

There is a wide range of standards in use today, and the most successful standards over time are developed through a consensus building process. The World Wide

Web Consortium (W3C) and the International Organization for Standardization (ISO) are two well-known examples of standards development organizations (SDOs). Many statistical agencies develop standards, and in fact, some have a standards division within the organization. Two typical statistical standards are the North American Industrial Classification System (NAICS) and the Standard Occupational Classification (SOC). NAICS, SOC, and other code sets (often under ISO, W3C, or other SDOs) are used to describe, classify, or code data that are collected by the agency. So, these standards are also metadata. There are also metadata standards. These address how one organizes or describes data. The Unified Modeling Language (UML) and the eXtensible Markup Language (XML) are two examples. This paper describes the standards- setting process, how standards influence the work of statistical agencies, the connection between standards and metadata, and the need for a standards strategy in the statistical agency.

Gillman, Daniel, (2002), "Decision Criteria For Using Automated Coding In Survey Processing," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Some survey data are classified into pre-specified categories during a process known as coding. If a computer assigns codes without human interaction, then this is called automated coding. Manual coding, computer-assisted manual coding, and interactive coding all require some level of human interaction.

The decision to employ automated coding in survey processing is not simple. There are many options and expenses to consider. Some of these are as follows: 1.) who develops the software; 2.) what is an acceptable error rate; 3.) how will errors be controlled; 4.) what percent of the cases must the automated coder classify; 5.) how much maintenance will a production system require; and 6.) what new resources must be developed to build an automated coder.

These criteria and others are described in this paper. A cost model is developed along with a description of the interactions between the criteria. Finally, some examples are given to show how the model might be employed by a survey organization.

Goldenberg, Karen and **Willimack, Diane**, (2003), "Measurement Differences in Key Economic Indicators," *Proceedings of the Section on Government Statistics*, American Statistical Association.

Economic indicators such as employment, earnings, and work hours may be commonly understood terms, but as collected in U.S. government statistical surveys they take on highly specific meanings. Operational definitions (and sometimes underlying concepts) vary from survey to survey and across statistical agencies, and present different cognitive and recordkeeping demands for respondents. What happens to the "same" measures when they are collected and

reported in different surveys? This paper examines the economic indicator concepts of total employment, production worker employment, work hours, and payroll as they are operationalized in self-administered establishment surveys conducted by the Census Bureau and the Bureau of Labor Statistics. We contrast questionnaires and collection procedures for the key variables across both surveys, and then compare published estimates of the same variables to assess whether they are statistically equivalent. To the extent that there are differences, we consider whether those differences can be attributed to the structure and the content of the survey questionnaires, and we discuss the implications of our research findings.

Goldenberg, Karen, and **Phillips, May Anne,** (2000) "Now that the Study is Over, What have You Told Us? Identifying and Correcting Measurement Error in the Job Openings and Labor Turnover Survey Pilot Test," *Paper Presented at the International Conference on Establishment Surveys II, Buffalo, NY, June 2000 (Proceedings Forthcoming).*

This paper describes qualitative research by the Bureau of Labor Statistics (BLS) to identify sources of measurement error in the Job Openings and Labor Turnover Survey Pilot Test. Over the year of pilot data collection, BLS refined the concepts being studied and modified the data collection form. BLS subsequently conducted in-depth debriefing interviews with a small group of Pilot Test participants and compared their data with intended survey concepts. The results of these interviews were used to identify and correct additional sources of measurement error, and to further refine the data collection forms.

Goldenberg, Karen L., and **Stewart, Jay** (1999), "Earnings Concepts and Data Availability for the Current Employment Statistics Survey: Findings from Cognitive Interviews," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Current Employment Statistics (CES) survey is conducted monthly by the Bureau of Labor Statistics (BLS) for the purpose of estimating month-to-month changes in total payroll employment, and in payroll earnings and hours for production and nonsupervisory workers. BLS is exploring the feasibility of using an earnings concept that is more inclusive than payroll, and collecting earnings and hours data for all employees. In a series of indepth interviews with CES respondents, we explored respondent understanding of key concepts and survey terminology. We debriefed respondents on their current reporting practices and inquired about their information systems. We found that respondents had some ambiguity in their understanding of our basic concepts. In addition, we asked respondents whether they could provide the proposed items. Although almost all respondents answered affirmatively, further inquiries suggested otherwise. By expanding "cognitive" interviews to include detailed information about data systems relative to specific concepts, we learned about data quality and

potentially avoidable sources of measurement error. These findings will be used to better understand the results of a data collection pilot study.

Guciardo, Christopher (2001), "Estimating Variances in the National Compensation Survey Using Balanced Repeated Replication," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The National Compensation Survey (NCS) currently uses Taylor series linearization to estimate variances of mean wages and total employment. Since Taylor series estimators depend on the form (ratio or total) of a parameter, they can become quite complex for certain other NCS parameters, such as pay relatives. Pay relatives are indexes relating locality to national pay: cells are weighted by national employment and hours. Consequently, future NCS variance estimation will be done using Fay's variation of balanced repeated replication (BRR). This paper presents BRR formulas and discusses how variance strata and PSUs are defined for different parameters. For means and totals, BRR is straightforward, except that locality and national estimates require different variance PSUs, variance strata, and Hadamard matrices. BRR is more complex for pay relatives, however, because they are functions of both locality and national parameters. Half-samples are defined using a hybrid Hadamard matrix that references both locality and national variance strata.

Guciardo, Christopher, Ernst, Lawrence, Ponikowski, Chester, and **Tehonica, Jason**, (2002), "Sample Allocation and Selection for the National Compensation Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Recently, the first of five rotating replacement sample panels was selected for the integrated National Compensation Survey (NCS) Program, which includes the Employment Cost Index (ECI), Employee Benefits Survey, and locality wage surveys. This paper explains, for these surveys, how the sample of establishments was allocated to and selected from industry strata and geographic PSUs, and discusses several interesting issues. For these surveys, we describe how the universe of establishments was divided into two groups: one consisting of larger units that will be in the sample for all five sample panels, and the other consisting of units to be selected independently for each panel. Also, we describe how we forced the ECI sample to be a subsample of the parent NCS sample, despite sampling requirements that would not seem to guarantee this outcome. Finally, we describe how we selected the sample from allocation cells that contain clusters rather than single PSUs, a clustering that was necessitated by the fact that for some industry strata the number of sample units allocated to a cluster is smaller than the number of PSUs in the cluster. Several other allocation issues are also discussed.

Horrigan, Michael and Herz, Diane (1999), "A Study in the Process of Planning, Designing and Managing a Survey Program: The case of time-use surveys at the BLS," *Proceedings of the Section on Social Statistics,* American Statistical

Association.

The BLS has recently set up a working group for the purpose of examining the possibility of conducting a time-use survey. This examination follows two significant activities on time use that the BLS has undertaken in the last year: (1) A pilot study of two alternative versions of a time-use survey using a telephone methodology, and (2) co-sponsorship (with the MacArthur Foundation) of the November 1997 conference on time use. Specifically, the working group has been charged with developing a report that addresses the feasibility of conducting a time-use survey using a subsample of the outgoing rotation groups from the monthly Current Population Survey. If viewed as technically feasible, the report will also develop a comprehensive management plan for the design and execution of such a survey. The working group is developing a report that will be delivered to Commissioner Abraham during the summer of 1998. The paper for the ASA session will provide an insider's view to the process of developing a survey strategy as well as a summary of the working group's recommended approach.

Huff, Larry, Eltinge, John and **Gershunskaya, Julie**, (2002), "Exploratory Analysis of Generalized Variance Function Models for the U.S. Current Employment Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

For the Current Employment Statistics Program, approximately unbiased and stable variance estimators are important for the empirical evaluation of standard design-based point estimators, and for production of related small domain estimators. In some cases, standard design-based variance estimators can be relatively unstable, which may lead to consideration of alternative variance estimators based on generalized variance functions. This paper presents an exploratory analysis of generalized variance function models for estimates of total monthly employment with domains determined by the intersection of metropolitan statistical area and major industrial division. Three topics receive principal attention: a.) a detailed description of features of the underlying sample design that are important in variance estimation; b.) graphical evaluation of potential biases in generalized variance function estimators; and c.) omnibus measures of the relative magnitudes of the fixed and random components of model lack of fit.

Izsak, Yoel, Ernst, Lawrence, Ponikowski, Chester, Paben, Steven and Tehonica, Jason, (2003), Redesign of the National Compensation Survey, *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Among the key products produced by the National Compensation Survey (NCS), which is conducted by the Bureau of Labor Statistics, are locality wage surveys. Consequently, a sample redesign of the NCS is necessary for each decade following the release of new metropolitan area definitions. NCS uses a rotating

panel design, with three stages of selection, namely: geographic areas; establishments, and occupations. We discuss some of the sample redesign issues that we are studying, including impact on between area variances of (1) single outside Core Based Statistical Areas (CBSAs) counties versus multi-county clusters outside of CBSAs, (2) micropolitan areas included in strata with metropolitan areas, or with strata of counties outside of CBSAs, or separately, and (3) alternative stratification variables. Additional issues are: number of sample establishments, number of sample areas, and allocation of sample areas among certainty metropolitan areas, noncertainty metropolitan areas, and the remaining areas; use of an overlap maximization procedure to select sample areas; sampling and variance estimation during the transition. Preliminary results are given.

Jacobson , Shawn (1999) ″Relationship Between Data Quality and Collection Date in the Consumer Price Index Housing Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In order to estimate shelter cost inflation, BLS collects Rent data from a sample of renters for the Consumer Price Index. Each rental unit is interviewed every six months; in each interview, the respondent is asked what the rent is in the current month and what the rent was in the previous month. One-month rent relatives are calculated using these reported rent values; six-month rent relatives are computed using current month rent data from preceding interviews. This paper gives the results of an exploratory data analysis based on rent data from 1992 through 1996. The results indicate that the proportion of six-month rent changes increases and the proportion of one-month rent changes decreases as the collection date increases throughout the month. Variances are computed using the stratified random-groups method, to test the significance of these trends. More recent data (from 1997 and 1998) is used to determine if these results hold in later time periods. Possible (psychological) explanations of the phenomena are considered.

Calhoun, Paul**, Jackson, Laura and Wohlford, John** (2001), "Organization and Management of Data Collection," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Atlanta Data Collection Center (DCC) of the BLS handles two BLS surveys. The Current Employment Statistics (CES) DCC opened in 1990 with three interviewers collecting employment data via CATI. CES staff and workload grew to 40 interviewers and over 42,000 monthly reports. The Job Openings and Labor Turnover (JOLTS) DCC, new in 1999, now has 18 interviewers who handle over 6,000 monthly reports.

This paper covers organizational structures where interviewers were assigned similar responsibilities and received individual supervision; placed in competitive teams with similar responsibilities; or put in specialty teams with different responsibilities.

DCC's exhibit life cycles where experience, efficiency and morale develop and flow. Management must adapt to this cycle to maintain maximum productivity. Performance reports and evaluations must match the structure. We have found that a team structure can be used to train and motivate staff, segment work for optimal performance, and improve morale. The life cycle of a DCC suggests a supervisory structure where a new center with new staff benefits from individual supervision while older centers may benefit from a change to teams.

Jain, Raj, (2003), "A Study of the Comparative Performance of the State Space Model-Based Method and the X-12 Arima Method of Seasonal Adjustment,"

Several Bureau of Labor Statistics series, including Unemployment Level of Civilians and the Consumer Price Index of Women's Apparel, are seasonally adjusted using the State Space Model-Based (SSMB) method and the X-12 ARIMA method of seasonal adjustment. The two methods are then evaluated and compared with respect to the quality of seasonal adjustment obtained.

Johnson, William, **Leaver, Sylvia**, **Larson, William** and **Shoemaker, Owen** (2003), "Determining the Optimal Pricing Frequency for Commodities and Services Items and Geographic Areas in the Consumer Price Index," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The U.S. Consumer Price Index (CPI) uses data collected with varying frequencies in order to produce a monthly all items - all U.S. price index. This paper describes research on the effect of changing the frequency with which various items are priced in different geographic areas on the quality of the resulting indexes. This research only examines the Commodities and Services portion of the CPI, which excludes Rent and Owners' Equivalent Rent. Data is used from an area where the CPI currently prices all items every month, to determine the effect of various schemes where some items are priced with less than monthly frequency. Costs of designs are modeled so that the various schemes can be compared with regard to cost versus variance and bias.

Johnson, William, (2002), "Redesigning the Consumer Price Index Area Sample," *Proceedings of the Section on Government Statistics,* American Statistical Association.

This paper describes the PSU selection process for the next CPI Revision. The U. S. Consumer Price Index (CPI) employs a multistage sample design that has been revised every ten years. The first stage consists of selecting primary sampling units (PSUs), which are formed from Metropolitan or Micropolitan Core Based Statistical Areas (CBSAs) based on preliminary definitions by the Office of Management and Budget. The selected PSUs are also used in the Consumer Expenditure Survey (CE) and the Telephone Point of Purchase Survey (TPOPS).

The process to select PSUs includes using variance models to compare various PSU sample designs, stratifying the PSUs, Keyfitzing, and controlled selection.

Johnson, William, (2000), "Exploratory use of Spatial Statistics to Inform the Choice of Variables for Stratifying Non-Self Representing Cities for the Next CPI PSU Sample Selection," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

After data from the 2000 Census becomes available, all non-self representing areas will be stratified prior to selecting the PSU sample for the Consumer Price Index (CPI). This paper discusses the findings from the use of spatial statistics and exploratory spatial data analysis to examine the presence of spatial correlation in 12-month price change for the all items CPI-U during the time period from 1988 through 1999. The models of 12-month price change considered in searching for stratifying variables for the sample design are examined for the presence of spatial lag and spatial error during this period. Earlier work done within the Bureau of Labor Statistics, which led to the selection of stratifying variables based on these models, is discussed in light of what is found.

Johnson,William, **Leaver, Sylvia G.,** and **Benson, Thomas S.** (1999) *"*Modeling the Realized Outlet Sample for the Commodities and Services Component of the U.S. Consumer Price Index," *Proceedings of the Section on Government Statistics,* American Statistical Association.

This paper describes the use of simulations to model the realized outlet sample for the commodities and services component of the U.S. Consumer Price Index as part of the effort to optimize the sample design. Outlet samples for different categories are drawn from independent frames obtained through the Telephone Point of Purchase survey. Outlets may occur in multiple categories or half samples, so the realized sample may have fewer unique outlets than the initial sample size specified. The efforts to model the actual number of unique outlets realized in spite of difficulties arising from the use of several sources of data are documented, and an attempt is made to evaluate the effectiveness of the resulting models. As the sample composition is changing over time from one drawn entirely from frames obtained through the Consumer Point of Purchase Survey to one drawn entirely from frames obtained through the Telephone Point of Purchase Survey an attempt is made to assess how the introduction of frames from this new source affects the realized sample.

Kohli, Martin, (2002), "Wassily Leontief and the Bureau of Labor Statistics, 1941-54:," *Proceedings of the Section on Government Statistics,* American Statistical Association.

A number of different interpretations of the development of input-output analysis exist. Some observers have argued that the Bureau of Labor Statistics merely tabulated data in accordance with Wassily Leontief's theories. This paper

evaluates such claims by examining the classifications and definitions used by Leontief and the Bureau in the early tables. Such an examination leads to the conclusion that the Bureau made a number of significant refinements to Leontief's original framework. The Bureau-modified table proved to be a valuable tool for assessing the accuracy of the GNP aggregates, and this prompted the Budget Bureau to institutionalize the making of input-output tables. Furthermore, the Bureau's work with Leontief had a number of effects on the agency itself. As a result of this work, the Bureau treated some of its price, quantity, and value measurements as part of a new framework; a consistent system of national economic accounts; and this revealed inadequacies in the statistical design of at least one program. To this day, statistical agencies continue to try to develop consistent measures of prices and quantities, using the framework pioneered by Leontief.

Kropf, Jurgen and **Brooks, Nicole**, (2003), NAICS Conversion of Establishment Data: How to Create History," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Current Employment Statistics (CES) Survey, an establishment survey, conducted monthly by the Bureau of Labor Statistics, provides detailed industry data on employment, hours, and earnings of workers on nonfarm payrolls. In June 2003, the basis for industry classification will change from the 1987 Standard Industrial Classification System (SIC) to the new 2002 North American Industry Classification System (NAICS). After the conversion to NAICS, SIC-based data will no longer be produced or published. Historical time series have been reconstructed as part of the NAICS conversion process. The NAICS-based reconstruction effort covers all CES published data types. The conversion techniques used in the historical reconstruction are discussed in detail, along with the inputs to the historical time series and the output--a NAICS-based historical time series. Also covered in this discussion are the specific methods used for exceptions to the regular conversion procedures.

Kropf, Jurgen , Getz, Patricia, and **Roosma , Michael** (1999), *"C*auses and Treatments of Non-Economic Fluctuations in Average Weekly Hours and Average Hourly Earnings Series," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Current Employment Statistics (CES) Survey is conducted monthly by the Bureau of Labor Statistics (BLS) for the purpose of estimating over-the-month changes in total employment, and in payroll earnings and hours for production and non-supervisory workers. In 1997, researchers noted the presence of fluctuations in the Average Weekly Hours (AWH) and Average Hourly Earnings (AHE) series, that appeared to be non-economic in nature and related to variations in the calendar. In this paper, we discuss the process used to identify and treat the observed distortions. First, the microdata were screened by an equal means test to

indicate significant differences in months with varying numbers of days per pay period. These tests identified problematic industries and reporters. Second, we generated estimates without the problematic reports; this resulted in near elimination of the fluctuations in AWH and mitigated the distortions in AHE. Next, we contacted reporters to inquire about their reporting practices and confirmed the source of variation in the payroll figures. Finally, we developed time series models with variables designed to identify, measure, and treat the effects of the varying length of pay periods. The modeling resulted in the successful treatment of affected industries, which now display seasonal adjusted series with virtually no calendar related fluctuations and significantly increased smoothness.

Kujawa, Laura, Stamas, George and **Crankshaw, Mark**, (2002), "Recent Experiences in Survey Coordination and Sample Rotation within Monthly Business Establishment Surveys," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

At the International Conference on Establishment Surveys in 1993, Esbjorn Ohlsson presented mehtods for coordinating sample selection across surveys and within repeated surveys. The approach assigns a permanent random number (PRN) to a unit as it enters the survey frame. The Bureau of Labor Statistics used these methods while redesigning its Occupational Employment Statistics and Current Employment Statistics surveys and developing a new design for the Job Openings and Labor Turnover Statistics survey.

This paper provides examples of what happens when thoretical concepts meet practical application. We discuss the process of assigning PRNs to records as they are added to the frame. We then focus attention on the methodology that was applied to: 1.) update business establishment surveys to capture a representative portion of these new establishments; 2.) update the samples to compensate for units that have gone out of business; and 3.) refresh a portion of the units for sample rotation purposes. Finally, we provide some solutions that were implemented to reduce solicitation cost and overcome other problems resulting from direct application of the theoretical method.

Kydoniefs, Leda and **Stinson, Linda** (1999) "Standing on the Outside, Looking In: Tapping Data Users to Compare and Review Surveys," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

It is not unusual for survey organizations to conduct periodic expert review (sample design, definitions, question wording, data collection procedures) of ongoing survey programs. But the question arises, is there another way to conduct such a review? What if attention was turned toward the individuals who use the data? This paper presents the methodological approaches one could consider in pursuing answers from users to the questions implied by "what do data users

want?" It also considers which approach provides us with the most extensive information about survey users given limited time and resources.

Leaver, Sylvia and **Larson, William**, (2003), "Estimating Components of Variance of Price Change from a Scanner-Based Sample," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

We present estimates of components of variance of price change for an experimental scanner-based Consumer Price Index for cereal for several publication areas. Components of variance for 1-, 6-, and 12-month lags are computed using a weighted restricted maximum likelihood estimation method. Estimates are contrasted among publication areas using two different random effects models, and findings are discussed with respect to approaches to sample design.

Leaver, Sylvia and **Larson, William**, (2002), "Assessing the Impact of Imputation on the Sampling Variance of the U.S. Consumer Price Index," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Imputation of basic-level price change, due primarily to sample attrition and item unavailability, occurs frequently in price index production. Though current variance estimation procedures do take basic-level imputations into account, it is of interest to assess the contribution to total variance that these imputations represent. In this paper, we present estimates of the sampling variance of price change for the U. S. Consumer Price Index which exclude the additional component of variation attributable to basic-level imputation . Variances are presented for one- and twelve-month price change lags, using a stratified random groups methodology. Estimates are contrasted with production variance estimates for the same series.

Leaver, Sylvia and **Larson William** (2001), "Estimating Variances for a Scanner-Based Consumer Price Index," *Proceedings of the Section on Government Statistics,* American Statistical Association.

In this paper we present estimates of the sampling variance of price change for an experimental scanner-based Consumer Price Index for cereal for the New York Consolidated Metropolitan Area. Variances are presented for 1-, 6- and 12-month price change lags, using a stratified jackknife methodology. Estimates are contrasted with variance estimates for the published price index for cereal for the same area. Sources of price change variability are identified and discussed.

Leaver, Sylvia G., Johnson,William, Shoemaker, Owen and **Benson, Thomas S.** (1999) "Sample Redesign for the Introduction of the Telephone Point of Purchase Survey Frames In the Commodities and Services Component of the U.S. Consumer Price Index ," *Proceedings of the Section on Government Statistics,* American Statistical Association.

This paper describes the methodology for the redesign of the sample for the commodity and services component of the U.S. Consumer Price Index to accommodate the introduction of outlet frames from the Telephone Point of Purchase Survey. This work represents a further expansion and revision of models developed for the 1998 CPI sample redesign. Models relating data collection costs and sampling variance of price change to item and outlet selection variables for the sample design were developed and estimated. With these models, data collection resources were allocated to minimize sampling variance of price change, subject to budgetary and operational constraints, using nonlinear programming techniques. Models for sampling variance and costs are given, and solutions to the design problem posed under varying assumptions are discussed.

Lent, Janice, (2002), A Note on the Effects of Extreme Price Values on Price Indexes," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

We examine the effects of extreme price values on the Fisher and Tornqvist index formulas. Using a simple model, we first consider the impact of outliers on the unweighted arithmetic, harmonic, and geometric means of a collection of values. Then, under the same model, we investigate the effect of a single extremely high or low price on the price index formulas (weighted means). Our exploration leads to some general rules regarding the relative robustness of the Fisher and Tornqvist indexes. We illustrate these with empirical results based on airfare data from the Department of Transportation's Origin and Destination Survey.

Lent, Janice, (2000), "Chain Drift in Some Price Index Estimators," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Since the market of available consumer goods is constantly changing, short-term price indexes may be directly measured more accurately than long-term indexes. Thus chained price indexes, which are essentially products of short-term indexes, are often used for estimating long-term changes in the cost of living. Differences between chained and direct indexes have been studied by Forsyth and Fowler (1981), Szulc (1983), and others. Chained arithmetic mean indexes, such as the Laspeyres and Sauerbeck indexes, are known to be vulnerable to upward bias relative to their direct counterparts. This bias due to chaining is known as "chain drift." In this paper, we use analytical and empirical methods (involving data from the US CPI) to examine the effect of chaining on some geometric mean estimators. We find that, for a straightforward Tornqvist index estimator, the effect of chaining is unimportant under reasonable assumptions.

Li, Bogong, (2003), "A Simulation Comparison Between Alternative Outlier Resistant Ratio Estimators," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Model-based predictive approach to finite population estimation provides insights into underlying mechanisms of social or economic phenomena. This feature is drawing the favor of specialists in many scientific fields to this method. Its use does not contradict, in most cases, the design-based estimators while taking full advantage of the theory and empirical results of generalized liner models. However, erroneous outlying observations resulting from the survey data collection may cause greater damage and affect the estimation by the model-based methods in a different way. This study proposes an outlier-resistant model-based estimator by bounding the influence function of possible erroneous outliers on estimated model parameters. It examines and compares the effects of outlying/extreme observations on finite population estimators by alternative estimators assuming the super population model is true.

Li, Bogong, (2002), "Small Area Modeling Research for the Occupational Employment Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Occupational Employment Survey (OES) is a yearly mail survey designed to produce estimates of employment and wages for more than 700 occupations. The OES sample contains approximately 400,000 establishments per year, accumulating a rolling sample of 1.2 million establishments every three years. The OES sample is stratified by geographic area, economic activity, and employment size class--with geographic area defined by State, Metropolitan Statistical Area (MSA) and balance of State area, and economic activity defined by three-digit Standard Industrial Classification (SIC) codes. While the sample is designed to produce reliable design-based estimates for large geographic areas, our research investigates synthetic and composite estimators for smaller geographic areas. We examine some standard small area models, as well as examine new methods which draw upon some of the unique features of the OES survey. We also examine a new method of variance estimation, which modifies the current Jackknife Random Groups approach.

Mason, Charles and Cheng, William, (2003), Linking Retail Establishments Reported in a Household Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Bureau of Labor Statistics' Point of Purchase survey collects information on the retail establishments where consumers purchase the goods and services covered by the Consumer Price Index. Among the data collected are the establishment names and addresses, which are used as a sampling frame for the selection of establishments priced in the CPI. Prior to selecting the sample of establishments, the raw data must be coded and collapsed so that multiple reports of the same establishment are combined. Additionally, address information is refined so that field economists can locate and initiate the selected establishments into the CPI. This paper reports on the recent change by the BLS from a

computer-assisted manual coding and collapsing procedure to a statistical linking processes that utilizes Soundex and additional scoring considerations.

Mason, Charles, Sangster, Roberta and Wirth, Cassandra (2001), "Comparison of Final Disposition Codes Used for Attrition Calculations for Telephone Samples," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Telephone Point of Purchase Survey (TPOPS) is one of the surveys used by the Bureau of Labor Statistics to create the Consumer Price Index (CPI). The data is used to create the establishment frame for the pricing of goods and services for the CPI. TPOPS is conducted quarterly over a four-year cycle. A sample for each panel is drawn via random digit dialing. The quarterly sample size is approximately 24,000 households, of which, 25 percent is new RDD sample.

This study focuses on the data used to compute an attrition rate for the TPOPS. The concern is that many decisions occur before a final case disposition code is assigned (e.g., Completed Interview, Ineligible). Temporary call dispositions are used that consist of all major work actions taken on each case. This includes interviewer actions, supervisor's actions, and programs set up in the instrument (e.g., call attempt limits). Ultimately all of the issues are resolved into a final disposition code. It is unclear how this process effects the attrition rate. We calculate the attrition rate based on the TPOPS final disposition codes. The limits of the final codes are examined and a revised rate is calculated.

Mason, Charles, (2000), "Results from a Random Digit Dialing Survey of Where Consumers Purchase Goods and Services," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics conducts a large household survey, the Point of Purchase Survey, which is used as the sampling frame for the selection of retail establishments priced in the computation of the Consumer Price Index (CPI). Within each of these outlets, BLS Field agents select unique items for repeated pricing from a category of goods and services independently selected from the expenditure data reported in the Consumer Expenditure Survey. In 1998, BLS converted the Point of Purchase Survey from a "personal visit" household interview survey based on a decennial Census address sample to a "computer assisted telephone interview" survey using a list assisted random digit dialing sample. This paper describes the new survey and the flexibility it provides the BLS in updating our CPI outlets. In addition it reports on the first production data results from the new telephone based Point of Purchase Survey (TPOPS) and examines the response rates and data quality; especially the outlet address information, which is especially important in this survey since it is our only means of locating the outlets. In addition, comparisons of the number of retail outlet reported between TPOPS and the older POPS survey are examined.

Mason, Charles, Sangster, Roberta and Wirth, Cassansdra (2001), "Comparison of Final Disposition Codes Used for Attrition Calculations for Telephone Samples," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Telephone Point of Purchase Survey (TPOPS) is one of the surveys used by the Bureau of Labor Statistics to create the Consumer Price Index (CPI). The data is used to create the establishment frame for the pricing of goods and services for the CPI. TPOPS is conducted quarterly over a four-year cycle. A sample for each panel is drawn via random digit dialing. The quarterly sample size is approximately 24,000 households, of which, 25 percent is new RDD sample. This study focuses on the data used to compute an attrition rate for the TPOPS. The concern is that many decisions occur before a final case disposition code is assigned (e.g., Completed Interview, Ineligible). Temporary call dispositions are used that consist of all major work actions taken on each case. This includes interviewer actions, supervisor's actions, and programs set up in the instrument (e.g., call attempt limits). Ultimately all of the issues are resolved into a final disposition code. It is unclear how this process effects the attrition rate. We calculate the attrition rate based on the TPOPS final disposition codes. The limits of the final codes are examined and a revised rate is calculated.

McConnell, Sheila, and **Goodman, William,** (1999), Recognition of More than Possible Trend in Time Series: Redesigned Screening of Microdata in the Current Employment Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

In June of 1995, The Bureau of Labor Statistics (BLS) announced plans to redesign the method of selecting establishments to be included in the sample of the Current Employment Statistics (CES) survey. The newly designed sample is random, unlike the current sample. Partially in response to the recent sample design, new methods to check the quality of reported data are being developed. The new data editing and screening system will include basic checks for internal consistency of records and will also check data against a variety of plausible patterns in reported data over time. In the latter type of test, an establishment's current reported values are compared to earlier data from the same establishment. The current values are screened using several tests. Passing just one of the tests results in acceptance of the data. This paper presents the editing and screening methods currently being tested and their performance.

Mikkelson, **Gordon**, **Allard**, **Mary**, Unger, Linda and Read, Glen (2003), "Implementing Improved Microdata Wage Editing on Universe Data at BLS," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Covered Employment and Wages Program, also known as the ES-202 program, is a cooperative program between the Bureau of Labor Statistics (BLS)

and the State Employment Security Agencies. The ES-202 program produces a comprehensive tabulation of employment and wage information for workers covered by state unemployment insurance laws and for federal workers covered by the Unemployment Compensation for Federal Employees program. Additional private and public sector activities are implemented or adjusted based on the wages from the ES-202 program. States and BLS ensure the accuracy and reliability of the wages using one primary wage edit. Staff use the edit results to review, research, correct, and explain suspect data. Recently, a new wage editing initiative was undertaken to improve the existing edit to better target suspect data. Two major components of the edits were examined: those records without sufficient historical data to properly function and those where adequate data exist but either were flagged without cause or were bypassed when they should have flagged. Several editing and screening options were empirically tested, and the results are presented.

Miller, Stephen, (2003), "A Look at Alternative Estimators for the Current Population Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

The Current Population Survey is a stratified multistage national household survey which, in addition to probability weighting and poststratification, employs an AK-composite estimator as part of the estimation process. One of the motivations behind AK-composite estimation is to improve estimates of monthly change while retaining good estimates of monthly level. We explore modifications to the AK-composite estimator which seek to further improve estimates of monthly change, by directly smoothing components of month-to-month movements. Other estimation methods are also examined which attempt to directly smooth estimates of month-to-month change in levels.

Mueller, Kirk, Manning, Christopher, Kropf, Jurgen, and Scott, Stuart, (2002), "Concurrent Seasonal Adjustment for Industry Employment Statistics," *Proceedings of the Section on Business and Economic Statistics Section,* American Statistical Association.

The Current Employment Statistics (CES) Survey, conducted monthly by the Bureau of Labor Statistics, obtains payroll employment, hours, and earnings from business establishments and produces industry-based estimates. Seasonally adjusted month-to-month changes in the CES national estimates are among the most widely watched economic indicators for both public and private sector policy makers. Accurate seasonal adjustment is an important component in the overall accuracy of these monthly data. The CES program will convert to NAICS industry coding in 2003. In the course of this conversion, concurrent seasonal adjustment is being considered to replace the traditional seasonal adjustment method which applies forecasted seasonal factors. This paper discusses data to be

available under the NAICS system, compares results from the two adjustment methods and discusses the implications of concurrent adjustment for users.

Mueller, Charlotte Balazik, Matthew, and Crankshaw, Mark, (2001), "Ensuring Data Quality in the Job Openings and Labor Turnover Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics (BLS) is producing a new survey collecting job openings and labor turnover data. The Job Openings and Labor Turnover Survey (JOLTS) collects data and produces estimates for job openings, hires, and separations. This survey is expected to provide valuable statistics to help evaluate the U.S. economy. The JOLTS data will help develop a complete picture of the economy when used in conjunction with the monthly employment level and the unemployment rate. In an effort to improve data quality and the resulting estimates, BLS has been editing and reviewing data starting with March 2000, the survey's first month of collected data. As with any new program, JOLTS has had to start from scratch to determine data edits and microdata review and how to handle all the output from these procedures. The edits and parameters were determined based on the nature of the JOLTS data, procedures of other surveys, and those of a JOLTS pilot study. Methodology for outlier detection and management was selected and implemented. Microdata review is conducted each month to identify problem records and correct data errors before final estimates are run.

Mueller, Charlotte, and Phillips, Mary Ann (2000), "The Genesis of an Establishment Survey: Research and Development for the Job Openings and Labor Turnover Survey at the Bureau of Labor Statistics," *Proceedings of the Section on Government Statistics,* American Statistical Association.

In late 2001, a new establishment survey will be making its debut at the Bureau of Labor Statistics (BLS). This survey--the Job Openings and Labor Turnover Survey (JOLTS)--will fill a current void in national economic data. For many years, the unemployment rate has served as an economic indicator of the unused labor supply. JOLTS intends to serve as a parallel indicator of unfilled labor demand. In the short time since 1999, JOLTS has been transformed from a concept to a live survey of 16,000 establishments across the United States. Monthly publication of rates and numbers of job openings, hires, and separations is the goal. This paper profiles the many research and development steps taken for creating this new survey. Looking outside BLS, we studied existing data sources similar in nature, and contacted academics for comments. Internally, we studied other BLS surveys and defined terms and data elements, established reference periods, and designed forms and interviewer scripts. Developmental testing included a pilot test and reinterview surveys.

Paben, Steven P., Elmore, Daniel, Ernst, Lawrence, Lettau, Michael, Ponikowski, Chester, Buszuwski, Mason, Lowell (2003), "Imputation of Benefit-Related

Data for the National Compensation Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

This study examined imputation methods for benefits data collected as part of the integrated National Compensation Survey (NCS) program, which is conducted by the Bureau of Labor Statistics (BLS). Among the components of NCS are the Employment Cost Index (ECI), which measures changes in wages and benefits via a Laspyres index; the Employer Cost for Employee Compensation (ECEC), which measures mean employer costs for wages and benefits; and the Employee Benefits Survey (EBS), which studies the incidence and plan provisions of employer-provided benefits. The integration of these surveys will allow NCS to link employer benefit costs to benefit incidence and plan provisions and calculate new measures, such as cost per participant. As part of this paper, we revisit the imputation methods used prior to the integration; evaluate their use as part of NCS; and introduce methods for handling item nonresponse for the new measures. We suggest regression methods for handling all missing cost related data and the nearest neighbor within-cell hot deck method for handling all missing incidence and provisions data.

Paben, Steven P., (2001), "The Effect of Some Design and Estimation Issues on the Variance Estimates of the Employment Cost Index," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Employment Cost Index (ECI) is a measure of the change in the cost of labor, free from the influence of employment shifts among occupations and industries. The calculation of the quarterly change in the ECI involves the multiplication of the previous quarter's cost weight in each industry-major occupational group cell by the estimate of the quarterly change to obtain a current cost weight for the cell. The cost weights themselves are therefore variable, and the variability generally is an increasing function of the number of quarters from the base period. The impact on the variance of the variability of the cost weights is examined. The variance of an alternative calculation of the ECI that is less affected by the variability of cost weights is also examined. Additionally, the ECI is currently in transition from a national-based sample to a geographic area-based sample. The impact on the variance of this design change, which adds an additional level of sampling while increasing the number of sample establishments is studied. Finally, the effect of using Fay's method of BRR rather than the standard BRR for operational advantages is investigated.

Paben, Steven P., (1999), " Comparison of Variance Estimation Methods for the National Compensation Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

A key use of the National Compensation Survey (NCS) is for the estimation of hourly mean wages for different localities. In NCS, hourly mean wages are calculated as the ratio of total annual pay to the total number of annual hours

worked. In this study, an artificial Metropolitan Statistical Area (MSA) was created by combining NCS data from 16 different localities to serve as a sampling frame for 100 simulated samples. Then, the variance of the 100 sample estimates was compared to variance estimates obtained with the linearized Taylor Series method of variance estimation and three different methods of replication, balanced repeated replication (BRR), a Jackknife method, and Fay's method. To compare the variance estimates of the different methods a number of evaluative statistics were calculated, including bias, stability, and 95% confidence interval coverage.

Paulin, Geoffrey *(*1999*), "*Applied Analyses of Multiply Imputed Income Data from the Consumer Expenditure Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

As is the case with many household surveys, income data collected in the Consumer Expenditure Survey are subject to item nonresponse. The current method for addressing this problem involves dividing participating families into "complete" and "incomplete" reporters. However, "complete" reporters need not provide a full accounting of income to qualify; in most cases, providing of data on one major source of income is sufficient. Multiple imputation has been proposed as a solution for item nonresponse to income questions in the Consumer Expenditure Survey. The Division of Consumer Expenditure Surveys (Bureau of Labor Statistics) has sponsored much research into the feasibility of developing a model-based imputation method to adjust for this problem. Experimental data have recently become available to internal researchers at the Bureau. How might these data be used in practice? And how, if at all, will results change when imputed data are used instead of "completely" reported data? This paper incorporates several types of analysis using these data and reports on the results.

Paulin, Geoffrey (1999) " Applied Analyses of Multiply Imputed Income Data from the Consumer Expenditure Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

As is the case with many household surveys, income data collected in the Consumer Expenditure Survey are subject to item nonresponse. The current method for addressing this problem involves dividing participating families into "complete" and "incomplete" reporters. However, "complete" reporters need not provide a full accounting of income to qualify; in most cases, providing of data on one major source of income is sufficient. Multiple imputation has been proposed as a solution for item nonresponse to income questions in the Consumer Expenditure Survey. The Division of Consumer Expenditure Surveys (Bureau of Labor Statistics) has sponsored much research into the feasibility of developing a model-based imputation method to adjust for this problem. Experimental data have recently become available to internal researchers at the Bureau. How might these data be used in practice? And how, if at all, will results change when

imputed data are used instead of "completely" reported data? This paper incorporates several types of analysis using these data and reports on the results.

Paulin, Geoffrey and **Vannett, Valerie** (1999), *"*Enhancing the Product Line: Issues in Publishing Variances in the Consumer Expenditure Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Division of Consumer Expenditure Surveys (DCES) is exploring new ways to improve data quality. One aspect of this has been the development of selected descriptive statistics for the average annual expenditures that are currently published. Tables have been developed that contain the mean expenditures, variance, standard error of the means, and coefficients of variation on an annual basis by demographic characteristics. The first section of this paper discusses the methods and issues surrounding the development of these descriptive statistics. The remaining portion of this paper details the uses of these statistics both within and outside the DCES. Within the DCES they will be used for quality control and as a research tool. At the same time, the public will have greater statistical information for the comparison of expenditures by demographic characteristics. Tables with descriptive statistics will be available upon request, and an example of a SAS program to produce these statistics will be added to the documentation for microdata available on CD-ROM.

Schwartz, Lisa K., Lynn, Siri, and Gortman, Jayme. (2001) "What's Work? Respondents' Interpretation of Work-Related Summary Questions."

The American Time Use Survey (ATUS) measures how people spend their time, what they spend their time doing, and with whom they spend it. The interview consists of several parts including an update of key CPS variables, a 24-hour time diary, and several summary questions that clarify and/or augment diary reports. This paper focuses on the development of effective employment probes and summary question to clarify which activities reported in the diary should be coded as work activities. Testing specifically focused on participants who work from nonstandard work environments, who work non-traditional work hours, or who hold more than one job. Based on the findings, it was recommended that during the time diary, interviewers should selectively probe for activities that were done for work but may not be included as work by the respondent. Researchers also recommended new wording of the paid work summary questions and recommended the development of a help screen to address concerns about confidentiality and perceived intrusiveness. The authors also recommended clearer definitions for several of the employment concepts.

Schwartz, Lisa K. (2001) "Minding the Children: Understanding how recall and conceptual interpretations influence responses to a time-use summary questions."

The American Time Use Survey (ATUS) measures how people spend their time, what they spend their time doing, and with who they spend it. The interview consists of several parts including an update of key CPS variables, a 24-hour time

diary, and several summary questions to clarify and/or augment diary reports. This paper focuses on the development of the ATUS secondary child care questions and the accuracy of the secondary child care measure. Two cognitive studies were conducted. Study One's objective was to clarify the concepts of secondary child care and determine the optimal child care question wording. Study Two's objective was to determine the extent to which recall and conceptual issues influenced responses to a secondary childcare summary question. Based on the results of both cognitive tests, the current wording of the summary question was deemed effective and recommended for implementation in ATUS. The authors also recommend that after the first year of production, BLS may wish to examine the relationship between the answer to "who was in the room with you? /who accompanied you?" and estimates of secondary childcare (times the respondent reported having a child under 13 "in [his/her] care."

Schwartz, Lisa K. and Fricker, Scott. (2000) "When Out of Sight Isn't Out of Mind: Measuring Passive Child Care."

The American Time Use Survey (ATUS) measures how people spend their time, what they spend their time doing, and with whom they spend it. The ATUS interview consists of several sections including an update of key Current Population Survey (CPS) variables, a 24-hour time diary, and several summary questions that clarify and/or augment the diary reports. This paper primarily focuses on the development of summary questions designed to measure secondary child care. Child care question testing examined two different question wordings to determine which wording yielded a more accurate secondary child care measure. The term "in your care" was chosen to denote secondary child care activities. It was also recommended that secondary child care measures be bounded by the time when parents and children are awake (except for children's napping time).   While not part of the title, the study also included a report on employment question testing. The questions focused on clarifying which activities in the diary might be accurate reports of work activities but not appear as such to coders, such as self-employment activities and work done outside of the regular work hours. The authors recommended that the ATUS ask a specific question to clarify activities done for one's job or business.

Shoemaker, Owen, (2003), Estimation and Comparison of Chained CPI-U Standard Errors With Regular CPI-U Results (2000-2001), ," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In 2001 the BLS calculated and published the first C-CPI-U set of indexes for the 12 months of 2000. This new C-CPI-U (Chained Consumer Price Index - Urban) is calculated and published every year, with a one year lag, using a Tornqvist formula, and can thus be labelled a "superlative" index. A set of weights are updated yearly, so that a unique set of monthly weights are available for both time t as well as for time t-n. We briefly outline the Tornqvist formula and then the methodology for estimating a set of standard errors for these new chained

indexes. We then compare, over the 24-month period of Jan 2000 through Dec 2001, these new superlative index results and their standard errors with the regularly published CPI results and their published standard errors.

Pfeffermann, Danny and **Sverchkov, Michael,** (2003), "Small Area Estimation under Informative Sampling Based on the Sample Distribution," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

It is now generally accepted that small area estimation should be model-based. Models and estimators considered so far assume either that all the areas are represented in the sample or that the sampled areas are selected by simple random sampling. Very few studies consider the case where the sampling of units within the selected areas is informative in the sense that the selection probabilities are related to the values of the target response variables even when conditioning on the model covariates. We will consider situations where the selection of the areas to the sample is with probabilities that are related to the true area means (informative selection of areas) and the sampling within the selected areas is also informative. In order to obtain valid estimates in this case, we will approximate the model holding for the sample data by accounting for the effect of the sampling process and apply standard inference methods to the sample model. Empirical results illustrating the biases that one could encounter when ignoring the sampling effects in the inference process, and the performance of estimators obtained under the proposed approach will be presented.

Sverchkov, Michael, and Pfeffermann, Danny**,** (2000), " Prediction of Finite Population Totals Under Informative Sampling Utilizing the Sample Distribution," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The sample distribution is the distribution of the response variable for units included in the sample. This distribution is different from the population distribution if the sample selection probabilities are related to the values of the response variable. Previous studies of the use of this distribution focus on inference on the parameters and shape of the population distribution. In this article we study the use of the sample distribution for the prediction of finite population totals under single-stage sampling. It is assumed that the population response variable values (hereafter the y-values) are random realizations from some distribution that conditions on known values of auxiliary variables (hereafter the x-values). Classical examples are linear regression and logistic regression models. The problem considered is the prediction of the population total Y of the response variable values based on the sample y-values, the sampling weights for units in the sample and the population values of the x-values. The use of the sample distribution permits the conditioning on all these values and the prediction of Y is equivalent therefore to the prediction of the y-values for units outside the sample. The prediction of the non-sampled y-values is achieved by

approximating the conditional expectation of the y-values given the xvalues for units outside the sample as a function of the conditional sample expectation (the expectation under the sample distribution) and the sampling weights. Several parametric and non-parametric procedures are considered for the estimation of the unknown model parameters required for the prediction process. Appropriate estimates for the MSE of the prediction errors are derived. Empirical results comparing the performance of the proposed predictors with other known methods are presented.

Pfeffermann, Danny, Tiller, Richard, and **Zimmerman, Tamara,** (2000), " Accounting for Sampling Error Autocorrelations Towards Signal Extraction from Models with Sampling Error," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association.

The time series approach to modeling periodic survey data requires information on sampling error autocorrelations (SEA), which, however, are not routinely produced by statistical agencies. Pfeffermann, Feder and Signorelli (1998) have proposed a simple way of estimating SEA from separate panel estimates. This paper examines the properties of this method in more detail and applies it to employment and unemployment series collected by the Current Population Survey (CPS) for 53 States and areas. We present the results of a simulation study designed to assess the performance of the SEA and three alternative variance estimators. On every pleasing result is that the SEA estimators computed from the separate panel estimates have smaller MSE than those computed from the true sampling errors. For variance estimation, the Bootstrap performs best for series of short length and is at least as good as the other methods for long series. To further improve the efficiency of the SEA estimators we borrow information across States using an iterative disjoint clustering algorithm. This algorithm groups States into 2 or 3 clusters. Pooling the estimated State SEA within these clusters results in substantial MSE reductions without introducing significant biases. Finally, we demonstrate how to fit relatively parsimonious ARMA models to the SEA, which is very useful for signal extraction applications.

Polivka, Anne, (2003), "Methodologies for Maintaining Data Comparability for the Current Population Survey: One Year into the Implementation of the 1997 Race and Ethnicity Standard," *General Methodology*, American Statistical Association.

In January 2003 the Current Population Survey (CPS) began using modified race and ethnicity questions to meet the 1997 Standards for Maintaining, Collecting and Presenting Federal Data on Race and Ethnicity issued by the Office of Management and Budget. These modifications included allowing individuals to indicate that they were of more than one race and directly asking if individuals were of Hispanic, Latino, or Spanish origin prior to asking about their race. To gauge the effect of these changes, the modified questions were asked in May 2002 as a supplement to the CPS. This paper uses the May 2002 data and exploits the

repeated interviewing of a portion of CPS households month to month and over the year to assess the effect of changing the race and ethnicity questions on major estimates derived from the CPS, including the unemployment rate and usual weekly earnings. To help data users construct historically consistent time series, this paper also explores several methodologies to predict how respondents under the old racial and ethnic classification scheme would be classified in the new paradigm.

Robertson, Kenneth W., and **Frugoli, Pamela L.,**(1999) " Statistical Issues for the Redesigned Occupational Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The purpose of this paper is to provide information on changes to the design and uses of the Occupational Employment Statistics (OES) survey. The survey is designed to measure occupational employment and wage rates by industry at the National, State, and Metropolitan Statistical Area (MSA) levels. During the course of the paper, the authors will describe several aspects of the survey. These will include the survey data and its uses, the sample design and estimators, and several statistical issues that were of interest during the redesign. The following issues are described: (1) combining data across years, (2) allocating the sample, (3) assigning sampling weights, (4) ratio adjusting the estimates to meet multiple estimation goals, (5) updating previous-year wage information, and (6) developing a procedure to calculate medians using grouped wage data from multiple years.

Robison, Ed, (2003), "Redesigned Current Population Survey Weighting," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Redesigned weighting for the monthly Current Population Survey (CPS) was implemented January 2003. New national and state coverage steps were added. A reconfigured second-stage iterative proportional fitting process forces weighted data to agree with independently derived population controls. A modified composite weighting procedure is another iterative process that forces weighted data to match sets of labor force controls generated with specialized composite formulas. The effect of each step and interactions among steps is analyzed. Particular emphasis is placed on the new state coverage step and the reconfigured second-stage state step. The steps afford much more detailed population control for states than prior procedures and dramatically stabilize monthly estimates for demographic subgroups within states.

Robison, Ed (1999), *"*Sampling and Reporting in Time-Use Surveys," *Proceedings of the Section on Social Statistics,* American Statistical Association.

Many variables potentially affect the way that people use their time. Individuals and families may allocate their time differently based on gender, income, and

presence of children, among other factors. In order to report information on time use by these and other variables, a time-use survey sample must be designed that will produce accurate estimates by many of these characteristics. The study of the allocation of time by households or individuals must also examine the tradeoff of time to leisure, paid employment, and other activities. Thus, measures need to be made that examine the use of time during the workweek (or work days) as well as during weekends (or nonwork days). This paper will describe the sample design and reporting issues that were faced when recommending a strategy for a national time-use survey.

Robison, Ed, Duff, Martha, Schneider, Brandon and **Shoemaker, Owen**, (2002), "Redesign of Current Population Survey Composite Weighting," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Weighting for the monthly Current Population Survey (CPS) includes a process called composite weighting. It is found that for highly correlated items, a lower-variance estimate for the current month can be obtained by also using data from previous months, suitably adjusted with an estimate of change. A composite estimate of this type has been in use for the CPS for decades, but a weighting procedure that could reproduce composite estimates from just a single monthly data file was not implemented until 1997. The procedure now in place successively computes ratio adjustments to weights so that estimates are forced to match three sets of composite estimates of employment, unemployment, and not-in-labor-force: 1.) by state, 2.) by ethnicity x gender x age, and 3.) by race x gender x age. A new process to be implemented in January 2003 retains the three-way iteration, but is designed based on a reevaluation of convergence properties and interaction.

Rosen, Richard, Harrell, Louis, Hertwig, Ralf and **Gomes, Tony**, (2003), "Efforts to Improve Response Rates in the Current Employment Statistics Program: Results from Two Experiments," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Achieving desired response rates in a timely and cost-effective manner is a top priority in the Current Employment Statistics (CES) program. CES is a monthly survey of about 300,000 business establishments that provides estimates of employment, payroll, and hours, for the nation, states and metropolitan areas. CES uses a number of collection methods with Touchtone Data Entry (TDE) the predominant method. Under TDE, delinquent respondents receive a nonresponse prompting (NRP) "reminder" call or fax message when it's time to report their data. This paper reviews two methodological tests to improve response rates: 1) use of incentives, and 2) an alternative NRP script. For the incentives test, the test sample received a small gift (a pocket size calculator or a mouse pad). For the alternative NRP method test, rather than an NRP reminder call or fax message, the test sample received a call and were asked to provide the data while on the phone. The paper analyzes response rates and discusses factors that may impact

them. Overall, the results do not appear to show any significant impact on response.

Rosen, Richard, Gomes, Tony and Hertwig, Ralf, (2002), "The Impact of Prompting on Response Rates: Experience with Touchtone Reporting in the CES Program," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Achieving desired response rates in a timely and cost-effective manner is a top priority in the Current Employment Statistics (CES) program. CES is a monthly survey of 250,000 business establishments that provides estimates of employment, payroll, and hours for the nation, states, and metropolitan areas. CES uses a number of collection methods. Touchtone Data Entry (TDE) is the predominant collection method, with over 60% of the sample. Under TDE, delinquent respondents receive one or two reminder messages each month. This paper measures the impact of these messages on response rates. Independent test and control samples were selected from the TDE respondent population. In one test sample, the advance notice message was withheld for a five-month period. In the other test, the nonresponse message was withheld for a five-month period. Both groups had a parallel control group for comparison. The results show that prompting respondents twice each month achieves the highest response rates. Witholding either message reduces the response rate by 13 percentage points. The paper compares response rates over the five-month period and explores mode effects associated with different prompting methods.

Rosen, Richard, Harrell, Louis, **Manning, Christopher** , and **Skuta, Doug,** (1999) *"*Data Collection Issues Related to Implementing the Redesigned Current Employment Statistics Survey," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Current Employment Statistics (CES) program is a monthly survey of over 375,000 business establishments. The survey produces estimates of employment, hours, and earnings for the nation, States, and metropolitan areas. CES data serve as a key measure of the health of the economy. Based on several years of research and the recommendations of a number of expert panels, in mid-1997 BLS launched sample solicitation for a new probability design to replace the existing quota sample. The new sample design is a simple random sample of Unemployment Insurance accounts. Initiating the new sample requires soliciting ongoing reporting from over 240,000 firms. This large-scale undertaking required development of new enrollment protols and procedures. Initiation of the new sample is accomplished primarily via CATI. Ongoing data collection utilizes such automated methods as Touchtone Data Entry (TDE), FAX, Electronic Data Interchange (EDI), and World Wide Web (WWW). In this paper , we discuss the issues involved in implementing data collection. We focus on developing the necessary infrastructure to perform enrollment and data colleciton, development of respondent materials, and management of survey activities across three Data

Collection Centers and multiple collection modes. Results to date and implications for other business surveys will be discussed.

Sangster, Roberta, (2003), "Can We Improve Our Methods to Reduce Nonresponse Bias in RDD Surveys?, " *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

There is general consensus that response rates are declining in RDD telephone surveys. The decline has been controlled in many surveys by increasing effort and thus the cost of data collection. In telephone surveys the typical types of compensation are refusal conversion and increasing the number of call attempts. Both methods tend to be costly and time-consuming, and increase respondent burden and frustration. Other strategies to improve response include research on how to deal with answering machines, best time and days to call, and adding advance letters into the survey process. None of these ideas have ever been pulled together into one coherent theory. This paper argues that it is time to revise our calling strategy for RDD surveys to better address how to reduce nonresponse bias. It proposes that we development a model that optimizes calling strategies with the express intent to reduce nonresponse bias.

Sangster, Roberta and Willits, Fern (2001), "Evaluating Numeric Rating Scales: Replicated Results," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The use of numeric rating scales is very common in survey research. However, studies have found that responses vary by changing numeric scale labels from all positive integers (0 to 10, continuum scale) to include negative and positive integers (-5 to +5, bipolar scale) (Schwarz et al., 1991). The intent of this research is to generalize Schwarz et al's finding across different questions and samples using replicated experiments. The data comes from 8 mail or self-administered surveys that together contain 31 experiments. In general, people are less likely to choose zero or negative integers from a bipolar scale than equidistant integers on a continuum scale. The consistency of the pattern is remarkable across the mail and self-administered surveys. The reluctance to select negative integers can mean that part of a bipolar scale will be virtually unused. In such cases, the spread of the scale values is attenuated, leading to a relatively high mean score and a reduced variance. However, this disappears for less skewed data. We suggest several ideas for designing and using continuum and bipolar measures based on these results.

Schmidt, Mary Lynn and **Smith, Dale** (2001), "Determining a Market Basket Structure for the U.S. Consumer Price Index Based on the International Standard Classification of Individual Consumption by Purpose (COICOP)," *Proceedings of the Section on Government Statistics*, American Statistical Association.

The Classification of Individual Consumption by Purpose (COICOP) as part of the 1993 System of National Accounts (SNA) is widely used throughout the world for classification and coding of personal consumption-related data associated with consumer expenditure surveys, personal consumption estimates for the national accounts and Consumer Price Indices (CPIs). For this study COICOP is used in establishing a new experimental market basket structure for the U.S. CPI. New weights are estimated and current Entry -Level Item (ELI) Clusters used for CPI price collection are reclassified according to COICOP. Relative importances at the four-digit COICOP level are estimated for use in comparison with corresponding data from other countries.

Scott, Stuart and **Zadrozny, Peter** (1999)*,"* Aggregation & Model-based Methods in Seasonal Adjustment of Labor Force Series *Proceedings of the Section on Business and Economic Statistics,* American Statistical Association.

One of the most closely-watched numbers produced by the U.S. Bureau of Labor Statistics (BLS) is the seasonally adjusted civilian unemployment rate. It is computed from eight employment and four unemployment series. A motivation for examining this issue of aggregation is that four of twelve components of the unemployment rate are for teenagers and four are agricultural employment, series with substantial sampling error. This study revisits work of Estela Dagum (1978).International interest in model-based seasonal adjustment has increased with availability of the TRAMO/SEATS software of Gomez and Maravall (1994). The methodology, which builds on the work of Burman (1980), carries out signal extraction on seasonal ARIMA models. Comparisons will be made between TRAMO/SEATS and X-11 methodology using Census's X-12-ARIMA (Findley et al, 1998).

Schwartz, Lisa and **Paulin, Geoffrey** (2000) "Improving Response Rates to Income Questions: A Comparison of Range Techniques**"** *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Income and other sensitive questions continue to yield high non-response rates on government and non-government surveys. Recent household surveys have implemented the use of follow-up questions with brackets as a partial solution to missing income data. In addition to gaining some information, brackets may provide more accurate imputed values. However, follow-up questions risk increasing respondent burden. The present study directly compares three alternative range techniques: the conventional response card, unfolding brackets, and a new procedure called respondent-generated intervals (RGI) (Press, 1999). Sixty participants were randomly assigned to one of the three range techniques and used that technique to respond to a series of income questions. Participants completed three rating scales that measured their willingness to provide income information, the ease with which they reached an answer and the intrusiveness of the range technique. A structured cognitive interview format was employed to further explore participants' subjective experience of the income interview.

During the cognitive interview, participants sampled and rated each of the other two range techniques under consideration. The results indicated that bracketing questions can be used to reduce nonresponse to income questions. All three techniques generated equally useful income data as indicated by interval width and ability to predict the exact value based on the midpoint of the range. The conventional and unfolding techniques, but not RGI, lost some predictive utility at the upper income levels. The conventional technique was the most preferred whereas the unfolding technique was the least preferred and the most intrusive. Factors underlying participants' preferences and their implications for survey design are discussed.

Scott, Stuart and Pfeffermann, (2003), "Evaluation of Two Variance Methods for X-11 Seasonally Adjusted Series," *Proceedings of the Section of Business and Economics Statistics*, American Statistical Association.

Pfeffermann (1994) and Bell & Kramer (1999) have proposed solutions to the long-standing problem of estimating the variance of X-11 seasonally adjusted series. The starting point is a decomposition of the observed series into trend, seasonal, irregular, and sampling error components. The Bell-Kramer method relies on models for the sampling error and the overall series. The Pfeffermann method makes more extensive use of the data and X-11 results, and is assisted by having a model for the sampling error. In addition, the variances are defined and calculated differently. A simulation experiment is carried out to evaluate performance of the methods, including a comparison of the underlying variance definitions. The models used in the simulation are based on a labor force series from the U.S. Current Population Survey for which sampling error information is available.

Searson, Michael, (2003), "Web Collection in the Covered Employment and Wages Program," *Proceedings of the Section of Survey Research Methods*, American Statistical Association.

The Bureau of Labor Statistics Covered Employment and Wages Program is developing a web-based collection program to address the data collection needs of its Annual Refiling Survey (ARS) and Multiple Worksite Report (MWR). The purpose of the ARS is to review and update the classification codes assigned to the 8 million worksites on this database. Approximately 1/3 of these worksites are reviewed each year. The purpose of the MWR is to dis-aggregate the employment and wages collected from tax reports (statewide) to the worksite level to meet the industrial and geographical needs of the program. Approximately 120,000 legal entities provide these data for 1.2 million worksites each quarter. This paper focuses on the different approaches that will be used to design and implement these web-based collection systems. Differences in the collection periodicity, employer contacts, survey content, employer size (number of worksites), solicitation procedures, testing environment, and other relevant issues will be used

to determine the best approach for both surveys. This cost/benefit analysis will help determine which employers within each survey are best suited for web-based collection.

Searson, Michael A. (2002), **"**Automation Strategies for the Covered Employment and Wages (CEW) Program," *Proceedings of the Section of Government Statistics*, American Statistical Association.

With potentially declining budgets and increasing workloads, the CEW program faces enormous challenges for the next decade. CEW is the cornerstone of the other statistical surveys of businesses at BLS, since it is used extensively as a sampling frame and population controls for employment levels. CEW uses the administrative records of the State's Unemployment Insurance (UI) system and supplements these data with information collected from two surveys. Since these data are administrative based, BLS has no option but to process all of these data.

This paper focuses on efforts to automate and reduce burden and decrease data collection and processing costs of the surveys---the Annual Refiling Survey (ARS) and Multiple Worksite Report (MWR). The ARS is used to review and update the industrial and geographical codes initially assigned to an employer. The MWR is used to collect employment, wages, and business identification information for each establishment from large employers. This paper addresses the current use of a pilot Touch-tone Data Entry system for the ARS and future research projects as well as the electronic collection of the MWR data, and future Web-based projects.

Searson, Michael A. (2001), "The Business Establishment List - Standard Statistical Establishment List Comparison Project*," Proceedings of the Section of Government Statistics*, American Statistical Association.

The Bureau of Labor Statistics (BLS) and the Bureau of the Census (BOC) each maintain large files of statistical information about businesses in the United States. These business lists are similar in content and program use, are costly for both agencies to develop and maintain, and are burdensome for businesses with multiple establishments. Comparing the two lists at both the macro and micro perspective provides the information necessary for assessing opportunities for list improvements and potential data sharing between the BLS and the BOC. Much of the initial research activities on this topic have focused on comparing and evaluating the SSEL and BEL at the microdata level. It is hoped that the research work will result in the evaluation of the following: the statistical concepts, coverage, content, and quality of both lists; the direct BOC and BLS program uses of respective list information, including data collection and product uses; and actions that could improve business list information, operations, or uses with or without new statutory authority. This paper will describe the BEL-SSEL research project and its current findings, and pose some possible questions for future research.

Searson, Michael (2001), "Strategies to Implement Electronic Collection of Multiple Worksite Report Data," Paper presented at the Federal Committee on Statistical Methodology Conference.

From a 1992 Response Analysis Survey of employers, the Bureau of Labor Statistics (BLS) determined that employer payroll/ tax filing practices fell into two distinct groups: 1) those that prepared their own tax reports using payroll/tax software that was developed internally or purchased from a software vendor 2) those that chose to contract with a third party (typically a payroll provider firm) to prepare their payrolls and file their tax reports. Based on this information, BLS staff developed an electronic data collection strategy to reduce their costs and improve the employer response rates for one of their largest surveys, the quarterly Multiple Worksite Report (MWR), which is filed by 112,000 legal entities representing 1.2 million worksites.

Shipp, Kenneth (2003), "Federal/State Cooperation in the Current Employment Statistics Program: An Update on Automating Data Processing," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics (BLS) Current Employment Statistics (CES) Program is a joint federal/state venture providing important and timely economic data for national, state, and area needs. Within this venture, BLS acts as a program manager and producer of national data while states produce independent state and area data using statistical methods developed by BLS. To ensure a consistent application of these methods, the Automated Current Employment Statistics (ACES) computer system is now used in every state agency. ACES performs all the functions necessary for the operation of the CES program in each state with a particular emphasis on estimation and benchmarking tasks.

This paper will update the development and maintenance of the ACES system and its benefits to both BLS and State agencies. Particular focus will be given to methodological improvements by State agencies using ACES as they completed implementation of the new probability design of the CES program and the introduction of data for the North American Industry Classification System (NAICS) earlier this year.

Chaiken, Meredith, Knott, James, Kofner, Aaron and **Shipp, Ken** (1999) "Increased Timeliness of Current Employment Statistics at the Bureau of Labor Statistics," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics (BLS) provides timely and accurate information reflecting the current state of the economy to the public. This monthly data includes the level of employment, aggregated as well as broken out by industrial division, at the national, state, and local level of analysis. These figures reflect

the total number of jobs in the economy at a precise moment in time; however, the lag between the collection of data and the publishing date may detract from the usefulness of the information. Recently, the BLS has taken steps to increase its timeliness without sacrificing objectivity or the quality of the estimates. Comparing the preliminary estimates to the final, revised data, we attempt to assess the effects, if any, of an incremental speed-up in production time. Using statewide nonfarm employment data from 1988, 1993, 1996, 1997 and 1998, we study the relative size of the revision of the estimates. Furthermore, we consider the effects at the industry level in the most recent three years. While there is some variation among years and industry, there is not significant evidence to indicate a compromise in quality of the data, while the benefits of a quicker production cycle are quite clear.

Shoemaker, Owen (2003), "Estimation and Comparison of Chained CPI-U Standard Errors With Regular CPI-U Results (2000-2001), *Proceedings of the Section of Survey Research Methods*, American Statistical Association.

In 2001 the BLS calculated and published the first C-CPI-U set of indexes for the 12 months of 2000. This new C-CPI-U (Chained Consumer Price Index - Urban) is calculated and published every year, with a one year lag, using a Tornqvist formula, and can thus be labelled a "superlative" index. A set of weights are updated yearly, so that a unique set of monthly weights are available for both time t as well as for time t-n. We briefly outline the Tornqvist formula and then the methodology for estimating a set of standard errors for these new chained indexes. We then compare, over the 24-month period of Jan 2000 through Dec 2001, these new superlative index results and their standard errors with the regularly published CPI results and their published standard errors.

Shoemaker, Owen (2001), "Estimation of Variance Components for the U.S. Consumer Price Index: A Comparative Study," *Proceedings of the Section of Government Statistics*, American Statistical Association.

For every new sample for the commodities and services (C&S) component of the U.S. Consumer Price Index (CPI), the Bureau of Labor Statistics attempts to produce a C&S sample design that allocates outlets and quotes in an optimal fashion. This item-outlet optimization C&S sample design requires the estimation of components of variance for the three factors in the design: non-certainty primary sampling units (PSUs), item-strata and outlets. The fourth component of variance is the error term. To produce these components of variance, a Random Effects model is chosen, with the independent random variable for the model an individual price change. Weighted Restricted Maximum Likelihood (REML) estimates are used to calculate the variance components. This paper compares an earlier set of variance components obtained from 1993-96 CPI data with an updated set of variance components obtained from 1997–2000 CPI data. We compare and critique the methodologies used as well as the empirical results obtained.

Shoemaker, Owen, (2002), "Estimation and Analysis of Variance Components for the Revised CPI Housing Sample," *Proceedings of the Section on Government Statistics,* American Statistical Association.

In 1999, the Bureau of Labor Statistics (BLS) introduced a new Housing Sample for the Rent and Rental Equivalency (REQ) estimators in the U.S. Consumer Price Index (CPI). The Housing Sample consists of, roughly, 10,000 sampled segments, composed of U.S. Census blocks, allocated in 87 Primary Sampling Units (PSUs) and collected in six panels every six months. In this paper, we model the 6-month price relative for both Rent and REQ, and analyze a random effects model that treats PSU and Segment as two random effects. We look at three years of data (1999-2001). We use the Restricted Maximum Likelihood (REML) estimation method to produce the variance components. Standard F-test procedures are applied to determine the significance of the effects in the model. Finally, the variance component results are compared to a set of variance components produced from the previous housing sample (1987-1998).

Shoemaker, Owen (2001), "Estimation of Variance Components for the U.S. Consumer Price Index: A Comparative Study," *," Proceedings of the Section on Government Statistics,* American Statistical Association.

For every new sample for the commodities and services (C&S) component of the U.S. Consumer Price Index (CPI), the Bureau of Labor Statistics attempts to produce a C&S sample design that allocates outlets and quotes in an optimal fashion. This item-outlet optimization C&S sample design requires the estimation of components of variance for the three factors in the design: non-certainty primary sampling units (PSUs), item-strata and outlets. The fourth component of variance is the error term. To produce these components of variance, a Random Effects model is chosen, with the independent random variable for the model an individual price change. Weighted Restricted Maximum Likelihood (REML) estimates are used to calculate the variance components. This paper compares an earlier set of variance components obtained from 1993-96 CPI data with an updated set of variance components obtained from 1997–2000 CPI data. We compare and critique the methodologies used as well as the empirical results obtained.

Shoemaker, Owen and Johnson, William (1999), *"Estimation of Variance Components for the U.S. CONSUMER PRICE INDEX ," Proceedings of the Section on Government Statistics,* American Statistical Association.

For every new sample for the commodities and services (C&S) component of the U.S. Consumer Price Index (CPI), the Bureau of Labor Statistics attempts to produce a C&S sample design that allocates outlets and quotes in an optimal fashion. This item-outlet optimization C&S sample design requires the estimation of components of variance for the three factors in the design: non-certainty primary sampling units (PSUs), item-strata and outlets. The fourth component of

variance is the error term. The total variance of these unit components of variance, divided by their respective number of PSUs, item-strata, outlets and quotes, is then minimized by the optimal number of respective outlets and unique quotes. To produce these components of variance a Random Effects Model was chosen, with the independent random variable for the model an individual price change. Weighted Restricted Maximum Likelihood (REML) estimates were used to calculate the variance components. This paper explicates the methodology for and the results from these estimates.

Shoemaker, Owen and Pathak, P.K. (2001), "The Sequential Bootstrap: A Comparison with Regular Bootstrap," *Communications in Statistics --- Theory & Methods* Vol. 30 (8&9) 2001

Based on Bradley Efron's observation that individual resamples in the regular bootstrap have support on approximately 63% of the original observations, C.R. Rao, P.K. Pathak and V.I. Koltchinskii (1997) have proposed a sequential resampling scheme. This sequential bootstrap stabilizes the information content of each resample by fixing the number of unique observations and letting N, the number of observations in each resample, vary. The Rao-Pathak-Koltchinskii paper establishes the asymptotic correctness (consistency) of the sequential bootstrap. The main object of our investigation is to study the empirical properties of the Rao-Pathak-Koltchinskii sequential bootstrap as compared to the regular bootstrap. In all our settings, sequential bootstrap performs as well or better than regular bootstrap. In the particular case where we estimate standard errors of sample medians, we find that sequential bootstrap outperforms regular bootstrap by reducing variability in the final bootstrap estimates.

Spletzer, James and **Pinkston, Joshua**, (2002), "Annual Measures of Job Creation and Job Destruction Created from Quarterly Microdata," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association

As firms expand and contract, jobs are created and destroyed. We are constructing annual job creation and destruction data using quarterly data on establishments from the Unemployment Insurance (UI) systems of the 50 states. The goals of this project are to complement the quarterly tabulations that the BLS currently publishes, while also producing data that will allow direct comparisons to existing statistics from the academic literature and to job flow statistics from the European countries. The longitudinal database resulting from this work will cover nearly all establishments in all industries and will provide an excellent source of data for research into labor demand topics such as employment adjustment and establishment survival.

Springer, Glenn, Dorfman, Alan H, Paben, Steven and **Walker, Martha** (1999), "Evaluation of Confidence Interval Estimation Methods for the National Compensation Survey," *Proceedings of the Section on Survey Research Methods,*

American Statistical Association.

The National Compensation Survey (NCS) is a Bureau of Labor Statistics (BLS) program which provides data on occupational wages. A theoretical investigation presented by Casady, Dorfman, and Wang (1994 , 1996) suggested that 95% confidence intervals for estimates, when based on the standard normal distribution and standard methods of variance estimation, tend to yield less than the actual 95% coverage. They presented nonstandard methods of constructing confidence intervals, which give more accurate coverage. These intervals tend to be longer than the standard intervals and depend mainly on the use of a t-statistic having degrees of freedom dependent on the available domain data. We modified this method to make it suitable to the multi-stage design of the NCS. Using NCS data, an artificial sampling frame was created and simulated samples were selected. The standard normal confidence intervals were compared to confidence intervals using the t-distribution with weighted degrees of freedom for estimates of means, totals and quantiles. Coverage properties for confidence intervals using the non-standard approach were found to be superior to the standard normal approach.

Zarate, Alvan, Hoy, Easley, **Stamas, George,** and Therriault, Gene D., (2000), "Disclosure Review Board of Federal Agencies: Characteristics, Defining Qualities and Generalizability," *Proceedings of the Section on Government Statistics,* American Statistical Association.

Federal statistical agencies are increasingly concerned about protecting the confidentiality of the data that they collect. In assuring that information released to the public does not put respondent privacy at risk, they have resorted to a number of mechanisms. Some agencies have relied on ad hoc disclosure risk assessment by one or more persons. Others have designated a single official to review such releases. Still others have asked other agencies to review their occasional releases. During the 1980s, when microdata had become a popular means of data release, the U.S. Census Bureau established a Micro-data Review Panel that worked successfully and became it's present Disclosure Review Board. In recent years, disclosure review boards (DRBs) have been established at several other agencies (NCES, NCHS, BLS). This panel will review the composition and functioning of DRBs at these agencies and discuss the necessary features for any DRB and how they might be generalized to other contexts. In doing so, consideration will be given to factors such as available statistical resources, types and levels of statistical and other skills required, and organizational location.

Stewart, Jay. (2002) "Contact Strategies in Time Use Surveys."

In most telephone time-use surveys, respondents are called on one day and asked to report on the previous day's activities. Given that most respondents are not available on their initial calling day, time-use surveys must have a subsequent attempt strategy to contact respondents. Research identified two subsequent attempts schedules: a convenient-day schedule and a designated-day schedule.

Most time-use research recommends the use of a designated-day schedule, but little empirical evidence exists to support this viewpoint. In this paper, computer simulations examine the bias associated with the convenient-day schedule and the designated-day schedule. The results support a designated-day schedule, and validate the recommendations of the previous literature. The convenient-day schedule introduces systematic bias: time spent in activities done away from home tends to be overestimated. More importantly, estimates generated using the convenient-day schedule are sensitive to the variance of the contact probability. In contrast a designated-day-with-postponement schedule generates very little bias, and is robust to a wide range of assumptions about the pattern of activities across days of the week.

Stewart, Jay, Goldenberg, Karen, Gomes, Tony, and **Manser, Marilyn,** (2000), " Collecting All-Employee Earnings Data in the Current Employment Statistics Survey: A Comparison of Two Concepts,"*Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics' (BLS) Current Employment Statistics (CES) program currently collects earnings and hours data for production and nonsupervisory workers. A review of this program concluded that user needs would be better served if CES collected earnings and hours for all employees using a more comprehensive earnings concept. In a pilot test we compared two approaches to measuring all-employee earnings: (1) gross payroll for the pay period that includes the 12th of the month (payroll includes overtime, but excludes irregular bonuses, commissions, or other nonwage payments) plus a measure of nonwage payments for the entire previous month, and (2) total earnings for the entire previous month (total earnings includes all payments made to employees). Two groups of 500 establishments each were asked to report earnings and the associated hours using concept (1) or concept (2). We found that response rates for gross payroll and total earnings were comparable after controlling for observable differences between the groups. The response rate for nonwage payments was lower than for either gross payroll and total earnings. To assess data quality, we matched reported earnings data to those collected from the Unemployment Insurance records. Again, there was no difference between the two concepts.

Stewart, Jay and **Joyce, Mary** (1999), "*W*hy Do We Need Time-Use Data?," *Proceedings of the Section on Social Statistics,* American Statistical Association.

Economists know a lot about individuals' behavior in markets. They have developed theories that explain how individuals respond to change in their economic environment. For example, theory predicts that changes in Social Security that reduce the value of benefits will induce workers to delay retirement. Empirical studies show that prediction is borne out by the data. But there are many areas where economists and other social scientists have well developed theories, but little or no empirical evidence. For example, once people retire,

theory predicts that they will engage in more nonmarket production, but there are no data with which to test this hypothesis. In other cases, interesting questions have gone unanswered because of lack of data. For example, how would the distribution of family income be affected if we accounted for the nonmarket production of non-working spouses? In this paper, we discuss some of the uses of time-use data in the context of these and other unanswered questions.

Stinson, Linda (2000), "'Day of Week' Differences and Implications for Time-Use Research,"*Proceedings of the Section on Social Statistics,* American Statistical Association.

One interesting assessment that can be made with time-use survey data is an analysis of the differences in activity patterns between various days of the week. This study is an attempt to understand the influences of the "day-of-week" variable by comparing daily activity profiles using an index that measures the similarity of activities between pairs of days. The distribution of these differences in time allocation should, in turn, provide contemporary guidance for designing and reporting survey data where "day-of-the-week" is important either as part of the substantive content of the study or as an important descriptive variable.

Stinson, Linda (1999), *"* Measuring How People Spend Time," *Proceedings of the Section on Social Statistics,* American Statistical Association.

Time-use studies typically have a single focus: the duration of human activities. That is, they ask respondents to report everything they did during a 24-hour period along with the starting and stopping times of those actions. This chronological reporting procedure avoids many of the pitfalls of other survey estimation procedures and is less subject to distortion due to social desirability bias. But there are many methodological considerations to take into account when designing a time-use survey. Decisions concerning reporting procedures and mode of data collection may influence data quality. Likewise, the choice of follow-up probes and the treatment of simultaneous activities can determine the amount of information available for accurate and reliable coding of activities. This paper will describe the methodological decisions with which we were faced when designing a time-use survey and introduce the choices that may be made.

Moore, Jeff, **Stinson, Linda** and Welniak, Ed. Jr. (2001), "Income Measurement Error in Surveys: A Review," *Journal of Official Statistics*, vol. 16.

Focusing on U.S. Government surveys, we briefly examine two of the more typical quality indicators-- "benchmark" comparisons of survey estimates to independent aggregate estimate, and nonresponse— but focus our attention primarily on response error research which compares individual survey respondents' reports to external measures of truth, often obtained from independent record systems. We also examine recent findings of "cognitive" research into respondents' understanding of the meaning of income questions,

their interpretations of the tasks which income questions present, and attempt to link what we know about income reporting errors to these cognitive processes.

Couper, Mick and **Stinson, Linda** (1999), "Completion of Self Administered Questionnaires in a Sex Survey," *The Journal of Sex Research*, vol. 36, pp.321-330.

This paper examines completion of four self-administered questionnaires (SAQs) in the 1992 National Health and Social Life Survey, a large national personal-visit survey on sexual behaviors and attitudes in the United States. Based on interviewer reports, only 66% of respondents self-administered all four SAQs, 15% self-administered none of them, and 20% did some of the SAQs themselves. Three factors that may effect the decision to self-administer the SAQs were explored: (a) capacity, (b) motivation, and (c) question sensitivity. The findings suggest it is important to measure who completes the SAQ and to attempt to maximize self-completion of SAQs in surveys.

Moore, Jeff, **Stinson, Linda,** and Welniak, Ed, Jr. (1999), "Income Reporting in Surveys: Cognitive Issues and Measurement Error." In Monroe Sirken, Douglas Herrmann, Susan Schecter, Norbert Schwarz, Judith Tanur, and Roger Tourangeau (eds.), *Cognitive and Survey Research*. New York: John Wiley & Sons, Inc.

Because income data are germane to a wide array of important policy issues, income questions are almost a constant in government-sponsored surveys. This chapter briefly reviews research on cognitive factors affecting income reporting and research concerning the actual quality of survey reports of income. It's intent is to link knowledge about cognitive factors specific to the reporting of income with knowledge regarding the magnitude and nature of the errors made by respondents in response to income questions.

Moy, Luann and **Stinson, Linda** (1999) "Two Sides of A Single Coin?: Dimensions of Change in Different Settings," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

In March 1998, OMB granted approval to test the wording and format of materials designed to collect aggregate `race' and `Hispanic origin' data according to new standards. The first step in the testing process was to develop three different forms, each approaching the problem of aggregate reporting in a slightly different way. The second step was to test the three forms to determine their strengths and weaknesses and, ultimately, to select a single form to present to OMB. The first wave of testing was conducted in cognitive labs at the Bureau of Labor Statistics (BLS) and the National Center for Health Statistics (NCHS) by professional staff from 3 different agencies: BLS, NCHS, and the U.S. General Accounting Office (GAO). After assessing the information collected in the lab about the three forms, the next step was to take the form into the field for

cognitive testing in the workplace. This paper presents the results of those two waves of testing. It also considers the differences and similarities in the results obtained from cognitive testing in formally different settings and reviews the possible sources of those differences and similarities in findings.

Schecter, Susan, **Stinson**, **Linda** and Moy, Luann (1999), " Developing and Testing Aggregate Reporting Forms for Data on Race and Ethnicity," Paper Presented at the Conference of the Federal Committee on Statistical Methodology.

An interagency team of survey researchers used cognitive methods to develop and test forms that will be used to report aggregated race and ethnicity data. This paper describes the methods and results of the cognitive testing for three unique forms. The paper also addresses data quality issues such as concerns about the incompatibility between the format of the individual source data and the aggregate data. Issues regarding differences between automated and manual data collectors and providers are also addressed.

Couper, Mick and **Stinson, Linda** (1999), "Completion of Self Administered Questionnaires in a Sex Survey," *The Journal of Sex Research*, vol. 36, pp.321-330.

This paper examines completion of four self-administered questionnaires (SAQs) in the 1992 National Health and Social Life Survey, a large national personal-visit survey on sexual behaviors and attitudes in the United States. Based on interviewer reports, only 66% of respondents self-administered all four SAQs, 15% self-administered none of them, and 20% did some of the SAQs themselves. Three factors that may effect the decision to self-administer the SAQs were explored: (a) capacity, (b) motivation, and (c) question sensitivity. The findings suggest it is important to measure who completes the SAQ and to attempt to maximize self-completion of SAQs in surveys.

Swanson, David, (2003), "Detecting Possibly Fraudulent or Error-Prone Survey Data Using Benford's Law," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The quality of any survey's results depends on the ability to collect data that accurately represent the underlying phenomena of interest. In some interview surveys two potential problems with data collection are inaccurate reporting by the respondent, and fabrication of data by a data collector who does not contact the selected sample unit (a process called curbstoning). For each of these cases, one may be able to identify problematic interviews by evaluating the distribution of the leading digits of the responses to the questionnaire. In the aggregate, the distribution tends to follow a pattern known as Benford's Law. Consequently, it may be appropriate to re-interview the cases that display a markedly different distribution of leading digits. This paper describes a potential application of this idea to the Consumer Expenditure Interview Survey.

Swanson, David, (2002), "Determining Within-PSU Sample Sizes for the Consumer Expenditure Survey," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Consumer Expenditure Survey is a nationwide survey of households conducted by the U.S. Bureau of Labor Statistics and U.S. Bureau of the Census to find out how Americans spend their money. The survey's random sample of households is designed to minimize the variance of its nationwide expenditure estimates. However, the survey has a competing goal of also obtaining accurate local area survey estimates to allow them to be published. This paper describes a method of sample allocation that minimizes the variance of the survey's nationwide expenditure estimates, while also providing sufficiently large sample sizes within each local area to support the publication of local area expenditure estimates.

Swanson, David C, Hauge, Sharon K,. Schmidt, Mary Lynn (1999) "Evaluation of Composite Estimation Methods for Cost Weights in the CPI," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

For each local area and for each item stratum, an estimate of consumer expenditure is needed to weight the market basket of goods and services for which the U.S. Consumer Price Index (CPI) is computed. These expenditure estimates, called *cost weights*, must be computed every time the CPI is revised. Due to small sample sizes in the Consumer Expenditure Survey, high variability is an inherent problem in producing such localized estimates. To alleviate this problem, the U.S. Bureau of Labor Statistics (BLS) employs composite estimation to reduce the mean squared error of the cost weight estimates at the index area/item stratum level. In this paper we summarize the research conducted at the BLS over the past ten years on different methods of composite estimation, and describe the method used in the CPI's 1998 revision.

Toftness, Richard, (2002), "Comparison of Variance Estimation Techniques for a Price Index with Non-Independent Weights," *Proceedings of the Section on Business and Economic Statistics,* American Statistical Association.

We present an economic approach to variance estimation of price indexes, applied to import and export trade. Instead of generating weights and price changes independently, we generate one total share value for each period for each industry, using a Gamma model of log trade dollar values, and then we calculate base period weights and quantities based on the Fisher model. To generate price changes, a spike at 0% is estimated; then positive and negative price changes are modeled separately, with the model chosen depending on the industry. The variance estimation methods explored are: the Ratio Biased, the Taylor Linearized, the Stratified Randomly Grouped, and the Stratified Jackknife.

Traetow, Monica, Kropf, Jurgen, and Strifas, Sharon, (2002), "Accounting for Business Births and Deaths in CES: Bias vs. Net Birth/Death Modeling," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Current Employment Statistics (CES) Survey, conducted monthly by the Bureau of Labor Statistics, estimates payroll employment, hours, and earnings. The survey originated as a quota-based sample, but currently is transitioning to a probability-based sample. Because business births and deaths cannot be captured in a timely manner by the CES survey, they are accounted for by modeling. The bias adjustment factor used for the quota sample is discussed and compared to the net birth/death model used for the probability sample. While the latter sample uses imputation of business deaths to account for business births, an ARIMA time series model, or the net birth/death model, estimates the residual not accounted for by the imputations. In contrast, the bias adjustment factors are derived solely from a regression-adjusted mean-error model. Although both models account for business births, the bias adjustment factor model also accounts for other elements of non-sampling error. The numbers derived from the net birth/death model are compared to the results from the bias adjustment factor model.

Tucker, Clyde, Biemer, Paul, and **Meekins, Brian**, (2003), "Latent Class Modeling and Estimation of Errors in Consumer Expenditure Reports, "*Proceedings of the Section on Social Statistics*, American Statistical Association.

This paper uses latent class analysis to make final models of the error structure in different categories of consumer expenditure reports collected in the BLS Consumer Expenditure Quarterly Survey (CEQ). Data from the CEQ are used to calculate cost weights used in the Consumer Price Index. Based upon the model, preliminary estimates will be made of errors in individual consumer reports by category. These estimates are predicated on the assumption that respondents can be assigned to a cell based on their method of completing the survey and their demographic characteristics such that the error within a given cell is approximately the same for all respondents in that cell. As a byproduct of this work, various technical issues surrounding the use of latent class analysis will be explored.

Tucker, Clyde, (2001), "Using the New Race and Ethnicity Data," *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Beginning in 2003, all or most government surveys will allow respondents to identify with more than one race. This change is required by the new race and ethnicity standards issued by the Office of Management and Budget in 1997. In fact, the 2000 census allowed for the reporting of more than one race. A number of questions have been raised about the tabulation and interpretation of these new data, including how to compare to past data series. Research has been done to

develop methods to address these analytical problems. This roundtable will discuss both the analytical issues and the methods for dealing with them.

Weber, Wolf, (1999), *″*A Method of Microdata Disclosure Limitation based on Noise Infusion and Outlier Substitution," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

Recent advances in data processing capabilities are enabling more researchers to analyze large data sets. While increased computing power may result in better research, it also adds to the capability of linking microdata file records to survey respondents, thereby compromising privacy. The Census Bureau's Disclosure Review Board (DRB) is responsible for minimizing the probability of respondent disclosure for Census sponsored or Census collected microdata. Recently the DRB began to enforce a new rule regarding the release of "highly visible" or sensitive variables. The motivation for this study is to suggest an alternate method that will minimize disclosure and will result in more accurate variance and regression based estimates than those that are obtained by the DRB approved method. Several forms of data disturbance were applied to Consumer Expenditure Survey microdata. This paper describes the method that was submitted to the DRB as a suggested alternative to the current disclosure minimization rules.

Wohlford, John, Phillips, Mary Anne, Clayton, Richard and Werking, George, (2003), "Reconciling Labor Turnover and Employment Statistics," *Proceedings of the Section on Government Statistics,* American Statistical Association.

The Bureau of Labor Statistics (BLS) provides two major monthly indicators of employment dynamics: the monthly employment change from the Current Employment Statistics (CES) survey and monthly hires and separations data from the new Job Openings and Labor Turnover Survey (JOLTS). On a conceptual basis, one would expect that the over-the-month difference in hires minus separations would equal employment change for an establishment. However, there are a number of concept, reference period, and employer payroll practice issues that impact this relationship. This paper looks at the difficulty in measuring monthly hires and separations and the methodology that BLS put in place to control for expected differences from employment data. Measures of the magnitude and seasonal nature of the differences will be presented along with what is currently known about the causes of these differences. The paper will conclude with a discussion of future research plans.

Wohlford, John and **Mueller, Charlotte,** (2000), "The Debut of a New Establishment Survey: The Job Openings and Labor Turnover Survey at the Bureau of Labor Statistics ," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The United States has had a strong economic indicator of unused labor supply--the unemployment rate--for decades. However, there has been no parallel measure for unfilled labor demand. In 1999, the Bureau of Labor Statistics (BLS) began developing a new mechanism for producing this long-needed indicator--the Job Openings and Labor Turnover Survey (JOLTS). This new survey will measure a job openings rate and corollary hires and separations rates from a sample of 16,000 business establishments. Late in 2001, monthly estimates of rates and numbers will be published for the nation and for broad North American Industry Classification System (NAICS) sectors. This paper provides a brief history of similar surveys previously conducted by BLS, and discusses the goals of the survey, the concepts used, the sampling procedures, the data collection techniques, the estimation process, and publication plans. It also reviews methodological and data collection research conducted in the development of this new survey. Present throughout the development of concepts, definitions and related methodologies is the overarching goal of providing estimates of a job openings rate directly comparable to the unemployment rate from the Current Population Survey. BLS expects this new indicator to greatly enhance policymakers' understanding of imbalances between the supply and demand of labor.

Zadrozny, Peter (2001), "An Estimated Autoregressive Model for Forecasting U.S. GDP Based on Real-time Data," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association.

Initial estimates of U.S. real GDP are made public 15, 30, and 45 days after the end of a quarter. These can be considered as increasingly accurate noisy observations of unobservable "true" GDP. Using these observations, we estimate the presumed true GDP AR process. We use a state-space Kalman-filtering setup to link unobserved true GDP to the three noisy-observation series. Sampling design considerations allow us to first calibrate the observation error covariances. Conditional on these, we estimate the AR parameters. We use the AR model to develop out of sample forecasts and evaluate the forecasts' accuracy. The application uses post World War II U.S. data.

Zadrozny, Peter, (2000), "Modelling Survey-Error Autocorrelations Subject to Time-in-Sample Effects for Model-Based Seasonal Adjustments," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association.

Survey-error autocorrelations have been ignored in model-based seasonal adjustments of published monthly economic data, causing the adjusted data's trends to be contaminated with survey-error trend components. Using Pfeffermann, Feder, & Signorelli's (JBES, 1998) notion of pseudo survey errors (SE), Tiller (Proc. ASA Bus. & Econ. Stat. Sec., 1996) estimated a conventional, time-invariant, 20 parameter, ARMA(2,17), SE model and used it compute uncontaminated model-based seasonal adjustments. The monthly survey data are generated by a "4-8-4" sampling pattern: respondents are sampled for 4 months,

are not sampled for the next 8 months, and then are sampled for 4 more months. The data come as panels of 8 "balanced" responses, one for each of 1 to 8 times in sample. The paper uses the sample design to develop a parsimonious 5-parameter model of each panel unit's responses. The model is periodic, with 8-month periodicity, and implies Tiller's time-invariant ARMA(2,17) process for average panel responses. The model is estimated using a full 8-unit panel data set representing Florida per capita employment. The paper contributes by demonstrating the feasibility of estimating a parsimonious survey-error model based on the 4-8-4 survey design, for the purpose of obtaining uncontaminated model-based seasonal adjustments.

Zadrozny, Peter and Chen, Baoline, (1999), *"Estimation of Capital and Technology with a Dynmaic Economic Model," Proceedings of the Section on Business and Economic Statistics,* American Statistical Association.

The paper models firms' decisions on capital and technology as outcomes of a dynamic optimization problem. Current stocks of capital and technology are unobserved and stochastic, determined by last period's net of-depreciation stocks, current observed investment and research flows, and unobserved disturbances. A quadratic dual-cost function approximates a constant-elasticity production function. Inputs of capital, technology, labor, and materials trade off along convex-to-the-origin isoquants. Production of saleable output, investment in capital, and research in technology trade off along concave-to-the-origin transformation surfaces. Concavity of the transformation surfaces imposes internal adjustment costs on investment and research, hence, makes capital and technology stocks quasi-fixed. The model is estimated using annual data, from 1947 - 1997, for U.S. total manufacturing. Estimation of capital and technology stocks involves two steps. First, the structural parameters of the model are estimated by maximum likelihood. The missing-data variant of the Kalman filter computes the likelihood function with respect to the observed variables, while allowing capital and technology to be unobserved over the sample period. Second, for given values of estimated parameters, the Kalman smoother obtains sample-period estimates of capital and technology.

Zimmerman, Tamara, (2003), "Evaluation of CPS National Variances," *Proceedings of the Section on Survey Research Methods,* American Statistical Association.

The Current Population Survey (CPS), a nationwide household survey conducted by the Bureau of the Census for the Bureau of Labor Statistics, provides official labor force estimates for the working-age noninstitutional population. Variances for the labor force estimators are based on a modified half sample replication method. Since the variances are based on a sample, the variance estimates have a considerable amount of error. One method to remove some of the sampling error in the variances is to group similar labor force items and fit a generalized variance function to the variance estimates. This paper evaluates the current variance

estimation procedure and suggests another procedure for generalizing variances for national labor force statistics.

Zimmerman, Tamara, Tiller, Richard, and Pfeffermann, Danny**,** (2000), "Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters," *Proceedings of the Section on Business and Economics Statistics,* American Statistical Association.

This paper proposes a simple but very general method for approximating the prediction Mean Square Error (MSE) of the state vector predictors estimators in a state-space model, when estimating the unknown model parameters from the observed series. As is well known, substituting the model parameters with the sample estimates in the theoretical MSE formula that assume known parameter values results in under-estimation of the true MSE. Methods proposed in the literature to deal with this problem are often inadequate, and may not even be operational when fitting elaborate models or when some of the parameters are close to their boundary values. The proposed method consists of generating a large number of realizations from the model fitted to the original observations, re-estimating all the model parameters using the same method as used for the observed values and then modifying the MSE values calculated from the newly generated data sets to obtain the desired MSE approximations. Application of the procedure to a model fitted to a time series of sample estimates of employment ratios from the Current Population Survey (CPS), with eighteen unknown parameters, estimated by a three-step procedure yields very accurate results.