Program and Abstracts

7th International Total Survey Error Workshop

June 2-4, 2013

Ames, Iowa

Hosted by

Center for Survey Statistics & Methodology
Department of Statistics
Iowa State University
Ames, IA 50011-1210

Sponsors

Center for Survey Statistics & Methodology and Department of Statistics, Iowa State U
National Institute of Statistical Science
RTI International
Westat, Inc.

Local Organizing Committee

Emily Berg, Jae-Kwang Kim, Sarah Nusser, Sandie Smith

Program Committee

Emily Berg, Paul Biemer, Edith de Leeuw, Dave Dolson, Brad Edwards, Wendy Hicks,
Alan Karr, Jae-Kwang Kim, Lars Lyberg, Peter Miller, Sarah Nusser

PROGRAM SCHEDULE

SUNDAY, JUNE 2, 2013

All events held at Reiman Gardens in the Speer Room, shuttles provided (see below)

3:00-3:45 PM SHUTTLES

Shuttles will run continuously from Gateway Hotel to Reiman Gardens
Students will meet participants in lobby
Early riders can stroll the gardens!

Late travelers will be shuttled from Des Moines airport to Reiman Gardens

4:00 PM SESSION I: PROCESSING ERROR - A NEGLECTED CORNER OF TSE
Speer Room
Chair: Sarah Nusser

Sarah	Nusser	Iowa State University	Welcome
Lars	Lyberg	Stockholm University	Data Processing Errors and Their Control
Alan	Karr	National Institute of Statistical Science	Integrating Disclosure Limitation Into TSE
Bryan	Stanfill	Iowa State University	Quantifying Recall and Processing Error when Utilizing the Compendium of Physical Activities in Physical Activity Recall Surveys

5:45 PM MIXER

6:30 PM DINNER

MONDAY, JUNE 3, 2013

All daytime events are in the Gateway Hotel Conference Center
All oral presentation sessions are 30 min (~20 min talk, ~10 min discussion)

7:45 AM **BREAKFAST**
Food will be placed outside the *Meadow Room*
Tables for eating are available in the *South Prairie Room*

8:30 AM **SESSION II: EVALUATING MEASUREMENT AND OTHER ERRORS**
Meadow Room
Chair: Paul Biemer

Wayne	Fuller	Iowa State University	Estimation of Measurement Error Properties: The Physical Activity Measurement Survey
Bruce	Meyer	University of Chicago	Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation
Emily	Berg	Iowa State University	Analysis of Measurement Error Models Using Parametric Fractional Imputation

10:00 AM **BREAK**
Refreshments will be placed outside the *Meadow Room*

10:30 AM **(II cont'd)**

Stephanie	Eckman	IAB and University of Mannheim	Use of GIS in Large-Scale Surveys

11:00 AM **SESSION III: TSE IN LONGITUDINAL SURVEYS**
Meadow Room
Chair: Dave Dolson

John	Dixon	Bureau of Labor Statistics	Using Paradata to Model Total Survey Error in the Current Population Survey
Mary	Mulry	US Census Bureau	Evaluating Recall Error in Survey Reports of Move Dates Through a Comparison with Records in a Commercial Database

MONDAY, JUNE 3, 2013 (continued)

12:00 PM **LUNCH**
South Prairie Room

1:30 PM **SESSION IV: NEW APPROACHES TO NONRESPONSE STUDIES**
Meadow Room
Chair: Peter Miller

Mike	Hidiroglou	Statistics Canada	Dealing with Non-response Using Follow-up
Jongho	Im	Iowa State University	Propensity Score Adjustment Under Non-ignorable Non-response Using Information from Paradata

2:30 PM **BREAK**
Refreshments will be placed outside the *Meadow Room*

3:00 PM **SESSION V: TSE IN AGRICULTURAL AND NATURAL RESOURCE SURVEYS**
Meadow Room
Chair: Emily Berg

Sarah	Nusser	Iowa State University	Sources of Survey Error in Land-based Surveys
Melissa	Mitchell	National Agricultural Statistics Service	Nonresponse Bias and Measurement Error in the Agricultural Resource Management Survey
Stephanie	Zimmer	Iowa State University	Combining Estimates from Several Sources for Estimating Acreage of Crops

4:30 PM **30 min free time and poster set-up**

MONDAY, JUNE 3, 2013 (continued)

5:00 PM **SESSION VI: POSTER PRESENTATIONS AND MIXER**
South Prairie Room

Nikolas	Mittag	University of Chicago	Imputations: Benefits, Risks and a Method to Handle Missing Data	**P1**
Jiwei	Zhao	Yale University	Variance Estimation after Multiple Imputation	**P2**
Andreea	Erciulescu	Iowa State University	Evaluating the Impact of Nonsampling Errors on Small Domain Estimates for the Conservation Effects Assessment Project	**P3**
Yang	Li	Iowa State University	Imputation Error in Erosion Estimations Based on a Longitudinal Survey	**P4**
Michael	Price	Iowa State University	Evaluating Errors in Administrative Data	**P5**
Wojciech	Jablonski	University of Lodz	Fieldwork in CATI Surveys: How Can We Improve the Data Quality	**P6**
Lasonja	Kennedy	University of Alabama at Birmingham	Survey Design and Data Analysis: Methodological Considerations to Reduce Error and Increase Instrument Utility in Quantitative and Mixed Methods Research	**P7**
Frederick	Lorenz	Iowa State University	Correlated Residuals in General-Specific Survey Questions	**P8**
Kyle	Vincent	The Bank of Canada	Estimating Population Size with Link-Tracing Sampling	**P9**

6:30 PM **SHUTTLES depart for dinner at the Corn Crib, a local farm**

TUESDAY, JUNE 4, 2013

All activities are in the Gateway Hotel Conference Center
All oral presentation sessions are 30 min (~20 min talk, ~10 min discussion)

7:45 AM BREAKFAST
Food will be placed outside the *Meadow Room*
Tables for eating are available in the *South Prairie Room*

8:30 AM SESSION VII: ASSESSING ERRORS IN MIXED MODE SURVEYS
Meadow Room
Chair: Brad Edwards

Z. Tuba	Suzer-Gurtekin	University of Michigan	Investigating the Bias of Alternative Statistical Inference Methods in Sequential Mixed-Mode Surveys
Jae-kwang	Kim	Iowa State University	An Imputation Approach for Analyzing Mixed-mode Surveys
Stanislav	Kolenikov	Abt SRBI	Mode Effect Analysis and Adjustment in a Split-sample Mixed-mode Web/CATI Survey

10:00 AM BREAK
Refreshments will be placed outside the *Meadow Room*

10:30 AM SESSION VIII: APPLYING THE TSE PARADIGM IN PRACTICE
Meadow Room
Chair: Lars Lyberg

Paul	Biemer	RTI International and UNC at Chapel Hill	ASPIRE: A System for Product Improvement, Review, and Evaluation
Paul	Lavrakas	Independent Consultant and Northern Arizona University	Applying a Total Error Perspective to All Quantitative and Qualitative Research Methods

TUESDAY, JUNE 4, 2013

12:00 PM	**LUNCH & OPEN MIKE**		
Alan	Karr	Open mike	Wild and woolly thoughts on advancing the total survey error paradigm
1:30 PM	**WORKSHOP CONCLUDES**		

ABSTRACTS

SESSION 1: PROCESSING ERROR-A NEGLECTED CORNER OF TSE

Data processing errors and their control

Lars Lyberg, Stockholm University, Sweden
Co-author: Gebrenegus Ghilagaber, Stockholm University, Sweden

Data processing is a set of activities aimed at converting the survey data from its raw state as the output from data collection to a cleaned and "corrected" state that can be used in analysis, presentation, and dissemination. The literature on data processing is small compared to the literatures associated with other survey processes such as questionnaire design, sampling, and data collection. It is not clear why this is the case. Data processing operations such as editing, data capture, coding, statistical disclosure control, and weighting contribute to the total survey error to varying degrees. Often these operations are mixtures of manual work and automation, which makes error structures complicated. Furthermore these error structures vary in severity.

Editing is really a quality control operation of the data collection whose purpose is to identify suspicious entries and, if necessary, correct them. It is not until recently that methodologists have become interested in cognitive aspects of manual editing. Editors contribute to the correlated response variance in ways similar to interviewers' contributions. Recently work on editor debriefings have started to shed some light on the cognitive aspects. Information from editors can help improve operations such as questionnaire design and the response process in general. Most interest on editing has focused on its effectiveness resulting in a steady flow of continuous improvements going from checking all entries to a selection of entries to battle what has been coined "overediting." Sometimes imputation procedures are used in this stage to compensate for missing data.

Data capture is the step where questionnaire or other input data are converted into computer-readable form. Keying has to a very large extent been replaced by scanning, where errors due to rejects and substitutes will occur. Most of the time these error rates are small but their consequences are not always known.

Coding is a classification process. Free-format descriptions of variables such as occupation and education are given code numbers either manually or by means of computer-assistance. Coding errors can be very serious and have been neglected by many statistics producers. Coding needs to be verified and dependent verification has been replaced by independent verification with adjudication in many organizations. Recently work has begun to study the cognitive mechanisms at work in coding, in much the same way as in editing, with the purpose of improving the process.

Statistical disclosure control is a set of measures that should protect individual data from disclosure. The set of measures is very extensive and the literature on the subject is relatively rich. The reason is that breaches of confidentiality could be devastating, not only for individual organizations, but for large

parts of the survey industry. The effects on total survey error of measures such as rounding and cell suppression are not very well understood.

Weights and their calculation are somewhat different from other processing operations. The error structure is simple in the sense that weights are either right or wrong but the impact of erroneous calculations on the total survey error are not thoroughly investigated. The common scenario is that it is discovered that calculations are based on unjustified simplifications and eventually corrected. In other cases statistical products have been using the wrong weights for years. In any case, seldom are there documented processes for checking the calculation of base weights and other weights.

The data processing operations are intertwined in various ways. Editing can discover errors not only in data collection but also in data capture, coding, and even in the final estimates depending on when it is performed. Statistical disclosure control limits the usability of statistical outputs. Coding errors can be devastating for different kinds of analysis where specific population categories are studied. Sometimes data processing operations are conducted together with other operations. For instance, it happens that interviewers act as coders and editors.

This paper sets out to describe what we know about the various data processing operations, how they can be controlled via verification and paradata and their impact on other operations and the total survey error.

Integrating disclosure limitation into TSE

Alan Karr, National Institute of Statistical Sciences
Co-authors: L. H. Cox, H. J. Kim and J. P. Reiter

Statistical disclosure limitation (SDL) is typically not viewed as a component of total survey error. In this paper we make the case that it should be, because most SDL procedures deliberately alter data values—that is, they introduce error. We illustrate why this matters by formulating models and reporting research on a number of key questions, among them:

1. Can, or should, we perform TSE-aware SDL? For instance, do variables subject to high measurement error require less intensive SDL? If the measurement error is understood, should SDL mimic it?
2. How do reducing TSE and performing SDL trade off, in terms of data utility, disclosure risk and cost?
3. How does SDL interact with edit and imputation, the two central (post-data collection) steps in reducing TSE? For example, should imputed values be “exempt” from SDL on ground that there is no need to protect them, because they are not real? Are complex edit and imputation procedures wasted when followed by SDL? What should be done when SDL creates edit rule violations?
4. Can SDL, edit and imputation be performed in a single, coherent process?

The research consists principally of computational experiments based on real and simulated data.

Quantifying recall and processing error when utilizing the compendium of physical activities in physical activity recall surveys

Bryan Stanfill, Iowa State University
Co-authors: Dave Osthus, Sarah Nusser, Wayne Fuller, Alicia Carriquiry and Greg Welk

The Compendium of Physical Activities (Compendium) was developed to link specific physical activities performed in various settings to an associated metabolic equivalent (MET) intensity level. Physical activity recalls (PARs) are commonly used instruments that gather physical activity information on individuals. Oftentimes, PARs and the Compendium are used together; the physical activity information from the PAR is converted into METs via the Compendium, resulting in MET intensities of physical activity for each individual. This process of gathering and converting information is not without error, however. The PAR process is susceptible to cognitive limitation for recalling activity from the past (recall error), while the conversion of physical activity into METs is susceptible to processing error. We develop a method to quantify these sources of error using data from the Physical Activity Measurement Survey (PAMS).

SESSION II: EVALUATING MEASUREMENT AND OTHER ERRORS

Estimation of measurement error properties: The physical activity measurement survey

Wayne A. Fuller, Iowa State University
Co-authors: Bryan Stanfill and David Osthus

That most variables in a survey are subject to measurement error is widely accepted. However, studies that quantify measurement error properties are infrequent because quantification requires the collection of additional data and requires model assumptions. We discuss estimation procedures in the context of a study designed to estimate measurement error properties. Two procedures, a monitor and a self report, were used to collect data on physical activity in the Physical Activity Measurement Survey (PAMS). Error properties of measures of physical activity are particularly important in the analysis of the link between activity and health. We propose estimation procedures for physical activity that involve identifying the different components of the measurement error variance for the two procedures. Adjustments for nonresponse were a part of the estimation process.

Errors in survey reporting and imputation and their effects on estimates of food stamp program participation

Bruce Meyer, University of Chicago
Co-authors: Robert Goerge and Nikolas Mittag

Benefit receipt in major household surveys is often underreported. This misreporting leads to biased estimates of the economic circumstances of disadvantaged populations, program take-up, and the distributional effects of government programs, and other program effects. We use administrative data on Food Stamp Program (FSP) participation in two states matched to American Community Survey (ACS) and Current Population Survey (CPS) household data. We show that nearly thirty-five percent of true recipient households do not report receipt in the ACS and fifty percent do not report receipt in the

CPS. Misreporting, both false negatives and false positives, varies with individual characteristics, leading to complicated biases in FSP analyses. We then directly examine the determinants of program receipt using our combined administrative and survey data. The combined data allow us to examine accurate participation using individual characteristics missing in administrative data. Our results differ from conventional estimates using only survey data, as such estimates understate participation by single parents, non-whites, very low income households, and other groups. To evaluate the use of Census Bureau imputed ACS and CPS data, we also examine whether our estimates using survey data alone are closer to those using the accurate combined data when imputed survey observations are excluded. Interestingly, excluding the imputed observations leads to worse ACS estimates, but has little effect on the CPS estimates.

Analysis of measurement error models using parametric fractional imputation
Emily Berg, Iowa State University
Co-author: Jae-kwang Kim

Maximum likelihood estimation of the parameters of a measurement error model is considered. By treating the true covariate as a latent variable, parametric fractional imputation of Kim (2011) can be used without relying on computationally heavy procedures such as Markov Chain Monte Carlo. The proposed method can be applied to situations where the calibration data are either external or internal. Parametric fractional imputation is illustrated in the context of a measurement error model with non-constant measurement error variances. Possible applications of this approach to estimation of the distribution of errors due to imperfections in a data collection instrument or the data processing environment will be discussed.

Use of GIS in large-scale surveys
Stephanie Eckman, IAB and University of Mannheim

The use of GIS tools in analyzing and conducting large-scale surveys has increased in the last several years and will likely continue to do so as the technologies become less expensive and easier to use. On the analysis side GIS tools can help us communicate findings to wider audiences and illuminate spatial correlations between the collected variables that are easy to overlook. On the data collection side, such tools can increase interviewer efficiency and reduce measurement error. For example, given the coordinate of a respondent's household, we can merge in the distance to the nearest supermarket, train station, etc. But how accurate are these calculations, and how do inaccuracies lead to bias in our analyses?
As we embrace these tools, however, survey researchers should maintain a healthy skepticism about their role. This presentation will review many recent examples of the use of GIS technology to improve the quality and/or cost-effectiveness of survey data, while also shedding light on how its use can introduce errors into surveys and the impacts of these errors on bias and variance in survey estimates.

SESSION III: TSE IN LONGITUDINAL SURVEYS

Using paradata to model total survey error in the current population survey

John Dixon, Bureau of Labor Statistics

This study examines nonresponse and estimates from the Current Population Survey (CPS). The CPS is administered to a household for 4 consecutive months, followed by a break of 8 months, and then interviewed another 4 consecutive months. These 8 interviews (panels) form the basis of this analysis. Possible panel effects in the CPS include changes in estimates of unemployment, sample attrition, and fatigue effects. Paradata, including respondent contact history recorded by interviewers, may help understand those effects. For example, are changes in employment estimates due to movers (households moving due to unemployment or retirement), attrition due to nonresponse, or measurement error (such as respondents better understanding of the unemployment concept in later interviews)? Can those who drop out of the survey be predicted from the previous panel paradata? How different are those who refused in the later panel, compared to those who initially refuse to participate in the survey at the first CPS contact? What differences are there in the respondent reported problems (such as too busy or not interested) over the course of the panels? A series of longitudinal structural equation models are used to investigate these questions.

Evaluating recall error in survey reports of move dates through a comparison with records in a commercial database

Mary Mulry, US Census Bureau
Co-authors: Elizabeth M. Nichols and Jennifer Hunter Childs, U.S. Census Bureau
Parvati Krishnamurty, NORC

The accuracy of a person's recall of the date of moving to an address is important for the U.S. Census since its goal is to enumerate the population at their usual address on Census Day April 1 of the census year. Measurement of the coverage of a census using a post-enumeration survey also relies on respondents accurately recalling when members of their household moved to the address, if any moved around Census Day. Past evaluations of coverage measurement surveys have shown that reporting errors concerning moves close to Census Day were the largest source of error. The U.S. Census Bureau sponsored a study examining the accuracy of respondent's memory associated with dates of moves by comparing the self-reported move month and year provided by the National Longitudinal Survey of Youth, 1997 (NLSY97) cohort to records in a commercial database. The NLSY97 interviews the cohort every year, but the study focused on responses the sample members gave in 2008 and 2009 when they were ages 23 to 29. The commercial database was not an ideal "gold standard" so both sources had their own error structure that presented challenges for a matching study. However, using regression models, the study found some evidence of memory error surrounding move dates. The data suggest that respondents in their mid-20s telescope move dates backwards by almost one month starting around 10 months from the move date.

SESSION IV: NEW APPROACHES TO NONRESPONSE STUDIES

Dealing with non-response using follow-up

Mike Hidiroglou, Statistics Canada
Co-author: Victor Estevao

The quality of surveys is affected by several types of errors. One of the most important ones is non-response. Non-response may introduce biases into the estimates that the surveyor may not even be aware of. Response rates may give us an idea of the extent of the bias (upper bound): generally speaking the higher the response rates the more likely that the bias is not important. However, one can never be sure as it is easy to construct scenarios where the bias is actually smaller for lower response rates.

There are two approaches for attenuating the potential bias associated with non-response. The first one is applied at the design stage by ensuring that a subsample of the non-respondents is followed up: Hansen and Hurwitz (1946) published the first paper on this procedure. The second approach is to incorporate auxiliary information that is related to the variable of interest in the estimation of the parameters. This auxiliary information may be in the form of estimating the probability of response for each unit included in the sample as in Fuller et al. (1994), or as direct auxiliary data incorporated in the estimation as in Lundström and Särndal (1999), or as combination of both as in Kott and Chang (2010). It should be noted that the dual use of a follow-up sample and auxiliary data is not addressed in the literature.

In this paper, we recognize that non-response will also occur in a follow-up. Given expected response rates for the respondents and the followed up sample of non-respondents, we develop an optimum allocation scheme that addresses the allocation of the sample ignoring auxiliary data. As suggested by Little (1986) we split the non-responding sample into response homogeneity groups (based on past knowledge) so as to minimize as much as possible non-response bias in the follow-up sample. We apply a simplified version of the procedure to the data set of Hansen and Hurwitz (1946). Given that we may not always be able to split the sample into response homogeneity groups, we study the impact of also using auxiliary data to reduce the bias.

Propensity score adjustment under non-ignorable non-response using information from paradata

Jongho Im, Iowa State University
Co-author: Jae-Kwang Kim

Paradata are data about survey processes, such as call records, interviewee attitudes and audit trails. The paradata are not main questions in a survey, but they offer additional information for each observation unit that can be useful in understanding the level of non-sampling errors such as nonresponse error.

For example the number of attempts to contact an interviewee can be used to adjust unit non-response when the response mechanism is non-ignorable. By assuming a parametric model for the conditional response probability that directly uses the study variable y as the covariate, the model parameters can be estimated from the conditional likelihood.

In this study, we consider propensity score weighting adjustment with non-ignorable non-response when there are such paradata from a follow-up sample data form. The proposed method provides a consistent estimator of the model parameters specified in the response model. We take some simulation studies and a real data example from a Korean household survey of employment to demonstrate the performance of the proposed method in this setting.

SESSION V: TSE IN AGRICULTURAL AND NATURAL RESOURCE SURVEYS

Sources of survey error in land-based surveys

Sarah Nusser, Iowa State University

Total survey error has not been widely considered for surveys of activities or conditions on the land. Land-based surveys are used to estimate characteristics of agriculture, natural resources and the environment. Most land surveys begin with an area sample that minimizes coverage problems, but the characteristics of other survey errors can differ from household surveys. An important driver is the data collection method, which may involve remote sensing (in the broadest sense), field-visits, and/or an interview with an owner/operator respondent. Nonresponse error will vary with the data collection method, especially for protocols requiring owner/operator consent or response. Remotely-sensed and to some extent field-visit surveys have fewer problems with nonresponse, but they may experience larger problems with identifying the location of the observation unit, which in turn can induce measurement error. We will provide a discussion of survey errors in land-based surveys, with illustrations from specific survey programs.

Nonresponse bias and measurement error in the agricultural resource management survey

Melissa Mitchell, National Agricultural Statistics Service
Co-authors: Morgan Earp (Bureau of Labor Statistics) and Jaki McCarthy (NASS)

With increasing nonresponse rates, biased survey estimates are a growing concern. Surveys conducted by the USDA's National Agricultural Statistics Service (NASS) are no exception. The annual Agricultural Resource Management Survey (ARMS) conducted by NASS provides important economic data on US farms, but has relatively lower response rates than many other NASS surveys. Using Census of Agriculture data, NASS has identified characteristics associated with nonresponse in ARMS and thus identified subgroups of establishments that are less likely to respond (Earp & McCarthy, 2009, 2010).

During the 2011 data collection of ARMS, a random subset of those subgroups was flagged to receive special recruitment efforts. Through the use of tree-models we were able to identify likely nonrespondents and use response propensity classes to examine efficient recruitment techniques for increasing response rates (Earp, Mitchell, McCarthy, & Kreuter, under review). While the assumption is that by increasing response rates, the quality of the data will be improved, we have yet to test that assumption.

This paper focuses on comparing the nonresponse bias as well as the measurement error across these treatment groups, as opposed to just comparing response rates. While providing a small token gift was found most effective in bringing in likely nonrespondents who would otherwise have biased the resulting survey estimates (Earp, Mitchell, McCarthy, & Kreuter, under review), the question remains what effect this treatment had on nonresponse bias and measurement error. If the nonrespondents who were converted to respondents using this treatment were not systematically different than the respondents, it is likely that the treatment did not reduce the amount of bias; however, if converted nonrespondents were systematically different from respondents we would expect to see a decrease in nonresponse bias, which we can test by comparing the relative bias of the mean across key estimates (i.e., production and demographic items) across the treatment and control group. And while we may find evidence of less nonresponse bias in the treatment group versus the control group, the question remains whether the converted nonrespondents introduced increased levels of measurement error. By comparing the reported versus edited values we can calculate an estimate of measurement error for key estimates that can then be compared across the control and treatment group to determine the effect of treatment on measurement error as well as nonresponse bias.

Combining estimates from several sources for estimating acreage of crops

Stephanie Zimmer, Iowa State University
Co-authors: Jae-Kwang Kim, Sarah Nusser, Cindy Yu, Shu Yang, Michael Price, Jonathan Lisic and Jeff Bailey

The National Agricultural Statistics Service (NASS) is committed to providing timely, accurate, and useful statistics in service to U.S. agriculture. One way they achieve this goal is through the June Area Survey (JAS). This survey uses an area frame to sample segments of land in the US during the first two weeks of June. The goal is to estimate planted acreage of crops in the US except Alaska. Because JAS is based on an area frame, the survey has no coverage error. However, the survey has a small sample size, which creates large sampling variance. Our goal is to combine other estimates of crop acreage with the JAS estimate to reduce the overall variance of the estimate. To do this, we need to model sources of error for each estimate.

One estimate can be obtained from NASS's Cropland Data Layer (CDL), which is classification of imagery to reflect land cover throughout the United States. The CDL classifies satellite imagery of the 48 contiguous states using training data from JAS as ground truth. The CDL also provides full coverage of the United States but is subject to classification error. We are working on estimating the classification error.

A second estimate can be obtained from the Farm Service Agency (FSA). Farmers can register fields with the FSA on a voluntary basis in all states. Farmers usually do this because some or all of their crops need to be insured or participate in some farm programs from the Farm Bill. Not all farmers will register their crops with the FSA, but it is very common for major crops. We treat the problem as a unit nonresponse problem by calling farmers who do not register their land nonrespondents. This problem is discussed in another abstract.

Once we adjust for nonresponse in the FSA data, we have three unbiased estimators of acreage. We propose combining these three estimates using a Generalized Method of Moments (GMM) which uses models to estimate and adjust for the errors in the data sources. It can also express the covariance between the estimates.

SESSION VI: POSTER PRESENTATIONS AND MIXER

P1 Imputations: Benefits, risks and a method to handle missing data

Nikolas Mittag, University of Chicago

Missing data is a frequent problem in economics that arises in many different forms and for many different reasons. Variables in an econometric model may be partly missing because the required information could not be obtained for some units, e.g. due to non-response in a survey, or the variables may be missing entirely, for example because the relevant questions were not included in a survey, the information is confidential or was difficult to obtain. The most common way to deal with missing data is to allocate predicted values to the missing observations, but some methods assign multiple values or avoid assigning explicit values altogether. Even though it usually refers to methods that assign explicit values, in lieu of a better term, I use “imputation” to refer to all methods that attempt to solve the problem of missing data. Common economic surveys already include imputed values for the units that did not respond which are usually predicted based on the respondents. This is not feasible if the variables are missing entirely, but if the missing variables are available in another dataset, the entire variables can be imputed. The main difference to partly missing variables is that the source data for the imputations is from a different data set or time period in the latter case. As I discuss below, this may raise some additional problems and make others easier to assess, but the problems in the two cases are sufficiently similar to treat them at the same time.

This paper focuses on the advantages and problems of imputations when the information used to impute a variable that is (partially) missing from the data is obtained from complete observations or an additional data set. However, the problems and advantages are expected to extend to other settings in which imputation methods are commonly applied, such as predicting factor or IRT scores or data augmentation to avoid selection or measurement error (e.g. Brownstone and Valletta, 1996). I first discuss the circumstances under which imputation methods can have benefits. In other cases, imputation should be avoided as it requires additional assumptions and may cause bias. I then line out common sources of bias and how they can be avoided. These benefits and problems have implications for how imputations should be done and I evaluate how well current missing data methods capture the benefits and avoid the problems. Finally, I introduce an imputation method based on an estimate of the conditional density of the missing variable that has several desirable properties, particularly if the data provider wants to make imputations available, and show that it performs well in practice.

P2 Variance estimation after multiple imputation

Jiwei Zhao, Yale University

In complex survey data, we usually encounter appreciable amount of missing values. The missing data mechanism can be missing completely at random (MCAR), missing at random (MAR), or nonignorable. Multiple Imputation (MI, Rubin 1987) is a well-known and well-established procedure to handle missing values and it is an important technique in the literature. In this paper, we consider a regression model with response variable subject to missing values and we concentrate on the variance estimation after MI. At first, we briefly review the results when the missing data mechanism is MAR (Wang and Robins, 1998). Under the MAR assumption, the estimates after MI are always less efficient than the ones before MI. In the following, we focus on the situation when the missing data mechanism is nonignorable. We first propose an estimation procedure before MI under some assumptions on the missing data mechanism. We then conduct MI based on the proposed estimates. However, different from MAR, the variance after MI is not generally necessarily larger than the one before MI. Hence, there is no definite answer to which one, before or after MI, is more efficient under nonignorable assumption. It is possible that MI could increase the estimation efficiency when the missing data mechanism is nonignorable. This is a different phenomenon comparing MAR and nonignorable missing data mechanisms. Finally, intensive simulation studies are conducted to illustrate the finite sample behaviors.

P3 Evaluating the impact of nonsampling errors on small domain estimates for the conservation effects assessment project

Andreea Erciulescu, Iowa State University
Co-author: Emily J. Berg

The Conservation Effects Assessment Project (CEAP) is a series of surveys intended to evaluate envirnomental outcomes associated with conservation practices. Four sources of error in CEAP are nonresponse error, location error, frame problems, and processing error. Nonresponse error occurs due to refusals and can be evaluated using auxiliary data available for the full sample. Location error, resulting from imperfections in data collection protocols and GPS instruments, refers to differences between the sampled location and the location at which data are collected. Location error was judged to be problematic in early CEAP surveys, and the magnitude of this source of nonsampling error is expected to decrease as data collection technology improves. Constructing a frame that covers the whole population of interest and does not cover domains that are irrelevant for CEAP is difficult. Only farming operations with land in cropland or certain kinds of hayland or pasture are eligible for CEAP, but current information on landuse is not available at the sample design stage. A fourth source of nonsampling error in CEAP arises because the original collected data (responses to survey questionnaires) are processed through a computer model called the APEX model. The APEX model calculates an erosion value and an erosion index for each respondent as functions of the survey responses and auxiliary information related to climate and soil characteristics. Imperfections in the APEX model may lead error in CEAP estimates. Possible ways to evaluate effects of nonresponse error and error due to frame inefficiency on small domain estimates based on the CEAP survey data will be discussed. Limitations in our abiblity to evaluate location error and error in the APEX model will also be considered.

P4 Imputation error in erosion estimates based on a longitudinal survey

Yang Li, Iowa State University
Co-authors: Emily Berg and Wayne A. Fuller

The National Resources Inventory (NRI), a longitudinal survey of nonfederal US land, provides the Natural Resources Conservation Service (NRCS) with estimates of soil erosion. Average erosion estimates as well as estimates of the acres of erodible land that exceed tolerance rates are published at various domains of interest. Historically, NRI estimates of erosion are based on the Universal Soil Loss Equation (USLE), a model that approximates average annual water erosion as a product of five measurable factors. NRCS is in the process of converting estimates of erosion from USLE to a more nuanced erosion model called RUSLE2. The RUSLE2 model is a computer algorithm that incorporates daily weather and detailed information on conservation practices. Although 2007 was the last year of data collection for USLE, the NRI will continue to produce estimates of USLE erosion through 2010. Extending the series of USLE estimates improves the “fitness for use” of NRI erosion estimates, as estimation of change over time is a priority for this longitudinal survey. USLE and RUSLE2 data from 2004 – 2007 are used to establish a regression model for predicting USLE. Erosion estimates for 2008 – 2010 contain imputation error arising from the use of predicted USLE instead of the true unobserved USLE. This presentation will discuss a proposed regression approach for predicting USLE as well as the issue of evaluating the relative magnitudes of the imputation error and the sampling variance in the erosion estimates.

P5 Evaluating errors in administrative data

Michael Price, Iowa State University
Co-author: Cindy Yu

There is increasing interest in the use of administrative data as input to the estimation process for survey data. Administrative data are subject to many of the same errors as survey data, and it is of interest to evaluate these errors and possibly adjust for them. We consider an example of Farm Service Agency (FSA) administrative data, which will be combined with crop survey data and other auxiliary information.

The FSA is in charge of administering the farm programs that the United States Department of Agriculture sponsors. As a result, the FSA obtains acreage information on crops and fields from farmers who are part of these farm programs or in crop insurance. The FSA provides summaries of this data to the National Agricultural Statistics Service (NASS). The goal of this project is to use this information to create an estimate of the number of planted acres of a particular crop in a particular county. This estimate will then be used as one of three components a Generalized Method of Moments (GMM) model that combines estimates from the FSA data, the Crop Data Layer (CDL), and June Area Survey (JAS).

The main problem with the data is that not all farmers register their land in the data base. There may also be some forms of measurement and processing error. We examine the nature of these errors in the FSA data and propose estimators to adjust for them. A difficult issue is how to benchmark the data to represent all crop acres. Another issue is whether the sign up rate varies with the type of agriculture commodities grown in a state.

P6 Fieldwork in CATI surveys: How can we improve the data quality

Wojciech Jablonski, University of Lodz

In the presentation, we will outline the results of a methodological study carried out in 2009 and 2010 among 12 major Polish commercial survey organizations: *4P research mix*; *ASM Centrum Badan i Analiz Rynku*; *ARC Rynek i Opinia*; *Expert-Monitor* (at present: *Kantar Media*); *GfK Polonia*; *IMAS International*;

IPSOS; *Grupa IQS*; *Millward Brown SMG/KRC*; *PBS DGA*; *Pentor Research International* (at present: *TNS Poland*); and *TNS OBOP* (at present: *TNS Poland*). Each of these companies has CATI facilities, and each carries out telephone interviews on a regular basis. The research was based on a standardized self-administered questionnaire for CATI interviewers. A total of 846 questionnaires were completed. Moreover, we conducted 32 in-depth interviews with CATI interviewers. Interviewers having over one year of job experience were selected based on sex, age, and level of education to ensure that the sample of participants reflected the demographics of the interviewers across all the participating companies. Additionally, in January 2013 a follow-up research was carried. During IDIs CATI survey managers from different companies were asked to comment on the results obtained in both quantitative and qualitative studies.

The presentation focuses on the selected findings of this project. It investigates the issue of different situations that take place during fieldwork, which can have an impact on the reliability of the data. Our intention is to present interviewers' opinion about the interview process. As we see it, they can act as valuable sources of information, and these perspectives should be taken into consideration while preparing and conducting telephone surveys.

One of the most interesting and unexpected findings was the interviewers' behavior when they encountered respondents who experienced difficulties understanding the questions asked. In such cases, interviewers may deviate from the prescribed protocol (reformulate the questions, ask them using language that is more easily understood, explain any terms which may be unclear) to ensure that respondents complete the survey. All interviewers are aware of the importance of standardizing the interview protocol. However, there seems to be unspoken consent to deviate from the rules across most research firms when the respondents appear to experience cognitive difficulties. It is worth stating that almost all interviewers indicated that the questions used in the CATI scripts are often formulated using complicated vocabulary and syntax. The questions have not been adapted to suit the intellectual skills of an average respondent. In the interviewers' opinions, if the researchers placed greater significance on the design of research tools by listening to the interviews and talking to CATI interviewers, there should be no reason for de-standardizing the procedure.

P7 Survey design and data analysis: Methodological considerations to reduce error and increase instrument utility in quantitative and mixed methods research

Lasonja Kennedy, University of Alabama at Birmingham

Researchers regularly develop measures to assess concepts of interest. It is not the answer (i.e. response) to a question that is of interest, but rather the extent to which the answer can be shown to have a predictive relationship to facts or subjective states that are of interest (Fowler, 1984). Proper survey research, therefore, requires instruments that are valid and reliable for the population examined.

A myriad of options exist for development of surveys, whether designed as prototypes or constructed from previously validated scales. The researcher must decide upon the type of instrument (form, electronic), theoretical basis for the tool, and a reasonable number of survey items. Additionally, the order of questions, stem for each item, and format for response options have to be determined. Although items from the aforementioned lists may be easily addressed, additional considerations must be given to conceptualization and validation of the tool. Most importantly, foundational applications are needed to maximize the utility of survey items and information collected.

During the following session the presenter will discuss important research methods to reduce error when designing novel instruments and demonstrate techniques useful for proper analysis of data collected from the tool. The presenter will provide basic demonstrations for those new to the field of survey design and survey research methods and demonstrate advanced techniques for participants skilled within the area. By the end of the session, the attendee should be able to:

1. Review basic elements of hypothesis testing
2. Discuss essential survey design techniques that may impact variability of scores and standard error (SE)
3. Demonstrate analytic methods to identify latent factors, a priori and a posteriori
4. List tools most useful in design and analysis for various expected outcomes

P8 Correlated residuals in general-specific survey questions

Frederick Lorenz, Iowa State University
Co-authors: Virginia Lesser (Oregon State); Laura A. Hildreth and Ulrike Genschel (ISU)

Survey researchers often ask a series of questions on a single topic. One common example is the "general-specific" sequence in which a general question (G) either precedes (GS) or follows (SG) a series of specific items on the same topic (S). Past research on general-specific questions has focused on the effects of specific items on responses to the general question. In this paper, we extend previous research by focusing on the relationships among specific items. In particular, we hypothesize that responses to any one specific item are conditioned by the immediately preceding item. We test this hypothesis by first comparing SEM models with and without correlated adjacent residuals, and then conducting permutation tests to evaluate whether the model with correlated adjacent residuals fits significantly better than one would expect by correlating any random set of residuals. The results show that the model with correlated adjacent residuals consistently fits significantly better than one would expect if random pairs of residuals were correlated. The paper concludes by discussing the implications of these new findings for our understanding of how respondents complete questionnaires.

P9 Estimating population size with link-tracing sampling

Kyle Vincent, The Bank of Canada
Co-authors: Kyle Vincent and Steve Thompson

There has been a growing interest in the study of hard-to-reach populations like those comprised of injection drug-users and the homeless. New link-tracing sampling designs and inferential methods for measuring attributes of such networked populations have therefore received an abundance of attention recently. In this talk we will present a novel strategy for estimating the size of networked populations when samples are selected via a link-tracing design. Preliminary estimates of the population size are obtained with the use of common mark-recapture estimation methods and are based on the conventionally selected members of the samples. A Rao-Blackwellization strategy is then implemented by incorporating the adaptively selected units into the analysis and will in turn achieve an improved estimator. Results from an empirical study of a networked population at risk for HIV/AIDS will be presented to demonstrate the efficiency of the new strategy.

SESSION VII: ASSESSING ERRORS IN MIXED MODE SURVEYS

Investigating the bias of alternative statistical inference methods in sequential mixed-mode surveys

Z. Tuba Suzer-Gurtekin, University of Michigan
Co-authors: Steven G. Heeringa and Richard Valliant

Sequential mixed-mode surveys combine different data collection modes in succession to reduce nonresponse bias under certain cost constraints. However, as a result of nonignorable mode effects, nonrandom mixes of modes may yield unknown bias properties for population estimates such as means, proportions and totals. The existing inference methods for sequential mixed-mode surveys generally assume that mode effects are ignorable. The objective of this paper is to describe and empirically evaluate some proposed multiple imputation estimation methods that account for both nonresponse and nonrandom mixtures of modes in a sequential mixed-mode survey. In particular, the multiple selection imputation models allow imputation of responses for alternative modes as if they responded in a given mode by controlling nonrandom mixes of mode. For example, if personal and telephone interviews are used, one step in the process is to impute values for the telephone cases as if they had responded by personal interview (PI) to produce a completed PI data set. Similarly, a completed telephone interview (TI) data set is created. The completed PI and TI data sets are combined for inference. Through simulations, the method is evaluated in terms of the bias reduction for varying degrees of mode effects and model fit. The American Community Survey (ACS) or the 1973 public-use Current Population Survey and Social Security Records Exact Match data will be used to conduct empirical and simulation evaluations. The focus of the empirical evaluations and simulations will be mean family income.

An imputation approach for analyzing mixed-mode surveys

Jae-kwang Kim, Iowa State University
Co-authors: S. Park and S.Y.Kim

Mixed-mode surveys are frequently used to improve survey participation but statistical tools for analyzing mixed-mode survey data are relatively underdeveloped. Motivated by a real survey in Korea, we consider an imputation approach to handling mixed-mode surveys. The proposed method uses measurement error models to explain the mode effects and then imputation to predict the counterfactual potential outcome in the measurement error model. The model parameters are estimated using the method of parametric fractional imputation of Kim (2011). The proposed method is applied to the survey of private education expenses in Korea.

Mode effect analysis and adjustment in a split-sample mixed-mode Web/CATI survey

Stanislav Kolenikov, Abt SRBI
Co-author: Courtney Kennedy, Abt SRBI

We analyze the results of a national survey collected in two modes: SAQ on the web with personal CATI follow-up of web non-respondents. We apply regression and implied utility-multiple imputation mode effect adjustments. Since some items may exhibit mode effects, such as social desirability bias, a split-sample design has been built into the study, with 13% of the cases randomized into CATI-only condition. Such randomization allows for a rigorous comparison of the item response distributions in the two modes. We analyze the behavioral and attitude items to identify the ones that may have been affected by the mode effect. A logistic model for Yes/No responses or an ordinal logistic model for Likert scales was fit to the data with explanatory variables that included demographic variables and the mode indicator for the subsample of the mode compliers. The regression mode effect adjustments consists of zeroing out the mode variable when forming the predictions based on the estimated regressions, and can be extended to the entire sample. Another mode adjustment is based on econometric framework of implied utilities in logistic regression modeling. We simulated implied utilities of the different responses, followed up by selection of the response with the greatest utility. This is essentially an imputation procedure for the response in the less reliable mode (CATI), and requires the framework of multiple imputations for an appropriate analysis. The variables that exhibited the strongest mode effects were found to be the self-reported incidence of major financial problems in the last 5 years (possible underreporting due to social desirability bias), persuading someone to vote for a political party or a candidate (possible overreporting due to social desirability bias), and contacting a public official (possible overreporting due to social desirability bias). The standard errors of the adjusted estimates have gone up, as was also expected.

SESSION VIII: APPLYING THE TSE PARADIGM IN PRACTICE (Chair: Lars Lyberg)

ASPIRE: A system for product improvement, review and evaluation

Paul Biermer, RTI International and UNC at Chapel Hill
Co-authors: Dennis Trewin, Australian Statistician and Heather Bergdahl, Statistics Sweden

In 2011 Statistics Sweden was challenged by its main stakeholder – the Ministry of Finance – to develop indicators that could show improvements in product quality over time. There are several quality frameworks that address different aspects of quality, such as organizational, process and product quality. The challenge, however, is to measure and monitor changes in product quality in a comprehensive and systematic way and to clearly and concisely present progress towards quality improvement to stakeholders. A tool, we call ASPIRE, was developed and tested in 2011 for the Accuracy Dimension of quality on eight key products. It was extended and enhanced in light of experience and reapplied in 2012 to (1) examine changes in Accuracy since the 2011 assessment for the products, (2) assess the quality for the other quality Dimensions, and (3) introduce new products that heretofore had not been evaluated. In this paper, we describe ASPIRE and how it was used in Sweden to set clear measureable goals for product quality and to measure quality improvements for a diverse set of statistical products.

Applying a total error perspective to all quantitative and qualitative research methods

Paul Lavrakas, Independent Consultant and Northern AZ University

This paper and presentation will focus on my long---held belief that the Total Error (TE) perspective should be the basis from which all social, behavioral, and marketing research practitioners and scholars conceptualize, monitor the implementation of, and interpret their research studies. TE is a comprehensive and systematic framework that encompasses all potential forms of "error" (i.e., sources of bias and variance) that may threaten and undermine the reliability and validity of any research study, including focus groups, IDIs, observational research, content analysis, experimental research, and survey research. I have been explicitly applying the TE framework essentially on a daily bias in all my work for the past 20 years since studying Groves' (1989) seminal text, *Survey Errors and Survey Costs*. My paper/presentation includes discussion of my belief that TE can and should be applied across a very wide range of research purposes and settings. These include: (a) planning, implementing and interpreting any original research study; (b) using results from research someone else has conducted to make important decisions; (c) evaluating the credibility of research studies reported by the news media; (d) structuring legal expert reports and testimony about research studies; and (e) writing RFPs and scoring proposals that are submitted. Despite what some appear to believe, in my view, TE is as relevant (and actionable) for qualitative researchers as it is to quantitative researchers.

The following figure (adapted from Groves, Fowler, Couper, Lepkowski, Singer and Tourangeau; 2004, p. 48) shows what I include in my TE perspective and how I organize my thinking about it. In addition to what Groves et al., included in their total survey error framework, I have added Specification Error (cf. Fuchs, 2008) which relates to the bias and variance that can result from misspecification of the contracts the researcher purports to measure. I also have added Inferential Error which relates to Campbell and Stanley's (1966) concept of Internal Validity; i.e., the extent to which, if at all, a research method/design of choice can be used to support cause-and-effect inferences/attributions.

As shown below, and as I have noted above, the TE perspective can (and I believe should) be applied to any form of social/behavioral research, including methods such as Content Analysis, which can be a very quantitative method, but often is deployed as a qualitative method.
Or for In-depth Interviewing (IDI), which is a highly qualitative research method:
My paper/presentation will argue that all research studies are improved i.e., they will provide higher quality (more likely to be accurate) findings when a TE perspective is used to conceptualize, implement, and/or interpret them.

For qualitative and quantitative researchers alike, the TE perspective provides an excellent framework for self-criticism and ideally will help researchers avoid over-interpreting and over-generalizing from a given study in light of an explicit and comprehensive understanding and acknowledgment of the strengths and limitations of its methods.