

Use <Ctrl-F> To Scan for Specific Author Names or Paper Titles

ABSTRACTS

(Shown in session order)

Thursday, 8:30-10:00

Developing Nonresponse Standards

Tom W. Smith, National Opinion Research Center, USA

For decades professional and trade associations, government agencies, and individual academics and practitioners have been discussing nonresponse standards. Debates have centered around 1) whether surveys should meet certain performance standards and, if so, what those standards should be and 2) how various features of surveys should be defined and what those definitions should be. Within the general debate over standards and definitions, there has been considerable discussion over the issue of survey nonresponse. This paper reviews and summarizes that discussion, indicates the current state of affairs, and examines possible future directions.

The review of the development of nonresponse standards considers in particular the following: 1) the CASRO definitions of nonresponse formulated by the Frankel committee, 2) the adoption of nonresponse standards by the federal government, 3) industry research by the Marketing Science Institute, 4) the research literature in general, 5) the international conferences on nonresponse organized by national statistical agencies, and 6) the recent efforts of the AAPOR Committee on Final Disposition of Cases Codes and the Calculation of Outcome Rates.

In addition, an analysis of the actual calculation and reporting of nonresponse rates in a) top journals (e.g. POQ, IJPOR, JOS, JMR, ASR, APSR, JASA, etc.), b) survey reports and releases (e.g. the Gallup Poll, Yankelovich Reports), c) newspapers and magazines, d) the basic survey documentation maintained in archives at ICPSR, the Roper Center, and the Harris Archive, and e) CPS reports of the Bureau of the Census is carried out.

The two main focuses are on how have the debate and the practice changed over the last 30 years and how does the practice agree with the standards that have been promulgated and/or debated. This discussion includes a description and analysis of the current state of affairs, especially the recent efforts of AAPOR.

Finally, the past and future course of nonresponse standards are examined in light of broad, general models of scientific and intellectual development and regulation. Among the paradigms considered are a) the unregulated, marketplace of ideas approach, b) the professional standards approach as represented by medicine, and c) the government regulation approach.

Thursday, 8:30-10:00

Large-scale Imputation for Complex Surveys

David A. Marker, David R. Judkins, and Marianne Winglee, Westat, Inc., USA

Much of the recent research into imputation methodology has focused on developing "best" procedures for a single variable or set of variables. In contrast, it is frequently necessary to impute for many variables from a single survey, with an even larger set of potential covariates and complex covariance structures among the variables to be imputed. Further, the imputations need to be completed in a relatively short time frame within a constrained budget. The analyst also is unlikely to be able to anticipate all of the important analyses for which the imputed data are to be used. This often prevents analysts from being able to produce "best" imputations for each variable. Instead, one tries to produce a set of imputed variables that minimize the attenuation of key relationships, hopefully reduces nonresponse bias, and satisfies the time and budgetary constraints.

This paper discusses the types of trade-offs that are faced when imputing for large-scale complex surveys. A number of strategies could be implemented in such situations. By providing examples from three U.S. government surveys, the paper demonstrates imputation strategies that can successfully be accomplished in this setting. The three surveys are the National Employer Health Insurance Survey, the Medical Current Beneficiaries Survey, and the Reading Literacy Study by the IEA (International Association for the Evaluation of Educational Achievement). In each survey it was necessary to develop imputation models in large data sets (tens of thousands of respondents) for dozens of variables, many of which were highly correlated with each other, requiring careful consideration of the sequence of imputation. Many of the variables had to be imputed satisfying arithmetic constraints. Item nonresponse rates varied from under five percent to over 50 percent following complex swiss-cheese patterns. After briefly discussing the complications in these datasets, the authors explain the strategies implemented to produce imputations that achieve many of the characteristics desired by "best" procedures. For one of the surveys, a follow up study was conducted to evaluate the effects of imputation on data analysis. Regression estimates based on the imputed data were compared against estimates based on complete cases, available cases, and an iterative estimation and maximization (EM) method of handling missing data.

Thursday, 8:30-10:00

Item Nonresponse

Item Nonresponse in Questionnaire Research With Children and Young Adolescents

Natacha Borgers and Joop Hox, University of Amsterdam, the Netherlands

Children are no longer neglected as respondents in surveys. Researchers are more and more convinced that information about perspectives, attitudes, and behavior of children should be collected directly from the children themselves, and not indirectly from their parents or other sources of information. Thus it has become more usual to let children participate in large-scale surveys.

So far, methodological knowledge on surveying children and young adolescents is scarce, and researchers must rely on ad-hoc knowledge from such diverse fields as child psychiatry and educational testing or on methodological knowledge on surveying adults. Regarding adults, there is increasing empirical evidence that both respondent characteristics (like cognitive development), and question characteristics affect response quality, as shown by research by a/o. Krosnick (1991) and Schwarz (c.f. Sudman, Bradburn & Schwarz, 1996). Because children and young adolescents are still developing their cognitive and social skills, which are necessary to answer questions, it is reasonable that answering questions in surveys brings along some additional problems concerning response quality.

In our study we aim at both the effects of respondent characteristics and question characteristics on the item-nonresponse in questionnaire research with children and young adolescents. The research population is children in the age of eight to eighteen years.

In a pilot study, presented at the SMABS '98-conference (Society for Multivariate Analysis in Behavioral Sciences), the results from three data sets were reported. First we have compared item-nonresponse of groups of children, which differ in aspects that indicate cognitive development. Variables we used as indicators for cognitive development were among others age and years of education. Secondly, we determined the effect of 25 question characteristics on item-nonresponse. This was done for every scale in each questionnaire separately and the results were combined using meta-analysis.

The results of this preliminary study indicate that only a small proportion of variance in item-nonresponse could be explained by child characteristics. Question characteristics explain a larger proportion of variance in item-nonresponse. However, these preliminary results had to be interpreted carefully because of limitations in the data sets. For more stable results, in the study we propose to present, the data set will be increased to at least five data sets, including at least 35 scales.

The theoretical background for our study is Krosnick's satisficing theory (1991), which was confirmed by our preliminary results: Cognitive sophisticated respondents are less sensitive for cognitive demanding questions than less cognitive sophisticated respondents.

To test the hypothesized interaction effect between question and child characteristics on item-nonresponse, we intend to use a multilevel logistic regression model. Multilevel analysis is appropriate here, because we have to distinguish between research units on two different levels: items and respondents.

In summary, our presentation will focus on the results of a multilevel logistic regression analysis, to answer three research questions concerning item-nonresponse in questionnaire research with children:

1. What are the effects of question characteristics on item-nonresponse?
2. What are the effects of respondent characteristics on item-nonresponse?
3. Which interactions between question and respondent characteristics are important in explaining item-nonresponse?

Item Nonresponse in a Survey with Sensitive Topics

Judy A. Hughes, Emory University, USA

Brenda J. Colley Gilbert, and Holly B. Shulman, Centers for Disease Control, USA

Nonresponse to sensitive survey questions is typical in survey research. It has been demonstrated in numerous survey research findings that nonresponse rates usually vary by respondent subgroups. In this study, we examined item nonresponse in the Pregnancy Risk Assessment Monitoring System (PRAMS) survey. PRAMS is an ongoing, state-specific, population-based surveillance system designed to identify and monitor selected maternal behaviors and experiences before, during, and after pregnancy. Currently, 18 states are participating in PRAMS, with surveillance covering approximately 35% of all U.S. births. Every month, in each participating state, a systematic stratified sample of 100-250 new mothers is selected from birth certificates. The data collection methodology for PRAMS is mixed mode, utilizing Dillman's Total Design Method. The PRAMS questionnaire is mailed to the sampled mothers 2 to 6 months after delivery. The questionnaire is mailed to nonresponders a second time, with phone contact attempts made for nonresponders to the second mailing. The PRAMS survey design includes oversampling of populations of women with high-risk of poor pregnancy outcomes. The data are weighted to adjust for survey design and non-response. Overall response rates to the survey range from 66-80%, and all items in the survey have response rates of at least 75%.

The 52-item core PRAMS survey involves topics of sensitive nature, including intendedness of pregnancy, physical abuse, and use of alcohol and tobacco before and during pregnancy. Initial investigation of the item nonresponse rates among 11 states for 1996 indicated that the following variables have high nonresponse rates ranging from 7-19%: mom's weight before pregnancy, number of alcohol binges 3 months before pregnancy, relative frequency of physical abuse, intendedness of pregnancy, and average cigarettes smoked per day 3 months before pregnancy.

Item nonresponse rates were compared by state and for all states combined for these 5 high-nonresponse variables. To identify factors influencing item nonresponse, the nonresponse rates will be presented by various maternal characteristics, including age, race, parity, education level, marital status, WIC status, and mode of contact. Significant predictors of item nonresponse will be determined through logistic regression modeling. Those variables significantly related ($p < .05$) to nonresponse will be included in

the final model.

The results of the analysis are useful to others involved in survey research on sensitive topics. Knowing the patterns of nonresponse to surveys involving sensitive topics and the specific subgroups that are most sensitive to these topics can lead to the use of new methods or adjustments to data collection procedures or during the analysis to increase response rates.

Nonresponse to Income Questions in Person-Based and Topic-Based Questionnaire Forms

Laura Loomis, U.S. Bureau of the Census

Two different interview forms were tested in a field experiment of CATI interviews for nonresponse follow-up to the American Community Survey (ACS). The "person-based" form presented the survey questions to the household respondent in such a way that the survey was conducted for the first household member, then the second household member, and each other household member, in turn. The "topic-based" version used a different approach, asking the household respondent about the same topic for every household member before moving on to subsequent topics in the survey. While most questions in the topic-based form had lower item nonresponse rates than the person-based form, the opposite was true for questions asking about wage and salary income and total income. An analysis of within-household nonresponse patterns found that respondents from topic-based households were more likely than those from person-based households to refuse to answer income questions for someone in their households, and also that topic-based respondents were more likely than person-based respondents to refuse to answer income questions for all of the household members they were asked about.

Thursday, 8:30-10:00

Survey Introductions and Interviewer Training

Effects of Style of Introduction on Nonresponse

Huub van den Bergh and Hanneke Houtkoop-Steenstra, Utrecht University, the Netherlands

In this study the effect of four different introductions on response rates in large scale telephone surveys in the Netherlands was investigated. Three standardized scripted introductions with different numbers of content elements, and a fourth agenda-based introduction were distinguished. In the latter the interviewers formulated their own introductions on the basis of a limited number of catchwords.

A total of 1831 first telephone calls by 132 interviewers were analyzed; only first calls were taken into account. In a multilevel model the three standardized scripted introductions did not differ much to either the response rates, the appointment or the refusal rates. However, the agenda-based introduction induced both higher response rates and higher appointment rates, and therefore lower refusal rates.

I am not Selling Anything: Experiments in Telephone Introductions

Rob van Leeuwen, UMT (Your Opinion Counts), Amsterdam
Edith de Leeuw, Methodika, the Netherlands

Approaching the new millennium, telephone interviewers are faced with new problems that threaten the validity of the method. The growing numbers of unwanted calls (e.g., tell-sell, telemarketing, push polls) influence the willingness to cooperate, and refusals and break-offs are increasing. Potential respondents now tend to expect a sales-pitch and not a serious and legitimate survey.

Telephone interviewers are more limited than face-to-face interviewers to convey legitimacy. A telephone interviewer can only use auditory communication, and the message should be very short to prevent break-offs. When random digit sampling is used, this precludes advance letters.

We conducted a series of split-run experiments in the Netherlands testing the efficacy of the "we are not selling" message in telephone introductions. To establish legitimacy at the outset we used the following introduction: Good-evening/morning/afternoon, this is from..... We are conducting a survey on... WE ARE NOT SELLING ANYTHING. To make sure we get a SCIENTIFIC sample, I would like to speak to (next birthday-selection method is then used). The message "We are not selling anything" is reinforced by the words "scientific sample" to convey that a serious, legitimate survey is being conducted.

Data were collected at 10 market research firms, all members of the Netherlands Association for Market Research. The topics varied in saliency and are a fair representative of the studies undertaken by the research institutes. In total 98.214 persons were CONTACTED, of which a random half heard the experimental introduction and the other half the firms' standard introductions (which is without "not selling" and "scientific sample"). At each market research firm a senior researcher completed a questionnaire, providing background data on each survey.

On average, the 'not selling'-introduction reduced the refusals with 2%. More details will be presented, and special attention will be given to interaction effects with important variables, such as topic of survey, saliency, and sample.

Developing New Models of Interviewer Training

Donald Camburn, Carol Gunther-Mohr, and Judith Lessler, Research Triangle Institute, USA

Nonresponse to household surveys is receiving substantial attention in the survey research literature. While interviewer training is cited as a key factor influencing nonresponse rates, the skills needed to solicit participation conflict with the skills needed to administer highly standardized questionnaires successfully. In addition, typical interviewer training programs devote only limited time

to the development of the skills needed to maximize respondent cooperation. The literature on training survey interviewers offers negligible insight into using interviewer training as a tool for reducing nonresponse to household surveys. To address this problem, a new paradigm is proposed for training interviewers to conduct face-to-face household surveys. While these models of training is designed to address nonresponse, it has the potential to reduce survey costs by decreasing overall training time and eliminating travel to central training sites. Because these training methods are technology-assisted, it is possible to evaluate the effectiveness of training programs and to validate training by testing trainees, carrying out cost-benefit analysis, and using experimental designs to assess outcomes.

Thursday, 8:30-10:00

Mode Effects

The Impact of Mode of Data Collection on Response Rates

Roger Tourangeau and Darby Miller Steiger, The Gallup Organization, USA

Most research on the differences among modes of data collection has focussed on the effect of mode on responses to the questions. Mode can, however, effect both the overall level of nonresponse and the characteristics of respondents and nonrespondents. This paper will reexamine the literature on mode effects from the vantage point of nonresponse. It will examine how the mode of data collection has affected overall response rates, response rates by subgroup, and item nonresponse rates.

Four main hypotheses will guide this literature review. The first is that modes of data collection that offer greater privacy (for example, modes that allow self-administration) produce higher response rates than those that are more public. The effect of privacy will be more apparent when the topic of the survey is sensitive and when the respondent is more likely to be embarrassed by his or her answers to the questions. For example, whether the questions are self-administered will have more impact on response rates in a survey on sexual behavior than in one on political attitudes. Moreover, the mode of data collection for a sex survey will have more impact on the response rates of single people (who are more likely to have multiple sexual partners) than of married people. The second hypothesis to be examined is that impact of the length of the questionnaire on response rates varies by mode. The length of the questionnaire will matter least in face-to-face interviews and most in mail surveys. The third hypothesis is that computer administration of the questions (as in audio CASI) will increase response rates among groups with greater experience with computers (such as the highly educated) but will reduce them among those with less prior computer experience (older, less educated persons). Finally, we will examine the hypothesis that computerization of the data collection process reduces rates of missing data and that self-administration increases them.

The paper will also present results from a new experiment that compares two modes of data collection by telephone (CATI and Interaction Voice Response, or IVR, in which recordings of the questions are played to the respondents) with a mail survey. The analysis will examine both unit and item nonresponse rates across the three modes. The questionnaire will consist of items drawn from the decennial long form. We anticipate that, because of the length of the questionnaire, both IVR and CATI are likely to produce higher "breakoff" rates (in which sample persons start the interview but fail to complete it) than the mail survey; on the other hand, the rate of complete nonresponse is likely to be higher in the mail condition than in the two telephone conditions. The demographic makeup of the three groups is likely to differ as well.

An Approach to Improve Validity of Response in a Sexual Behavior Study in a Rural Area of China

Hongjie Liu and Roger Detels, University of California, Los Angeles, USA

A survey of sexual behavior was conducted among rural residents of Anhui, China in 1997 to determine if using a tape recorder with earphones to administer sensitive questions increased the acceptance and accuracy of the survey. Validity of the responses was evaluated by determining the internal consistency of responses to related questions, comparing reported sexual activities and frequencies reported by the husband and wife within a couple and by resurveying a cluster. Among the 1,269 individuals selected, 185 were not at home. Most (62%) had migrated temporarily to cities. Among the 1,084 who were at home, 1,057 (97.5%) participated in the survey, compared to 70-80% in similar surveys in China. The percent inconsistency among related questions was < 1%. The discordance within couples was 5% for premarital sex, and 3% for oral-genital and/or anal-genital sexual intercourse, and 19% for coital frequency within the past four weeks. A repeat survey gave similar results. The authors suggest that use of a tape recorder and earphones and an anonymous answer sheet contributed to the very high acceptance rate for this sexual behavior survey and enhanced the validity of the responses. This strategy may enhance the response rate for similar surveys in other countries.

The Effect of Survey Techniques on Response Rate in the Italian Household Budget Survey

Giuliana Coccia, Claudio Ceccarelli, and Fabio Crescenzi, Statistics Italy

In 1997, the National Institute of Statistics of Italy redesigned the Household Budget Survey (HBS).

The revision of the survey is due to a general underestimation of the consumption expenditures. Underestimation was originated by a scarce cooperation of respondents causing a reduction in quality of answers. For this reason the new survey point mainly on field-work and enhancing of survey techniques. So an experiment was performed on some Municipalities. In this three different samples of households were interviewed: a) PAPI sample with local interviewer; b) PAPI sample with professional interviewer; and c) CAPI sample with professional interviewer.

In this paper, after a short presentation of the experiment, the effects in terms of total non-response and item-non-response will be analysed with reference to the three techniques.

Thursday, 8:30-10:00

Hard to Reach Populations

Hospital-Based Supplementation as a Means of Improving Response in a Hard-To-Reach Population

Holly B. Shulman and Chris Johnson, Centers for Disease Control and Prevention, USA

The Pregnancy Risk Assessment Monitoring System (PRAMS) is an ongoing, state-specific, population-based surveillance system of selected attitudes, behaviors, and experiences during the period around pregnancy and the child's early infancy. Study participants are identified off the vital records birth certificate file and sent a questionnaire two to six months after their delivery. The basic methodology for PRAMS consists of a mailed survey with mail and telephone follow-up contacts for nonrespondents. This mixed mode methodology, utilizing Dillman's Total Design Method, yields reasonable response rates in most populations with the notable exception of high-risk, minority women living in urban areas. To address this response problem, a hospital-based data collection activity was developed targeting this particular subpopulation to supplement the standard mail/telephone methodology. With the hospital-based supplementation of PRAMS, targeted women were contacted in the hospital at the time of the delivery of their baby. They were asked to fill out the questions on the survey pertaining to the periods prior to pregnancy, during pregnancy, and during delivery. Contact information was obtained from the mother so that she could be recontacted two to four months postpartum to fill out the questions pertaining to the time after delivery. The basic premise of this methodology is that women will be more likely to participate if they are contacted in person. In addition, by obtaining contact information and having the woman be aware of the study, it is more likely that these women can be successfully contacted for follow-up and will be willing to participate.

Hospital response data for 1993-96 is presented from 5 PRAMS states that implemented the hospital-based supplementation methodology. Response rates for the in-hospital portion of the questionnaire ranged from 71% to 95%. Full response (the in-hospital portion plus the postpartum follow-up portion) ranged from 46% to 70%. Two states, New York and Georgia, have comparable data on mail/telephone response rates for the targeted subpopulation for 1996-1997 when hospital-based supplementation was discontinued. In New York, 74% of the sample responded to the hospital portion of the questionnaire and 57% responded to both parts. By comparison only 49% of the sample responded using the mail/telephone methodology. In Georgia, 83% responded in hospital, 65% responded to both parts of the survey, while only 58% responded using the mail/telephone methodology.

Our experience indicates that for hard-to-reach populations, the delivery hospitalization offers an excellent opportunity to reach new mothers and their infants. A significantly higher percentage of these women participated in the hospital than by mail and telephone contacts. Using the hospital contact at the time of delivery as an opportunity to obtain locating information for future follow-up of the mother or infant provides modest gains in response. Although these results are encouraging, the costs of doing in-hospital surveillance is considerably greater than for conducting mail or telephone surveillance. The cost in dollars and resources must be weighed against the corresponding gains in response rates. Our experiences may be useful in determining the feasibility of such approaches for other research efforts.

Effectiveness of Monetary Incentives and Lottery-Type Incentives in Mail Surveys

Victoria A. Albright, Westat, Inc., USA

Deborah A. Kitchell, The College of William & Mary, USA

Effectiveness of Monetary Incentives and Lottery-Type Incentives in Mail Surveys. - The discussion will cover response rate findings from a recent survey of 6,000 individuals who sought help through the Washington State Employment Service, with different incentives offered in two mailings: the first mailing included \$1 or \$5; and the second mailing included a chance to win a \$100 Wal-Mart gift certificate, with one certificate for every hundredth return.

Call-Screening: What Problems Does It Pose For Survey Researchers?

Robert W. Oldendick, University of South Carolina, USA

Michael W. Link, Research Triangle Institute, USA

Increased availability of caller identification (or Caller-ID) services as well as telephone answering machines and the use of such devices to screen telephone calls poses a serious potential threat to the representativeness of samples in telephone surveys. Using Caller-ID or telephone answering machines as screening devices, respondents can determine the identity of those making incoming calls and choose whether or not to answer the telephone. This poses a three-fold problem for survey researchers. First, those who screen their calls contribute to the nonresponse problem by making it more difficult to reach an increasing segment of the public by telephone. Call screening threatens the representativeness of samples and conclusions drawn from those samples to the degree that call screeners differ significantly from non-call screeners in their attitudes, opinions, and behaviors. Second, Caller-ID services pose a new and unique obstacle in that it is nearly impossible for researchers to know whether or not an unanswered number is a valid household number. This introduces yet another source of possible error in determining valid response rates. Finally, call screening more generally has the potential to increase significantly the practical costs (both in terms of time and money) of conducting survey research via the telephone, by increasing the number of calls it takes to obtain completed interviews.

Using data from five random-digit dialed surveys of the South Carolina adult population, this research examines these propositions. First, we provide estimates of the extent to which Caller-ID services and telephone answering machines are used to screen calls and the demographic characteristics associated with self-reported screening behavior. Next, we look at the effects of call screening practices on the number of call-attempts it takes to complete an interview and the incidence of first time refusals.

The findings indicate that the demographics of those who have Caller-ID services and who use such services to screen calls differ in important respects from those associated with answering machine use. In effect, Caller-ID services have expanded the demographics of the call-screening population and, hence, increased the potential threat of non-response. Perhaps more importantly are the different conclusions one can draw from the self-reports of call screening compared to more objective call

history information. While we find little relationship between self reports of call screening behavior and either the number of attempts it takes to complete an interview or the incidence of first-time refusals, we do find a significant link between increased numbers of answering machine events in call histories and the increased likelihood the case will ultimately result in a refusal case. The conclusions, therefore, are mixed. Although call screening is becoming a widely recognized problem for survey researchers, we still understand little about the actual dynamics at play in such cases.

Thursday, 10:30-12:00

Trends in Household Survey Nonresponse: A Longitudinal and International Comparison

Edith D. de Leeuw, Methodika, the Netherlands
Wim de Heer, Statistics Netherlands

Survey nonresponse, that is, the failure to obtain participation of all sampled units in a survey, is as old as survey research itself. Non-participation in surveys appears to be increasing in both Europe and the United States. However, international comparative data on nonresponse trends are scarce. The international survey on household survey non-response, reported here, started in 1991 and has been regularly updated.

Data on response and nonresponse for government survey agencies are now available for the following 16 countries: Australia, Belgium, Canada, Germany, Denmark, Finland, France, Hungary, Italy, The Netherlands, Poland, Sweden, Slovenia, Spain, UK, and USA. Some countries even provided time series going back to the end of the Seventies and the early Eighties. In addition, data were collected on response related factors, such as survey design, and field work strategies.

In this presentation we focus on two large surveys, the Labour Force Survey and the Family Expenditure Survey. The data show a downward trend in response in several European countries. However, other countries appear to remain stable. When survey design and fieldwork strategies are incorporated, it becomes clear that often these apparently stable figures are 'bought' with an increased fieldwork effort reducing the number of noncontacts.

We relate the nonresponse figures, refusals and noncontacts, to differences in survey design and field work strategies across countries. We end with general response inducing principles and recommendations based on our findings.

Thursday, 10:30-12:00

Poststratification and Weighting Adjustments

Andrew Gelman, Columbia University, USA
John Carlin, Royal Children's Hospital and University of Melbourne, Australia

To adjust for known or expected discrepancies between sample and population, the standard methods are poststratification and weighting. We use two simple examples to illustrate how these two approaches--which are often confused with each other--differ. We then review the unified notation developed by Little for weighting and poststratification and discuss the practical and theoretical difficulties of each approach. Some practical methods for obtaining estimates and standard errors under both methods are discussed. Finally we discuss work in progress on a more general approach to model-based inference that accounts for the design information used in weights and poststratified estimates.

Thursday, 10:30-12:00

Interviewer Effects 1

Interviewer Strategies and Attitudes

Lilli Japec and Peter Lundquist, Statistics Sweden

To assure quality in the data collection process in interview surveys at Statistics Sweden a project was initiated that will look at different aspects of data collection e.g. interviewer training, collection of process data and call scheduling strategies.

In the past ten years nonresponse rates have increased in interview surveys at Statistics Sweden. It is mainly noncontact rates that have increased while refusal rates have remained quite stable.

At Statistics Sweden there are 122 field interviewers located across the country (interviewing from their homes) and 30 interviewers located in a centralised telephone interview facility. Normally the interviewer has some information about the sampled person before making her/his first call attempt. How this information is used varies among interviewers. We believe that interviewers are to some degree guided by their personal work experience, and that they develop their own informal rules on how to plan their call attempts.

In this paper we will analyse data from: A questionnaire that was sent out in May 1998 to all field interviewers at Statistics Sweden. The questions concerned interviewers contact strategies and attitudes. The questions on interviewer attitudes is also part of the International Interviewer Questionnaire coordinated by Joop Hox, University of Amsterdam.

The Swedish Labour Force Survey (LFS) from January-February 1997 and June 1997-June 1998. It is a monthly panel survey with a sample of 16 000 persons per month. The mode of data collection is telephone and the interview takes about 5 minutes to complete.

The Swedish Survey of Living Conditions (SLC) from 1997. The survey is carried out every quarter and the sample size is 6 000 persons per quarter. It is a face to face interview that takes about 1 hour.

We will use multilevel analysis to combine information on interviewer level from the questionnaire and on respondent level from the LFS and SLC.

The aim of the study is to identify and measure critical components in the call scheduling and interviewing process that cause a lot of variation in noncontact and refusal rates.

The Effect of Interviewer Characteristics on Survey Response Rates

Jean Martin and Roeland Beerten, Office for National Statistics, UK

Social Survey Division (SSD) of the Office for National Statistics (ONS) has carried out a survey of interviewer characteristics and attitudes in relation to response as part of the international study being coordinated by Joop Hox. Preliminary results from the survey were presented at the 1998 International Workshop on Household Survey Nonresponse(1).

This paper examines the effect of interviewer characteristics, attitudes, expectations and behavioural strategies on interviewer response rates, both contact rates and co-operation rates, taking account of features of survey designs and area characteristics which also affect response.

More specifically we address the following questions:

- a) Are there differences in response (contact or co-operation) between interviewers due to differences in demographic characteristics and/or length of experience?
- b) Do interviewer attitudes or behavioural strategies affect either contact or co-operation rates, controlling for any differences due to demographic characteristics and length of experience?
- c) How much of the variation in response rates can be explained by interviewer attributes and how much by area characteristics and survey design features?

The survey covered all SSD interviewers carrying out in-home interviews on household surveys: around 150 working exclusively on the Labour Force Survey and around 380 working on other surveys. We have response rates for each assignment for each interviewer over a year and examine response rates for 6 different continuous surveys (including wave 1 of the LFS). The surveys differ very substantially in survey design features which are reflected in very different average response rates – from 80% to the LFS for which one adult can answer on behalf of other adults in the household to 62% for the Family Expenditure Survey which requires each adult in the household to be interviewed and to keep a 2 week diary of all expenditure.

Most interviewers work in a variety of different areas and response rates differ significantly between areas. Thus both survey and area characteristics need to be taken into account in determining the effect of interviewer characteristics on response.

We plan to use a multi-level modelling approach to the analysis, with interviewers at the higher level, interviewer assignment at the lower level. We will include information about interviewers (demographic characteristics, experience, attitudes, behaviour etc.) at the top level and survey identifiers and area information derived from census small area statistics at the lower level. Multi-level modelling will allow us to determine how much of the variation in response rates is accounted for at different levels and how much is explained by the variables included at the different levels.

Not only will the analysis described allow us to gain more insight into interviewer attributes which affect either contact or co-operation rates, but information about the relative contribution of survey and area differences could be very valuable in judging interviewer performance and setting interviewer targets for particular surveys.

Interviewers Attitude and Response Rates

Tamara Knighton, Jennifer Hubbard, and Sylvie Michaud, Statistics Canada

An international survey is conducted among different countries. The survey gathers data on interviewer attitudes and its impact on non-response. The questionnaire has been designed to measure the interviewer behavior and procedures, the first time they establish contact with a respondent. In Canada, the questionnaire has been administered last February to interviewers participating to the National Longitudinal Survey of Children and Youth. The survey is a longitudinal panel that has been implemented in 1994. A sample of over 20,000 children of 0-11 has been selected and they have been traced and interviewed every two years. In 1998, a new large sample of more than 10,000 children has been selected from an administrative source. The attitudinal survey will be related to those sets of new interviews. However, in most cases, the same interviewer workforce has been used for both the new sample and the longitudinal sample. So it may be possible to see if the factors that are important to explain non-response in a first contact also show up as contributors to attrition.

The paper will first describe the National Longitudinal Survey of Children and Youth, giving an overview of its design and content. The interviewer attitudinal survey questionnaire will then be presented. Some summary results from that survey will be presented, in the context of the NLSCY. The third section will provide information on response rates and how it is linked to different aspect of the attitudinal survey. A comparison will be made between response rates obtained from the new sample and the longitudinal cohort. Some characteristics of attritors will be examined in more details to try to see if attrition is affected by interviewer attitude in general. Some conclusions will be drawn, along with a presentation of future research plans.

Thursday, 10:30-12:00

Estimating Nonresponse Rates

The Effect of Whole Household Nonresponse on Dual System Estimation in the Census 2000 Integrated Coverage Measurement Survey

Richard A. Griffin and Robert D. Sands, U.S. Bureau of the Census

Census coverage measurement surveys, like all surveys, are subject to nonresponse of various types. Nonresponse due to whole household non-interviews has the most potential for introducing bias into estimates since no information is obtained about the household or its residents. For Census 2000, the Post Enumeration Survey (PES) results will be used in conjunction with the Census counts to produce Dual System Estimates (DSE) of the population that eliminate the differential undercoverage, in particular the longstanding disparity in census coverage of racial and ethnic groups. In the 1995 Census Test in Oakland California a low initial response rate of 86% was augmented by a nonresponse conversion operation that successfully converted about 60% of the nonresponse cases and achieved a final response rate of 94.4%. For Census 2000, PES interviews will be attempted for six weeks and the goal is 100% response. However, at the end of this activity there will be some nonresponse cases. A formal nonresponse conversion operation will then be conducted for two weeks using the best available enumerators. All nonresponse cases are included in the nonresponse conversion operation and we expect to obtain an overall national average response rate of 95% with local variation.

Nonresponse increases the error of DSE in two ways: (1) reduced total number of interviews increases the variance of estimates; and (2) nonrespondents are believed to be different than respondents which adds bias to the estimates. This paper will use nonresponse conversion information from the 1990 Census PES, 1995 Census Test Integrated Coverage Measurement (ICM) Survey, and the 1998 Census Dress Rehearsal (ICM and PES) for a sensitivity analysis on mean square error of DSEs as a function of initial response rate after the first six weeks of interviewing, percentage of nonrespondents successfully converted, and difference in Census/PES match rates between respondents and nonrespondents.

The production nonresponse adjustment for PES spreads the weights of non-interviewed units over interviewed units in the same block cluster and type of basic address (single unit, multiple units, other). There are collapsing rules if the number of interviewed units is too small compared to the number of non-interviewed units. This paper will use multivariate logistic regression modeling to determine significant correlates of capture in the census. Alternatives to the production nonresponse adjustment using results from this modeling are discussed and considered using simulations from both the 1995 and 1996 Census Tests.

Estimating Response Rates in Random-Digit Dialing Surveys that Screen for Eligible Subpopulations

*Trena M. Ezzati-Rice, Victor G. Coronado, and Robert A. Wright, Centers for Disease Control and Prevention, USA
Martin R. Frankel and David C. Hoaglin, Abt Associates Inc., USA
John D. Loft, Research Triangle Institute, USA*

The response rate is a primary measure of survey quality. The purpose of this paper is to compare methods for calculating response rates for Random-Digit-Dialing (RDD) surveys that require screening households for an eligible population. One objective is to help ensure greater comparability in estimates of response rates for this type of survey.

An important component of overall response rates in RDD surveys is the estimate of the eligibility rate among sampled telephone numbers that cannot be contacted and screened during the field period. As suggested in recent recommendations published by the American Association for Public Opinion Research (AAPOR), depending on survey design specifications, it may be important to distinguish between final dispositions for known households that could not be screened for eligibility and dispositions for those cases where residential status is unknown. Separate estimates of eligibility rates may be necessary for each of these final dispositions.

This paper extends an approach suggested by Massey for estimating the response rate in RDD telephone surveys with screening. That paper suggested an alternative to calculating overall response rate as the product of the screening completion and the interview completion rates, using as an example the telephone call outcomes from the first three quarters of the National Immunization Survey (NIS). Currently in its fifth year of data collection (1998), the NIS measures immunization coverage of children 19 to 35 months old nationally and in each of 78 State or metropolitan areas. Because the target population for this RDD survey--households with children 19 to 35 months old--represents only 4% to 5% of all households in the U.S., extensive screening is required to identify eligible households. Beginning with the first quarter of the survey, the observed eligibility rate has been lower than the estimated eligibility rate by a factor of about 20%. Massey suggested that the lower eligibility was due to under reporting of eligibility by household respondents. Calculating response rates using the eligibility rate observed in the NIS resulted in an overestimate of response rates. As an alternative, he proposed a formula for overall response rate--ORR--that uses an estimate of the household eligibility rate (later designated as the eligibility benchmark) derived from sources external to the NIS. The eligibility benchmark is used to estimate the number of eligible households in each quarterly sample nationally and in each of the 78 areas in order to calculate overall response rates nationally and for each area.

This paper reports how this approach has been implemented in more recent NIS data collection years. In particular, the eligibility benchmark has been updated on an annual and quarterly basis to account for national trends in birthrates and household formation. The paper compares ORR with other estimates of response rates to illustrate the sensitivity of response rate calculations to various assumptions about the proportion of eligible households in the sample. As the survey research profession searches for standardized approaches to response rate calculations for complex surveys, the issues described above will have important applications.

Methods for Calculating Nonresponse Rates in the National Health Interview Survey

Frances M. Chevarley and Richard Coles, National Center for Health Statistics, USA

The National Health Interview Survey (NHIS) m estimation, match rate) oversampled Blacks and Hispanics in its 1995-2004 sample design in order to increase the sample size for the Black and Hispanic populations. One way the sample design ensures a larger number of Black and Hispanic Households is by having higher sampling rates in areas with larger proportions of Blacks and Hispanics. Another way the sample design ensures more Black and Hispanic Households is by designating some households as "screener" households. For the "screener" households the interview is completed only for those households with Black and/or Hispanic members. For the regular "interview" households, the interview is completed for all of the households. In this paper unit nonresponse is investigated by looking separately at nonresponse rates for the "screener" and "interview" households and by making assumptions concerning the nonresponding "screener" households. Before the 1995-2004 sample design, all of the nonresponding households were used in the nonresponse rate calculations because "screener" households were not used in previous NHIS sample designs.

As a first step, assumptions concerning the nonresponding households for the "screener" households are investigated and used to estimate the numbers of the nonresponding "screener" households thought to contain Black and/or Hispanic members. For the second step, nonresponse rates are calculated for the "interview" and "screener" households separately, before combining these rates to produce nonresponse rates for all of the households combined. Nonresponse rates are calculated based on the different assumptions that were investigated concerning the nonresponding households for the "screener" households in the first step. The calculation of nonresponse rates calculated for the whole United States will be compared with those calculated initially for lower sampling areas and then aggregated to a nonresponse rate for the whole United States. Trends in nonresponse rates would also be investigated for the different nonresponse rates that are calculated. Because nonresponse rates play a critical role in post-survey adjustments, the effect of the different nonresponse rates on the NHIS nonresponse weighting adjustments and ultimately on the NHIS data estimates, time permitting, is also being planned.

Thursday, 10:30-12:00

Empirical Calibration

The Use of Administrative Registers to Reduce Nonresponse Bias in Household Surveys

Linda Geuzinge and Bart F.M. Bakker, Statistics Netherlands

Nonresponse is a severe problem in social statistics and social surveys. It makes survey estimates questionable, because of a potential and hardly measurable bias. It is a largely unknown factor and it is not known to what extent post-survey adjustment can deal with nonresponse bias. In recent years a growing concern developed with respect to this problem. From an international survey on nonresponse it appeared that response rates for the Netherlands are very low if compared to other countries. For instance, the Dutch Labour Force Survey reached a response rate of 60% in 1995, while the response rate for Labour Force Surveys in other European countries varied between 82% and 93% that year. The response rate for the Health Survey in 1995 was 58%.

The question remains whether a low response rate implies unreliable results. As the nonresponse error is a function of the nonresponse rate and the difference between average scores among nonrespondents and respondents, increasing response rates do not always reduce the nonresponse error. Improvement of response rates might not be enough. Reducing the nonresponse bias must be the ultimate goal.

The nonresponse bias can be reduced by smart methods of weighing. In the paper we will present such a method and the method will be tested empirically. The method contains the following steps. Firstly administrative registers with data on the whole population are collected. These administrative registers are about jobs, social security benefit and the Dutch population register. Secondly, these registers are exactly matched to each other and to the response and nonresponse of the particular survey. Now we have the disposal of information on nonrespondents. This is particular because in nonresponse-analysis before we did not have this kind of information about people who did not reply. For example, we know of someone has a job or not. From the literature we know that people without a job most of the time have a worse state of health or a higher medical consumption than people with a job do. At the same time it is possible that the people with a job are fewer at home when the questioner will interrogate the respondents for the survey. They can be selective for the response. Thirdly, with the use of principal component analyses, the most important variables in the survey are combined in a few factors. Fourthly, we search for variables in the registers that correlate high with those factors and are associated with the chance of response. We will ascertain if people who did not answer are different of the people who did answer on the basis of certain auxiliary variables from the registers. When this is the case we will investigate the cohesion between these auxiliary variables and the target variables. When it comes to cohesion we can conclude that selectivity of the nonresponse assign the parts in the Health Survey. Fifthly, we use the variables from the registers which satisfy both conditions to weigh the response of the survey. We make a correction for selectivity of nonresponse in the Health Survey also. The first results will be presented at the conference.

A Comparison of Household Grossing Methods and Their Effect on Nonresponse Bias

Denis Down and Dave Elliot, Office for National Statistics, UK

Non-response in most household surveys is in units of whole households, so that it makes sense to apply any differential weighting at the household level. This means that the grossing factors for all the individuals in a household are the same. Control totals, however, are often numbers of persons categorised by age and sex. Methods have been developed to match the survey totals to such control totals using household grossing factors. Most of them work by minimising a distance function between initial weights, which may be all the same, and the final weights. The methods are described as "calibration" weighting.

A different approach to matching population totals using household weights has been developed for the Survey of English Housing (SEH), which is sponsored by the Department of the Environment, Transport and Housing in England. The method categorises households according to the age of the youngest person and moves progressively from households with young children up the age ranges. The author has named the approach as the "cascade" method. In its application to the SEH it is elaborated to deal

separately with young adults living with older people (typically their parents) and those living alone or with other young people. Young adults are under-represented in the SEH, as in most surveys. The survey accepts proxy data for household members who are not present during the interview, so it is young adults living alone or with other young people who will be under-represented, rather than those living with their parents. The elaboration allows this distinction to be made. A separate stage is used to achieve the correct numbers by sex as well as age. There is also an initial weighting based on the number of calls needed to make the initial contact with the household, a measure of how hard it was to find anyone at home.

The paper discusses the advantages and disadvantages of the calibration and cascade methods, referring particularly to the SEH. Advantages of the cascade method are that:

- it is simple to implement using only spreadsheets;
- it is relatively transparent in its behavioural implications;
- it can incorporate explicit assumptions on patterns of non-response;
- it has not proved necessary to impose limits to avoid negative weights;
- it works more strongly than the calibration method to compensate for
- non-response bias on variables that are important for the SEH.

Advantages of the calibration method are that:

- it is better suited to matching populations by a large number of age bands;
- none of the grossing factors are calculated as residuals, as happens in the cascade method;
- as a result extreme or unstable values can be avoided more easily.

The Geographic Dimension in Adjusting Household Surveys

Laszlo Mihalyffy, Hungarian Statistical Office

Adjusting techniques ensuring agreement between adjusted estimates from household surveys and the corresponding actual values coming from some exogenous source of information are widely used by statistical agencies in different countries. The purpose of adjustment is usually to compensate for non-response or coverage error, or possibly for both. A wide class of adjustment techniques combines demographic information on individuals with total number of households of which the individuals in consideration are the members. Since in most cases national surveys are considered, an obvious third dimension in adjustment is represented by some geographic breakdown. Adjustment schemes based on these principles are used in the consumer expenditure surveys in the United States and in Italy, and, among other things, in the Labour Force Survey (LFS) in Hungary. In these applications, adjusted sample weights are computed by the method of least squares, generalized iterative scaling or by Newton method.

If geographic breakdowns in adjustment schemes are defined in terms of administrative units - which seems to be an obvious strategy -, the differences in response rate within those units are ignored. In adjusting the Hungarian LFS, e.g., the geographic units are the capital city and the 19 counties, though within the latter, big cities have definitely lower response rate than rural areas. This implies among other things that for a county with half-a-million inhabitants of which 200,000 live in the county seat, adjusted sample estimate of total population agrees with the corresponding actual value (which is an updated census count), while that of the county seat amounts only to 170,000. The goal of the paper is to suggest a better strategy for specifying area units in adjustment schemes.

Thursday, 10:30-12:00

Characteristics of Nonrespondents

Type and Extent of Nonresponse in Customer Satisfaction Research and Strategies to Compensate for Nonresponse

Rajesh Srinivasan and Steven Hanway, The Gallup Organization, USA

Non-response, be it item non-response or at the unit level, has been and will continue to be an inevitable part of survey research. Recent trends in response rates raises its significance in terms of the resulting bias. While most government research tries to reduce non-response either by design or through a substantial effort by way of callbacks, commercial research is and will always be at the mercy of financial and time constraints that interfere with the objective of high response rates. The resulting impact on survey estimates is rarely if ever the subject of research.

Traditional methods of estimating non-response bias involve additional data collection efforts by sampling non-respondents, while adjustment for non-response may involve collecting frame-related information on non-respondents. While face-to-face surveys enjoy the luxury of better frame information, albeit at a higher cost, in Random Digit Dial (RDD) surveys frame information is mostly limited to geography such as exchange, zip code or Census tract. Surveys of customers on the other hand have the potential to be rich with information both on respondents and non-respondents, information that goes beyond simple geography. From that perspective, customer satisfaction surveys are a good candidate for understanding the type and extent of non-response and subsequently designing strategies for non-response adjustment.

Gallup currently conducts customer satisfaction surveys for companies in telecommunications, financial services, retail, entertainment and health care. Our objective in this research is three fold: to do a comparative analysis of both the type and extent of non-response that is pervasive across industries as well as within any given industry; to evaluate the application of non-response adjustment strategies on key survey estimates; and to include the effort required to increase response rates in a cost model so as to be able to conduct sensitivity analyses.

Nonresponse in the Telepanel of CentERdata*Juan Felix and Dirk Sikkel, Tilburg University, the Netherlands*

In 1997, CentERdata, a research institute at Tilburg University, has initiated the so-called CentERdata Quality Project. As a participant in this project, I am currently investigating the nonresponse in the telepanel of CentERdata. This telepanel consists of approximately 2000 households in the Netherlands who are provided with a PC and modem at their home. Every week, the household members receive a questionnaire by modem and phone.

Nonresponse in the telepanel of CentERdata occurs at different stages. First, during the recruitment stage of new members (initial nonresponse). Second, between subsequent research waves (wave nonresponse). Third, when panel members leave the panel (attrition). Here we will focus on the first type of nonresponse. The central question of our research is: what are the characteristics of initial nonrespondents in a panel survey? The answer to this question will provide us with information for the correction of nonresponse bias.

To study the characteristics of the nonrespondents we use data from subsequent telephone interviews (cati). The data were collected during the recruitment of new panel members. In the recruitment stage of new panel members we can discern four different steps. First, respondents are interviewed by telephone (cati). Second, 'potential' panel members are stored in a basic 'sign up' file. Third, on a weekly basis new panel members are selected from the basic sign up file. Fourth, the selected potential panel members receive an introduction letter and are interviewed by phone (cati) for a second time. The data used in this paper are based on the telephone interviews that were conducted during the first and last step.

Besides the 'basic' background variables like age, gender, income, household size and political preference we have collected a lot of additional information. These are for example home-ownership, number of rooms in the house, paid job or not, traveling time from home to work, number of working hours per week, number of sick days during last three months, suffering from chronic disease, estimation of personal health condition, visited theater during last twelve months, visited cinema during last twelve months, membership of a sports club, ever been victim of a burglary, afraid at home or not in the evening, and afraid in own neighborhood or not.

On the basis of the collected data we were able to examine whether respondents who do not intend to join the panel are atypical. We have used bivariate and multivariate analyses with panel membership as the dependent variable. The results of these analyses will be used to correct for the nonresponse bias. If the initial nonresponse is selective, one of the possible corrections for the bias is weighting the sample of panel members with poststratification.

The Effects of Differential Unit Response Rates Across Plans in a Health Plan Satisfaction Survey*Sherman Edwards and John Rauch, Westat, Inc., USA*
Alan Zaslavsky and Paul Cleary, Harvard University, USA

The Medicare Consumer Assessment of Health Plans Survey (MCAHPS) is a mail survey with telephone follow-up, asking Medicare beneficiaries enrolled in HMOs about their experiences and satisfaction with their health plans and providers. In the Fall of 1998, surveys were sent to 600 beneficiaries in each of 314 health plans. Sampled individuals received a prenotification letter, an initial survey mailing, and a thank you/reminder post card. Nonrespondents to the first mailing were sent a second questionnaire, and those still not responding were contacted by telephone to attempt to complete the survey or, if a telephone number was not available, received a third survey by Federal Express. All nonrespondents were called at least three times with some refusal conversion attempted. The overall response rate for MCAHPS was 78 percent, with individual plan response rates ranging from 36 to 95 percent.

After the data collection period, a sample of respondents and a sample of nonrespondents were administered a 10-question subset of the CAHPS instrument by telephone. The target was 1,000 completed interviews in each cell. The sample was allocated equally between plans in the top and bottom response rate quartiles.

While the overall survey response rate was good, the variation in response rates by plan is worrisome. The data collection strategy was to treat all respondents the same in number and types of contacts rather than attempting to make plan-level response rates more comparable by concentrating follow-up efforts on plans with lower response rates.

This paper examines the differences between early respondents and late respondents, and between respondents and non-respondents, in terms of both demographic characteristics and responses to selected satisfaction measures. Both bivariate and multivariate methods will be used. The central research question is, do differences in response rates across plans result in differential nonresponse bias? If so, should more efforts at obtaining cooperation be expended for low response rate plans than for those with higher response rates?

Thursday, 1:30-3:00**The Causes and Consequences of No Opinion Responses in Surveys***Jon A. Krosnick, Ohio State University, USA*

Most approaches to the analysis of survey data treat non-response as a problem that requires repair, so significant efforts have been devoted to designing techniques for minimizing non-response, for imputing missing data, and for weighting data to compensate for interviews never conducted. Remarkably, though, many survey researchers have believed that there is one type of non-response that should be encouraged rather than discouraged when measuring subjective phenomena. According to number of important textbooks on survey methods, respondents should be offered -- a "no opinion" or "don't know" option so as to discourage reporting of what Converse (1964) called "non-attitudes." During the 20th Century, a great many studies have examined the causes and consequences of item non-response in the form of "don't know" or "no opinion" answers, and this paper provides a complete

review of all this work. Numerous studies show that most people who select "don't know" options would instead have offered meaningful, reliable, and valid opinions if encouraged to do so by omitting the "no opinion" response option. Thus, people do not seem to say "don't know" because they lack substantive opinions. Rather, they do so because of a number of other factors, including most importantly survey satisficing and the desire to avoid presenting themselves in socially undesirable ways to researchers and/or interviewers. It is very rare for respondents to say "don't know" when answering typical survey questions because they truly have no information with which to form a judgment. Therefore, in order to maximize data quality in surveys on subjective phenomena, item non-response should be discouraged, just as we attempt to minimize other sorts of non-response as well.

Thursday, 1:30-3:00

Using Administrative Records to Impute for Nonresponse

*Elaine Zanutto, University of Pennsylvania, USA
Alan Zaslavsky, Harvard University, USA*

Administrative records are a relatively inexpensive source of detailed information that have a variety of statistical uses, especially as technology facilitates manipulation of increasingly large datasets. In particular, administrative records can be used to impute for both unit and item nonresponse. Because they provide detailed coverage at relatively low cost, records are also particularly useful for small-area estimation.

Administrative items may be closely related to survey items, but the records may differ subtly in coverage, content, and reference period from the survey. Consequently, simply substituting administrative records for the missing responses can introduce bias into the survey estimates. Instead, we propose modeling the relationships between respondents' survey responses and their administrative data. In this process, the latter serve as covariates for prediction of the values that would have been obtained on the survey. This estimated relationship can then be used to impute values for the nonrespondents.

We begin with a general discussion of administrative records, focusing on their strengths and weaknesses for statistical uses. We then give a theoretical framework for using them to impute for survey nonresponse, drawing on general concepts of imputation for missing data. We conclude by applying these concepts to a method for imputing the characteristics of nonrespondents to the U.S. decennial census.

Thursday, 1:30-3:00

Nonresponse in Web Surveys

Vasja Vehovar, Zenel Batagelj, Katja Lozar, and, Metka Zaletel, University of Ljubljana, Slovenia

Nonresponse and noncoverage are the most critical features of Web surveys, particularly when targeting the general population. Frequently, in fact, just a few percents of target population have a nonzero probability for the inclusion into the sample. In addition, the response rates are often extremely low. The Web surveys thus radically increase the gap between the art of designing useful survey and the scientific approach to sample surveys. This presents, no doubt, an additional challenge to the survey profession and particularly to its standards.

For proper discussion, however, the nonresponse problems in Web surveys must be decomposed into three components: the Web survey mode, the related selection/solicitation procedures, and the specific technological limitations that may accompany this survey mode.

The authors present an overview of the research on nonresponse in Web surveys. In particular, they discuss the framework for participation in Web surveys, the role of selection procedures, the impact of technological limitations, and the importance of the design features of Web questionnaires. Also addressed are the specifics of the so-called validation problem.

Thursday, 1:30-3:00

Follow-Up Studies

Telephone Reminder Calls and Mail Survey Response Rates

Edward P. Freeland and Peter Furia, Princeton University, USA

Our study compares unit and item response rates in a mail survey that includes telephone reminder calls for a subset of cases. Unlike previous studies that examine response rates for mail-only and telephone-only surveys, our experiment seeks to shed light on the effectiveness of reminder calls for mail surveys.

Our study utilizes a simple random sample of registered US voters (N=1000), which provides us with personalized addresses and listed telephone numbers for each individual in the sample. An 80-item, 8-page TDM questionnaire on local and global affairs is mailed to each sample member. The initial mailing is followed by a reminder postcard one week later, and a second copy of the questionnaire after three weeks.

Three days after the second mailing of the questionnaire, the non-respondent sample with telephone numbers is split into two groups. One group receives a "reminder" call in which sample members are encouraged to complete the questionnaire and return it by mail. The experiment is designed to test the marginal utility of supplementing the TDM procedure with reminder calls for households with listed telephone numbers.

In keeping with TDM procedures, a final questionnaire and request is sent to all remaining non-respondents approximately seven

weeks after the initial mailing. At this point, our sample is again subdivided into "cross-cutting" categories based on the type of outgoing postage used in the final mailing. Per TDM recommendations, one half of these remaining individuals are sent the questionnaire via a \$3 Priority Mail envelope, whereas the other half receive it via regular postage and the same type of envelope as before.

In sum, we compare unit and item response rates across the following four groups:

Group 1: Questionnaire 1; Postcard; Questionnaire 2; Questionnaire 3

Group 2: Questionnaire 2; Reminder Call; Questionnaire 3

Group 3: Questionnaire 2; Questionnaire 3 (priority)

Group 4: Questionnaire 2; Reminder Call; Questionnaire 3 (priority)

Our experiment tests whether Groups 2 and 4 each achieve higher response rates than Group 3 (the standard TDM mail survey group). In the case of Group 4, we expect this difference to be significant, representing a best-case scenario in terms of "mixed-mode" survey response. In the case of Group 2, we expect to demonstrate a more economical telephone follow-up option that equals or outperforms standard TDM not only in unit and item non-response but also in cost. The survey will be conducted in early Spring 1999.

Can the Use of a Second Mailing Increase Mail Response Rates in Historically Hard-To-Enumerate Census Tracts?

Erin M. Whitworth, U.S. Bureau of the Census

The U.S. Census Bureau conducted a dress rehearsal in 1998 in Sacramento, CA and in selected counties surrounding Columbia, SC. In areas that rely on the U.S. Postal Service to deliver census questionnaires, a second questionnaire was delivered to all households in an attempt to improve the mail response rate. Previous research by Dillman, Clark, and Sinclair (1995) shows an increase in the mail response rate (as much as 10 percent) in the presence of a second mailing.

This paper will explore the effectiveness of a second mailing in historically hard-to-enumerate Census tracts. Hard-to-enumerate tracts are identified by using a database developed by staff at the Census Bureau. Bruce and Robinson (1998) have demonstrated the value of such a database in predicting rates of mail response. Each tract is assigned a score indicating its enumeration difficulty. This score is based on a summary of socio-economic variables that have been shown to be correlated with non-response and difficulty in enumeration. These variables include percent of housing units with no phones, percent in poverty, percent on public assistance, percent unemployed, percent linguistically isolated and the like. These socio-economic variables, together with demographic variables, were obtained from the 1990 Decennial Census.

The value of targeting Census tracts to receive a second mailing is assessed in this paper. Increasing mail response in such areas can help to bridge levels of differential mail response which may in turn have an impact on improving data quality including coverage. The potential effects on levels of item non-response are also examined.

Nonresponse in Urban, Inner City Samples: Implications for Health Research

Joyce Moon Howard, Angela A. Aidala, Ann Brunswick, Jack Elinson, Columbia University, USA

Urban residence is one of the most universal correlates of nonresponse in surveys. Urban dwellers, especially those in central city areas, are more likely to be lost to survey due to non-contact or refusal. This poses an especially serious threat to health surveys and epidemiological studies. Health conditions, as well as risks for poor health outcomes are not uniformly distributed throughout urban communities, but are concentrated in segments of the population whose living arrangements (e.g. high crime areas) or lifestyle patterns (irregular work schedules) increase the likelihood of nonresponse. Thus rates and proportions on variables of interest can differ significantly among those included and not included in our study.

While relatively high rates of nonresponse among central city residents have often been noted, survey findings have seldom been adjusted for the missing data due to the above described urban nonresponse. What is the likely bias in prevalence estimates? Are there differential patterns of association between risk and outcomes between those included and not included in our survey which could affect our understanding of etiology of disease as well as of its course and its social and biological consequences?

This paper presents an investigation of the potential consequences of nonresponse among minority residents in a central city in the northeastern United States. Data are derived from two separate longitudinal studies that are noteworthy for high sample completion rates: for both studies, over 90% of those interviewed at each study wave were successfully recontacted and reinterviewed. In addition, both studies have kept extensive documentation of field protocols and the effort expended to bring each case to completion. The Young Adult Health Project (YAHP) is an ongoing investigation of health and health risks in a cohort of African Americans who were adolescents in Central Harlem and still reside in New York City (n=364 at 4th wave in 1990). The second study is the Community Health Advisory & Information Network (CHAIN) project which is a longitudinal study of need for services, service utilization, and quality of life of a representative sample of individuals living with HIV/AIDS in New York City (n=420 at 4th wave in 1997).

Two analytical approaches have been used. In a series of cross sectional analyses the sample has been divided into those who were reached relatively early in the study period, and those who were especially "hard to reach" - requiring six months or more of repeated contact attempts. We also compare individuals who easily cooperated with the survey and those who were reluctant or initially refused to participate and were subsequently converted. Our assumption was that the latter individuals would most likely be missed using conventional survey methods which allow for fewer repeat visits, shorter time in the field, and less extensive conversion protocols. In addition, the longitudinal data permitted analysis for predictive factors that mitigated against inclusion in the fourth wave.

Findings indicated that those most likely to be excluded from study were significantly different on certain background characteristics, health related attitudes, health conditions and health risk behaviors. Analyses of these differences expand our

understanding of factors underlying nonresponse and provide guidance for adjustments in both data collection protocols and post survey statistical adjustments.

Thursday, 1:30-3:00

Methods for Reducing Nonresponse

Nonresponse Rate Reduction: Recent Developments

Lars Lyberg, Statistics Sweden

Keeping nonresponse rates on decent levels is important to survey organizations. There are many causes for nonresponse and for each cause the survey designer is referred to a number of "prescriptions" that could be used. The general experience is that each prescription can help decrease the nonresponse rate by a few percentage points only. Therefore a number of different prescriptions have to be administered in order to substantially decrease the nonresponse rate in a survey.

In this paper we will provide an overview of recent developments of various prescriptions or methods to decrease nonresponse rates. Areas that will be covered include measures to increase data protection, offering respondents a choice regarding data collection mode, contents and appearance of advance letters, designing questionnaires and questions so that they seem easy to deal with, follow-up procedures directed at certain segments of the sample, the use of incentives and other measures to reduce respondent burden and the role of interviewers.

We will discuss the possibilities to combine measures across the areas mentioned. Examples will be given of such "battery" solutions. Also measuring key data collection process variables in a continuous way will provide a basis for changes that could lead to decreased nonresponse rates. Examples of process variables that could affect nonresponse rates are provided.

Nonresponse in Time: Some Lessons from Finland

Kari Djerf, Statistics Finland

Increasing nonresponse has been one of the major problems in large surveys organisations for many years. Statistics Finland like many others have been aware of its various consequences on data quality. Therefore we have collected data on the whole sample set from various sources, especially the administrative registers. Those data have been used for analysing the nonresponse bias as well as for weighting.

However, it is not always sufficient to analyse only the results. We have time series of our continuous large surveys since the early 1980's. The surveys include the Finish monthly Labour Force Survey (LFS), the Income Distribution Survey (IDS), and Household Budget Survey (HBS). Our LFS shows fairly modest increase of nonresponse until the mid-1990's when it suddenly increased by 4-5 percentage points. Never-theless the yearly nonresponse level is still not very high, about 13%.

Quite opposite, the HBS is known to be very burdensome for the respondents. Nonresponse rates are generally over 30 per cent and relatively stable despite many field strategy efforts. The nonresponse patterns of the Income Distribution Survey have been much more volatile, ranging from 15 to 30.

In this study we try to evaluate various sources ultimately leading to nonresponse. It will be based on information on the survey climate issues, survey design, and the strategic survey decisions taken by the organisation, data on response burden and the nonresponse structure. We will use both the cross-sectional and the time series methodology to find out "a baseline nonresponse rate" for each survey, and additionally the factors which affect on it. Our analysis will provide us with information how to change the design of those important surveys and fighting against nonresponse. It may also add some new ideas for the international cooperation on nonresponse.

Efforts Aimed at Reducing Survey Nonresponse in a National Health Survey

Trena M. Ezzati-Rice and Meena Khare, National Center for Health Statistics, USA

Robert S. Murphy, Westat, Inc., USA

The success of sample surveys and the quality of estimates from such surveys depends to a large extent on a high proportion of identified sample persons participating in the survey. No statistical adjustment procedure can substitute for efforts to collect survey data with high levels of response. The best guard against potential nonresponse bias in a sample survey is to utilize various field procedures to achieve and maintain high cooperation rates. The National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, conducts several large-scale national population based health surveys. The National Health and Nutrition Examination Survey (NHANES), a periodic national survey conducted by the NCHS, is designed to provide national statistics on the health and nutritional status of the U.S. civilian noninstitutionalized population. The NHANES is unique in that it consists of both a household interview and a 4 hour standardized physical examination conducted in specially equipped mobile examination centers that are transported from one survey location to another. NHANES collects a wide range of medical history, demographic, and socioeconomic information on all selected persons and a range of physical and biochemical measurements on all interviewed persons. The most recent NHANES, the NHANES III, was conducted from 1988 to 1994 and included a number of special procedures designed to maximize response rates. This paper focuses on these efforts to increase responses rates both at the interview phase and examination phase of the NHANES. In particular, this paper will provide a summary of the methods used to initially increase awareness of and participation in the interview phase of the survey including community outreach and publicity, working with local health officials, advance letters, bilingual interviewers, attention to the gender and ethnicity of the interviewers, training and re-training of interviewers including refusal conversion training and repeated callbacks, etc. Additional special measures were incorporated to maximize participation in the examination phase of the survey including selection of multiple persons from each household, a \$50 monetary incentive to all examined respondents, arranging for transportation to the examination center

via the use of local taxi companies or escort service by household interviewers, transportation reimbursements, locating the examination centers within a 50 mile radius of the sample segments at a convenient and safe location, free health information brochures, a home examination especially targeted for the older population, a report of major medical findings to each survey participant, free examination to accompanying family member or friend, etc. Despite these special procedures designed to maximize response rates, NHANES III experienced unit or total nonresponse. Of the 39,695 selected sample persons, 86% were interviewed in their home and 78% were examined in the mobile examination centers. This paper will conclude with a summary of the patterns of nonresponse for both the interview and examination phase of the NHANES III.

Thursday, 1:30-3:00

Weighting and Imputation

An Assessment of the Impact of Non-response Weighting in Sample Survey Data

Jai Won Choi, National Center for Health Statistics, and Balgobin Nandram, Worcester Polytechnic Institute and National Center for Health Statistics

When there are missing data in government surveys, it is common to use weights to estimate a population characteristic. One important example is non-response which causes bias in weighted data. The weighting is often done by inflating each observation by the ratio of the sample size to the number of respondents to make up the non-respondents in the sample. There is no bias when the responses of the non-respondents are the same as those of the respondents in the unit. However, the bias could be a serious problem if the actual answers of respondents differ from those of non-respondents. We show how to estimate the bias caused by ratio estimation and the amount of variance due to this bias. We use an example based on the National Health Interview Survey to illustrate our findings.

Representative Weights and Imputation for the 1997 German ISSP: An Application of the Conditional Minimax Principle

Siegfried Gabler, Sabine Haeder, and Michael Wiedenbeck, ZUMA, Germany

The International Social Survey Program (ISSP) is a continuing annual survey research program on topics important to social science research. It currently has more than 30 member countries. The 1997 ISSP survey in Germany was conducted as a respondent panel of the 1996 ALLBUS (German General Social Survey). However, not all of the ALLBUS respondents were willing to participate in the ISSP survey one year later: The response rate amounted only to about 55%. Thus, the dataset consisting of both surveys remained incomplete. Unfortunately, these unit nonresponses in the ISSP were not at random but systematic (Mohler, Harkness and Wiedenbeck 1998). This may lead to biased estimates for population parameters from the ISSP sample. Our goal is to improve the quality of this dataset by using a model for replacing missing data in the ISSP from ALLBUS responses. As a consequence, estimations based on the completed ISSP data should be closer to the true values in the population than without imputation. Of course, this is difficult to verify. Therefore, we show it indirectly: The success of our efforts will be measured by a check back from ISSP to ALLBUS. We predict the distributions of some variables of the ALLBUS with the help of the imputed ISSP data and compare the "true" and the estimated distributions.

The conditional minimax principle defined by Gabler (1988) yields the decision-theoretical justification for our approach. This method was already successfully used for the construction of individual weights (Gabler and Häder 1998). It was extended by Gabler (1998) to an imputation method which gives the same estimates for the population parameters as the weighting procedure. The results depend on the researcher's a-priori knowledge on relations between the variables. Our method is flexible to incorporate this knowledge.

Adjusting for Wave Nonresponse in School to Work Panel Data

Antonella D'Agostino and Laura Neri, Statistics Italy

Wave non-response is one of the main problems in panel surveys where information is collected from the same sample elements at several different points of time. While weighting are usually used to compensate for total nonresponse, imputation is usually adopted for item nonresponse; however, the choice of a compensation strategy for wave nonresponse is not obvious. It depends on some reference survey factors, such as: the number of wave of missing data, the amount of missing data, the nonresponse pattern, the analysis to be conducted and the availability of auxiliary variables with high predictive power for missing values. In this work the reference sample survey is an Italian panel survey (LEVA) conducted at a local level by Region Lombardia Statistical Office; the project follows several cohorts of students to analyse their transition from school to working life over a period of six/seven years after the compulsory education, for this reason LEVA has a very complex nonresponse pattern. The missing data is made more serious by the high nonresponse rate caused by the mail questionnaires used for collecting data for the intermediate waves. Concerning the analysis to be conducted on the observed data, it is important to specify that the interest is related to the transition from school to working life over a long period after compulsory education.

The attention is focused on the first three waves of the panel survey, and in particular on categorical variables indicating school/not school for each wave. The reference variables present a non negligible amount of missing values, and the rates are different across the waves; so to perform a serious data analysis, it is necessary to compensate for missing values. From a longitudinal perspective wave non-response can be viewed as a set of item nonresponses in the longitudinal record; this idea suggests compensating missing value using an imputation strategy. In building the imputation strategy, we have to bear in mind that the same variable recorded on the same individual in different times are not independent., so the imputation strategy must inherently be multivariate in nature. The imputation procedure makes the best of the available information, so we start the procedure exploring the available data and we go on using a multivariate imputation techniques. The basic strategy is to create imputation through a sequence of univariate logit regressions. The imputations are defined as drawn from the predictive distribution specified by the

above mentioned regression model; the sequence of imputing missing values could be continued in a cyclic way, each time overwriting the previous drawn values in order to preserve the interdependence structure for the variables.

Thursday, 3:30-5:00

The Influence of Interviewer's Attitude and Behavior on Household Survey Nonresponse: An International Comparison

Joop Hox, Utrecht University

In interview surveys, be it by telephone or face-to-face, the interviewer is one of the most important means to improve response. There is considerable empirical evidence for variation in response rates between interviewers. But, what makes interviewers differ?

There is little evidence that interviewer attributes, or personality characteristics, influence response rates. But, there is some evidence that interviewer attitude and behaviour do play a role. Furthermore, there is empirical evidence for differences in response rates between countries.

An international research project was started at the international workshop on household survey nonresponse in Mannheim (1996). Data on the interviewer level are available for 10 different countries and from 32 surveys. The data come from both official statistics and research institutes, and both face-to-face and telephone surveys are included. Country, agency, and interview mode are added as background variables.

Three research questions are central in this international comparison: (1) 'Do interviewers in different countries have different attitudes towards the interviewer role?', (2) 'Does interviewer attitude predict interviewer response rate within and across different countries?', and (3) Does (self-reported) interviewer behaviour add to the predictive power for interviewer response rate within and across countries? The main results will be presented.

Thursday, 3:30-5:00

Weighting Adjustments For Ignorable Nonresponse

Jelke G. Bethlehem, Statistics Netherlands

Unit nonresponse may cause estimates based on survey data to be biased. Therefore, it is vital to reduce the amount of nonresponse in the fieldwork as much as possible. Nevertheless, in spite of all these efforts, a substantial amount of nonresponse usually remains. To avoid biased estimates, some kind of correction procedure must be carried out.

One of the most important correction techniques for non-response is adjustment weighting. It means that every observed object in the survey is assigned a weight, and estimates of population characteristics are obtained by processing weighted observations instead of the observations itself.

This chapter gives an overview of a number of adjustment weighting techniques. It starts by giving the general theoretical framework for sampling based on the Horvitz-Thompson estimator. By introducing models for nonresponse it shows how this phenomenon can effect estimates.

In case of ignorable nonresponse, weighting techniques can reduce the bias of estimates. Some weighting techniques are discussed: post-stratification, linear weighting, and multiplicative weighting. Also, attention is paid to the general calibration framework which incorporates all these techniques as special cases. The chapter concludes with an application of weighting to real survey data.

Thursday, 3:30-5:00

Special Populations and Nonresponse 2

Nonresponse in Surveys of the College Educated

*Ron Fecso, National Science Foundation, USA
Patrick Flanagan, and Katherine Bartley, U.S. Census Bureau*

The Division of Science Resources Studies of the National Science Foundation sponsors surveys of college and graduate degreed individuals. The Census Bureau has conducted one of these surveys, the National Survey of College Graduates. This paper explores the nonresponse patterns of these more highly educated people with nonresponse patterns of the general population. Current trends and directions for future work are discussed.

Nonresponse Patterns Among the US Aged Population: Data From the National Long Term Care Survey.

Larry S Corder and Kenneth G. Manton, Duke University, USA

The aged 65+ and the oldest-old aged 85+ population are among the fastest growing age groups in the US population. Further they have the highest per capita need for acute and long term care (LTC) services. This combination has recently raised considerable concern among policy-makers about the long term fiscal status of the Medicare program and the desire to provide for means to pay for long term care services.

Most national surveys are not explicitly designed to assess the aged population's health characteristics and use of health services.

Measurement problems associated with the oldest-old and maintenance of high response rates is generally assumed to be negatively associated with increasing age. Age-related increases in functional impairment and intensity of health services have been widely demonstrated.

We present information from the longitudinal National Long Term Care Surveys (NLTC) (1982, 1984, 1989, 1994, 1999) that illustrates the pattern of nonresponse among the surprisingly heterogeneous aged population on a period and cohort specific basis across the life of the longitudinal panel. We then employ grade-of-membership analysis (GoM) to examine the changing correlates of nonresponse across the period 1982-1999 for the aged population. It is widely assumed that a weak but important negative association between age and response exists in surveys. This observation, based on a wide variety of general population surveys is examined in detail on both a period and cohort basis with large ($n = 2000$) 85+ and very large ($n = 6000$) samples in each year of the NLTC (longitudinal sample with age in cohort in each survey year).

Tabular results suggest a modest effect of age on reporting which may be influenced by level of physical functioning. Thus, as health and physical functioning improve among the aged, we expect a decline in period nonresponse to ages 95+; where about 1 of 3 is disabled and mortality is high. Further, we expect a decline in the rate of increase in nonresponse with each succeeding entry cohort.

The NLTC which incorporates notions of saliency and tailoring in its' design employs substantial resources to locate and interview the aged population with great success. The results suggest that trends in health and functioning influence nonresponse in a measureable way just as other access/environmental barriers also affect it.

Over time, proxy reporting levels have dropped in the NLTC along with declines in the reports of health and functional limitations. Our conclusions concerning improvement in years of healthy life lived are based on high (about 95%) response rates. Confidence in these conclusions plays a direct role in the Social Security and Medicare solvency debate.

Characteristics Related to Response: Experience From an 18-state Mixed Mode Surveillance System

Brenda Colley Gilbert, Holly B. Shulman, Laurie Fischer, and Mary Rogers, Centers for Disease Control and Prevention, USA

The Pregnancy Risk Assessment Monitoring System (PRAMS) is an ongoing, state-specific, population-based surveillance system designed to identify selected maternal before, during, and after pregnancy among stratified samples of mothers who have recently given birth to a live infant. Currently, 18 states are participating in PRAMS; surveillance covers approximately 35% of all U.S. births.

Every month, in each state, a sample of new mothers is selected from birth certificates to receive a 14-page mailed questionnaire. There is a second and sometimes a third mailing attempt before PRAMS staff telephone the nonresponders.. The mixed mode data collection methodology utilizes Dillman's Total Design Method.

Using 1996 data from 11 states, we examined overall response, response by mode, and response by several maternal and infant characteristics. Overall response rates ranged from 66-80%. To investigate the relationships between maternal and infant characteristics, and the likelihood of response, for each state we used logistic regression. We examined mother's age, race, ethnicity, marital status, maternal education, parity, and birthweight. Each of these variables was included in the model for each state before reducing the models down to only those variables significantly ($p < .05$) related to the likelihood of response.

The characteristics of women most likely to respond to the PRAMS survey were extremely consistent across states: response rates were higher for first-time mothers, higher for married women, higher for white women, highest among women with 12 or more years of education, and higher among women with a normal birthweight infant for all states. In all states, the strongest predictors of response were parity and education. Ethnicity, birthweight, and age were not good predictors of response in most states.

Based on these findings, we conclude that the mail/telephone methodology employed in PRAMS is an effective means of reaching a high proportion of sampled women who have recently given birth. These findings may be useful to other researchers involved in surveillance of pregnant women or to those using mail and telephone surveys to reach other populations.

Thursday, 3:30-5:00

Privacy and Nonresponse

Don't Call Me: The Impacts of Protecting Privacy on Nonresponse

Gary T. Henry and Dana K. Rickman, Georgia State University, USA

Increasing numbers of unsolicited phone calls have begun to stimulate legislators to provide higher levels of privacy protection for individuals and families. One method of increasing privacy is to allow individuals to restrict telemarketers, organizations seeking contributions, and survey researchers from placing calls to residents who want these protections. For political pollsters and survey researchers, these restrictions threaten to increase response bias due to non-response. But are the individuals who would opt to have their telephone numbers excluded the same individuals who currently refuse to participate?

Two surveys, one a national survey of 1108 adult residents and another of 820 Georgia residents, were conducted to examine the impact of potential non-response. A slight majority in both surveys agreed that households should have the right to block calls, although only 20 percent of the national survey respondents indicated they would actually block scientific polls and 27 percent would block political polls.

The surveys were also used to examine the impact of restricting calls on the probability of response and survey estimates. Respondents were asked how likely they were to agree to participate in surveys. These were converted to probability estimates and

regressions were run using a variety of demographic and substantive variables as explanatory variables. Then identical regressions were run after the dependent variables were adjusted to reflect a zero probability of responding for those who would exclude their phone numbers. Self described conservatives were the only group for whom the negative bias would increase significantly, if allowed to exclude their phone numbers. To estimate the effects on several demographic and substantive variables, such as political knowledge, estimates were developed including

then excluding those individuals who would exclude their phone numbers. Results are presented unweighted and weighted using gender, race, and age post-stratification weights. Biases in estimates are observed, but post-stratification weights add a measure of correction.

Identity Thieves, Warrantee Cards and Government Surveys

Eleanor Gerber and Melinda L. Crowley, U.S. Bureau of the Census Susan R. Trencher, George Mason University, USA

Concerns about privacy have direct relevance to survey nonresponse, both in terms of refusal to participate in surveys and item nonresponse. There is a general belief that respondents are increasingly concerned with revealing information about themselves. It has been suggested that such concerns play a role in response rates in all modes of data collection.

The current paper reports on a qualitative study of belief structures that influence survey respondents' perceptions of and reaction to survey information requests. The premise of the research is that there is an increasing awareness of demands for information and a sense that individuals must take steps to guard against unnecessary revelation of information and its misuse by others. These beliefs and practices are widespread cultural phenomena, although emotional investment in the issue varies widely with respondent. The main focus of this research is how respondents conceptualize the way that they manage information about themselves and their families. Among the issues discussed are the following: 1. respondents' perception of risks to which they expose themselves when they reveal different kinds of information; 2. sources of information and knowledge that influence these beliefs; 3. the information management strategies respondents employ to reduce their risks or to protect themselves; and 4. social and cognitive factors that can encourage respondents to reveal personal information under different kinds and amounts of threat. Respondents' reasoning about these issues is the central focus of the research. Reactions to surveys are examined in the context of other requests for information by marketers, polling, and other institutional record keeping. Beliefs about effects of new technology on personal privacy are also explored.

Sixty respondents who had just ended participation in the Current Population Survey or the Survey of Income and Program Participation were interviewed by a team of ethnographers in four locations around the nation. Special efforts were made to interview respondents who dropped out of the survey prior to finishing a complete survey rotation, in order to capture the viewpoints of respondents who were uncomfortable with survey requests and demands. A semi-structured qualitative interview protocol included several tasks to elicit respondents' beliefs and accounts of behavior. Open-ended prompts elicited respondents' experience with information requests, sources of knowledge about privacy matters and information management strategies. A series of vignettes focused on reactions to a variety of information collection instruments. In addition, several card sorting tasks shed light on how respondents rate their trust in various sponsorship for data collections, and the sensitivity of requests for particular kinds of information. In addition, ethnographers observed first time interviews with respondents in their first month in sample, focusing on the spontaneous emergence and negotiation of privacy issues. Field interviewers were informally debriefed by field interviewers for their perceptions of the role of privacy issues in achieving cooperation with surveys.

Survey Quality and Mobile Phones

Vesa Kuusela and Veijo Notkola, Statistics Finland / Survey Research Unit

Due to the increasing popularity of the mobile phones, the structure of telephone ownership is changing rapidly in many countries. In Finland, the amount of the households having a conventional telephone has decreased during the past ten years. Simultaneously the amount of mobile phones has shown a very rapid growth (see Kuusela, 1997). In the end of year 1998 there were more mobile phones than conventional phones. Almost two thirds of households have at least one mobile phone and many households have more than one mobile phone. Occasionally, mobile phones have replaced conventional phones (Kuusela, 1997).

In telephone surveys where the sampling frame is the list of telephone numbers, there will be some under coverage if only conventional phones are included, because of the changes in the telephone coverage. Even if the mobile phones are included, there may still be some under coverage because mobile phones are not listed as frequently as conventional phones. Other way around, the selection probabilities of some households will often be larger than the others because there is more than one mobile phone in the household.

Because of the mobile phones, people who were previously hard to reach may be contacted much easier now. However, the respondent may answer the phone almost anywhere and therefore he or she may not be willing to answer any questions, but making an appointment is possible. This helps both the telephone and face-to-face surveys.

Another question, however, is conducting an interview in a mobile phone. Except that it is more expensive, it may affect the quality of a survey, as well. The respondent may be in a somewhat uncomfortable situation and he or she therefore answers heedlessly. Moreover, people are sometimes reluctant to speak long in a mobile phone and therefore they answer hurriedly.

Mobile phones may have many effects on the surveys and on the survey practice, because of a variety of reasons. However, there is no empirical evidence concerning the potential effects.

To gather information about the possible effects of the mobile phones in Finland, questions dealing with the use of a telephone were added at the end of the LFS questionnaire. LFS is a telephone survey but if a respondent does not have a telephone he or she will be interviewed face-to-face. The material consists of nearly ten thousand cases gathered during the winter 98/99. The purpose of the study is to analyse what kind of the effect mobile phones may have on the non-response, and what kind of effects

mobile phones have on the quality of a survey in terms of the duration of interviews. The effects on some central variables are also considered.

Thursday, 3:30-5:00

Nonresponse in Election Polls

Sampling and Nonresponse Biases in Election Surveys: The case of the Last General Election in Quebec

Claire Durand, Andre Blais, and Sebastien Vachon, University of Montreal, Canada

During the last electoral campaign in Quebec, Canada, all the polls published in the media had a similar estimate of vote intentions. The Parti Quebecois (PQ), a centre-left party dedicated to Quebec sovereignty, was clearly ahead, by an average of five points in the last six polls of the campaign. The PQ won the election, held on November 30, 1998, but with a lesser share of the vote (43%) than the contending Liberal Party (44%), a centre-right federalist party. Pollsters and many observers have contended that the discrepancy between the polls and the actual vote could be explained either by a last minute shift in favor of the Liberals or by differential turnout.

We rely on a number of sources of data in order to sort out the possible causes for such a discrepancy. A post election poll was conducted among fifteen hundred respondents who had answered one of three electoral surveys conducted during the penultimate week of the campaign by two Quebec pollsters (CROP and CREATEC). The response rates for the campaign surveys varied from 50% to 60% and the reinterview rate in the post election survey was 83%. An analysis of the data from three surveys carried out by Crop during the four-weeks campaign was performed in order to estimate the impact of item and survey non response. A study of voting sections with a high percentage of collective households allows us to estimate the voting behavior of residents of collective households. Two Statmedia studies conducted in 1997 and 1998 provided information on the sociodemographic characteristics of respondents from unlisted and doubly listed telephone lines. Finally, three Crop surveys carried out after the election allows us to compare the voting intentions of respondents from listed and unlisted telephone numbers.

The results of the post election survey do not support the late shift and differential turnout hypotheses. The most likely explanation for the discrepancy between vote intentions as revealed in the polls and the actual vote is to be found in sampling and non response biases. Analysis of item non response as well as survey non response shows that there is a consistent tendency for non respondents to be supporters of the Liberal Party. An analysis of sampling frame biases also show that Liberal supporters are likely to be under sampled. Finally, adjustment weighting also tends to increase the bias against the Liberal Party.

It is pointed out that these biases are not specific to the Quebec situation and are likely to increase with demographic and technological changes.

Religion and Nonresponse: Polling the Peace Process in Northern Ireland

Colin McIlhenny and Amanda Fleming, PricewaterhouseCoopers, UK

In 1998 the electorate in Northern Ireland went to the polls in the historic referendum on the Good Friday Agreement and later to elect a new parliament.

This paper assesses religion as a potential variable influencing non-response. The author as Director of International Research for PricewaterhouseCoopers has been responsible for designing a series of opinion tracking surveys commissioned by the BBC and Sunday Times newspaper. The paper will present evidence highlighting any patterns of non-response which emerged via the polls as the public counted down to referendum day.

The electorate in Northern Ireland is one of the most divided in the world with the voting age population of just over 1 million bitterly split on the key constitutional issues of the day. Religion is the key variable in predicting party support at the polls. This paper breaks new ground in assessing its role as a potential predictor of survey response/non-response.

Thursday, 3:30-5:00

Panel: Does Nonresponse Really Matter?

Friday, 8:30-10:00

Multivariate Imputation of Coarsened Data On Households' Income and Net Worth

Steven Heeringa, Roderick J.A. Little, and Trivellore E. Raghunathan, University of Michigan, USA

The Health and Retirement Survey (HRS) is a longitudinal study of the relationship of individuals' physical, mental and financial well being. Key to many analyses of the HRS data are detailed measures of the financial characteristics of respondent households. Survey questions that ask respondents to report dollar values for financial variables such as income, assets, and liabilities are subject to high rates of missing data. To counter these high rates of item-missing data, the HRS uses special questionnaire formats designed to collect an interval scale observation whenever a respondent is unwilling or unable to provide an exact response to a financial amount question. The use of these special question formats significantly reduces the rates of item-missing data for HRS financial variables, but results in a mixture of actual value responses, interval censored responses and completely missing data.

We develop and apply a multivariate Bayesian approach to multiply impute the "coarsened" observations for financial variables. The paper will describe our multivariate imputation algorithm and will include an extensive discussion of results from simulation studies and real data applications designed to evaluate the multivariate approach to imputation of coarsened financial measures.

To aid practitioners who must choose between methods, a major section of the paper is devoted to comparing the performance of the Bayesian multivariate approach to that for univariate imputation methods that are common in current survey practice.

Friday, 8:30-10:00

To Answer or Not to Answer: Decision Processes Related to Survey Item Nonresponse

Paul Beatty, National Center for Health Statistics, USA
Douglas Herrmann, Indiana State University, USA

Survey item nonresponse (including refusals to answer questions and "don't know" responses) creates obvious challenges for survey analysts. The prevalence of item-nonresponse is largely determined by questionnaire designers, who choose the specific words that define the respondent's task. The resulting questions vary in clarity and complexity of the expected response task, factors which may strongly influence respondents' willingness and ability to provide substantive answers. Questionnaire designers may also send direct signals about the appropriateness of item-nonresponse (e.g., by making "don't know" responses more or less explicitly available).

This paper combines theories of the response process with data from an experimental questionnaire, in an effort to understand how respondents decide whether or not to provide a substantive response to a survey question. Based on previous literature on the response process and our own research, we propose that the decision to respond or not respond to a question is driven by three factors: the availability of the information requested (cognitive state), the respondent's perception of the level of accuracy required by the questioner (an adequacy judgment), and the respondent's motivation to provide the information requested (communicative intent).

Cognitive states include available (the requested information can be retrieved with minimal effort), accessible (information can be retrieved with effort or prompts), generatable (information is not literally known, but may be estimated), and inestimable (information is not known and there is virtually no basis for estimation). While we would expect cognitive state to be highly influential in the decision to respond to a question, we note that theoretically, any cognitive state can lead to either a substantive response or item non-response; this depends on the respondent's understanding of the question requirements, and motivation for both cognitive effort and disclosure of information.

After presenting our model of these decision processes, we discuss results from a questionnaire designed to test our model. This began with autobiographical memory questions intended to draw upon a variety of cognitive states. For example, we expect that a respondents' date of birth will be available for almost everyone, while an estimate about lifetime consumption of fruit would be generatable at best. Several factual and subjective questions were also included. Afterwards, respondents answered cognitive questions about those responses, including a self-assessment of their cognitive state; confidence in the accuracy of their answers; identification of questions in which there were any doubts as to whether they would provide a substantive response; and factors involved in the decision to respond or not.

Respondents generally provided a great deal of information about the basis for their answers. We obtained measures of the relative difficulty of the questions, which questions created a dilemma of response decision, and what factors were weighed in that decision. We also explored how these factors co-varied with alternate instructions about whether or not "don't know" responses were acceptable. The paper concludes with a discussion of the theoretical implications and practical applications of this sort of assessment.

Friday, 8:30-10:00

Imputation and Variance Estimation

How to Find the Best Imputation Technique? Tests with Three Methods

Seppo Laaksonen, Statistics Finland

The paper develops the imputation method which takes advantage both of a multivariate regression model and a nearest neighbour hot decking method. This method is successfully applied to a complex case where a variable being imputed is of a ratio-scale type and consists of a high number of non-known zero values. The results obtained by means of the method are compared with two other techniques, (i) random hot decking and (ii) a two-step model based method. The latter one first takes advantage of logistic regression and then of standard regression imputation. Our results do not give a unique conclusion. Regression based nearest neighbour hot decking is best on average but the two-step model based method has also some advantages. The paper also makes an attempt to estimate variances which take into account the fact that some data are imputed. This method provides an additional variance component, called imputation variance. The paper also presents a diagnostic test for the quality of imputations; this test checks how many times the same donor is used in imputing missing values.

Imputations Based on Box-Cox Transformations: Properties of Survey Estimates

Claes Cassel, Jan Selén, and Peter Lundquist, Statistics Sweden

Weighting methods are a standard approach for unit or total non-response. For missing data for specific variables, that is for item non-response, methods where one imputes (fills in) values are common. There is a number of imputation methods such as mean imputation, hot and cold deck imputation, substitution and regression imputation. Composite methods combine approaches, stochastic imputation as opposed to deterministic imputation incorporates a random component, while multiple imputation results in more than one value for the missing item by repeated stochastic imputation.

Here we will concentrate on regression imputation. We will explore a rather recent suggestion where transformation is an option for the modelling. Of special concern is the Box-Cox transformation, which encompasses the semilog specification and the linear

specification as special cases. For imputations of income, as an example, usually either of these two model forms is used by the statistical agencies. We will examine the two alternatives in a more general transformation framework, explore the properties of some alternatives and discuss guidelines for a choice and estimation of a transformation. A discussion of the evaluation of imputations is included.

The estimation of the precision of survey estimates when some variable values are imputed cannot be achieved by standard methods. Following a suggestion by Särndal (1992) where the variance of an estimated total is expressed as a sum of the sampling variance and the variance due to imputation, we give a general formula for the variance which is valid in the case of some estimated transformations.

The approach is tested on data from the Swedish household budget survey 1995. The households included in the survey were also included in an earlier survey the same year. Therefore data for some non-response in the household budget survey is known, which is advantageous for the comparison of imputation alternatives. Our interest is concentrated on household income.

Friday, 8:30-10:00

Survey Context and Participation

The Influence of Environmental Characteristics on Survey Cooperation: A Comparison of Metropolitan Areas

Brian A. Harris-Kojetin and Scott S. Fricker, Arbitron, Inc., USA

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 (Groves and Couper, 1998). Over these diverse contexts, there may be differing norms and expectations for interacting with strangers and complying with a request for help. More specifically, there also may be different expectations for survey interviewers requesting participation and for potential respondents interacting with them. These attitudes and norms may affect behavior towards survey interviewers in general or as representatives of the particular sponsor or data collector. The end result may be varying response rates across areas of a country.

There are many examples of differences in nonresponse across different areas within a country, particularly distinctions observed between urban and rural areas (e.g., Groves, 1989; Groves and Couper, 1998; Goyder, Lock, and McNair, 1992; House and Wolf, 1978; Steblay, 1987). However, there is much less documentation of the varying social and economic conditions that may underlie these environmental differences in response rates. Recently, Groves and Couper (1998) found that in addition to urban-rural differences, characteristics of the community, such as block-level population density and the percentage of persons less than 20 years old, were related to survey cooperation rates. Other research in social psychology has focused on specific characteristics of communities to try to understand some of the aspects of the environment that underlie differences in people's helping behavior. For example, Levine and his colleagues (1994) examined six different types of helping behavior in 36 cities and identified demographic, social, and economic characteristics of these communities that was related to the level of helping behaviors observed.

In this chapter we plan to review the theoretical and research literatures in social psychology, sociology, and survey methodology that provide insights into the environmental characteristics that may influence survey cooperation. Furthermore, we will be operationalizing indicators of the demographic, social, and economic environment from a number of sources, including Census data, to construct composite indicators that reflect social psychological attributes of metropolitan areas in the United States. We will then examine how well these indicators are related to differing levels of survey cooperation rates across metropolitan areas in the United States.

We currently have average cooperation rates for all metropolitan areas from the Current Population Survey, the U.S. monthly household labor force survey. We hope to obtain these data from some other major surveys across a wide a spectrum as possible, including government, academic, and commercial surveys. First, we will examine the pattern of cooperation rates for different surveys across the metropolitan areas to see the extent to which different surveys conducted by different organizations experience similar relative cooperation rates in specific metropolitan areas. Secondly, we will construct models utilizing the cooperation rates from the different surveys as dependent variables in regression models that utilize demographic, social, and economic characteristics of the metropolitan areas as predictors. Finally, we will discuss the implications of these findings for theories of survey cooperation and for improving data collection procedures.

Nonresponse Bias in Household Expenditure Surveys

John King, Office for National Statistics, UK.

Response and non-response to a large household expenditure survey are studied using geo-demographic codes. Geo-demographic information available for the drawn sample is used to identify and describe responding and non-responding households. Information from the sample on household expenditure, income and some other characteristics is used to estimate the effect of non-response. Overall non-response bias in some income and expenditure variables appears to be small. This appears to be the net effect of larger biases in different directions. The non-response pattern is compared with that of other large household surveys. Comparisons with results from other grossing-up methods are also indicated. Implications for fieldwork practice to reduce non-response are also discussed.

Nonresponse in Interviews: Stable or Instable Patterns of Action?

Joerg Blasius, Univeritaet zu Koeln, Germany

In recent years, there has been an increasing discussion about low response rates in social science surveys. In order to obtain higher response rates one can contact non-respondents again, this time by another interviewer, trying to induce them to participate in the interview. However, if answers are not given because of the interviewee's principles not to take part in any interview, another try is probably unsuccessful. If the reasons for non-responding are due to specific situations, another contact does make sense.

Within a project supported by Deutsche Forschungsgemeinschaft (German Research Foundation), the author tried to maximise response rates both of a face-to-face interview and a telephone interview which ran at the same time. They also tried to find out as much as possible about the interview situation. In order to achieve this aim all non-respondents were visited again or phoned a second time, respectively, and asked again for their participation in the interview. This procedure had the following result: in the face-to-face interview 30 % of the people contacted could be convinced to answer the questions, in the telephone interview even 50% of the non-respondents could be converted. In addition, the interviewees were asked to answer questions about the course of the interview. Last, but not least, immediately after the interview had taken place, all people interviewed within the two studies received a letter asking them to fill out a questionnaire on the interview situation; here, the response rate was over 80 % in both studies.

Based on this material it is possible to give information on converted non-respondents and on the most important parameters which finally induced people to participate in the interview, after all. Furthermore, we want to investigate if there is a difference in response patterns between those who were willing to participate in the interview from the start and "converted" non-respondents.

Friday, 8:30-10:00

Nonresponse in Establishment Surveys

Exploring the Relationship Between Survey Participation and Survey Sponsorship

Jaki Stanley McCarthy, Kathleen Ott, Jay Johnson, National Agricultural Statistics Service, USA

The type or particular organization sponsoring a survey data collection is typically thought to have an impact on survey response. It has been suggested that government or university sponsorship increases survey participation, evidenced by the relatively high response rates enjoyed by many government sponsored surveys.

However, there is very little quantitative evidence to support the claim that respondent knowledge of organizational sponsorship directly affects survey response.

A series of questions was asked of agricultural establishments selected for the National Agricultural Statistics Service's quarterly agricultural surveys in one state in order to examine the relationship between knowledge and attitudes of respondents toward NASS surveys and survey participation. The questions were about the respondents' identification of the agency (at the local and national level), their perceptions of the agency and its data, the effect of data on the respondents, and their opinions regarding responding to NASS surveys. These questions were asked of both respondents and non-respondents to the surveys in contacts throughout 1998 and 1999.

Initial results indicate that both survey respondents and non-respondents generally claim to recognize NASS and have some understanding of the data products that are produced as the result of their reported data. Results also suggest that it is not recognition of the agency that is related to NASS response rates, but the perceptions of how NASS reports and services impact farm operators. Distinct differences were found in attitudes of respondents and non-respondents for some of these measures. Differences were also found between different types of operations, although not across operation size.

Findings are intended to guide promotional and public relations activities that will be targeted toward potential respondents and suggest data collection procedures that will increase survey participation. The opinion questions will continue to be asked of survey respondents to gauge changes in attitudes toward the Agency as these activities continue.

An Analysis of the Relationship Between Survey Burden and Non-Response: If We Bother Them More, Are They Less Cooperative?

Jaki Stanley McCarthy and Dan Beckler, National Agricultural Statistics Service, USA

In surveys of certain populations, individuals may be contacted on numerous occasions over time. This is particularly true in surveys of establishments, where large or unique operations may be selected with near certainty for recurring surveys and may be included in samples for multiple surveys. Cooperation in any particular survey may be affected by the number and frequency of times an establishment has been selected for surveys by that organization in the past.

The USDA's National Agricultural Statistics Service (NASS) contacts farms and ranches in the United States for many surveys. For example, the Quarterly Agricultural Survey collects data on inventory and production; the Farm Labor Survey collects information on hours worked and wages; the Agricultural Resource Management Study collects information on production practices, chemical and pesticide use, and farm economics. A particular farm or ranch may be selected for any or all of these surveys, both within a single year and over multiple years. This paper examines the relationship between response on one NASS survey and the reporting burden placed on agricultural operations by NASS in the past.

Measures of the survey burden that NASS places on its respondents will be made, including the number of other NASS surveys they have been contacted for, the length of time since they were last contacted for a NASS survey, the type of information they have been contacted for in the past. Comparisons of these burden measures will be made between respondents and non-respondents for the 1998 June Agricultural Survey in one state.

NASS has also been collecting information from respondents and non-respondents about their opinions and attitudes toward the agency and its work. The relationship between these measures and their NASS reporting burden will also be examined.

Implications of the relationship between survey burden and response will be discussed. Recommendations for practices to increase cooperation that may be employed by agencies who impose large reporting burdens will also be made.

"Which Form Will I Complete (First)?: Competing Questionnaires and Unit Nonresponse*Laurie Schwede and Sharon Birch, U.S. Bureau of the Census**Ingrid Goldstrom and Ronald Manderscheid, Department of Health and Human Services,**Joseph Moone, Department of Justice*

Unit response and nonresponse rates are often used to assess the quality of self-administered questionnaires. A relatively high unit nonresponse rate is most often considered to indicate problems endogenous to that particular survey's data collection, such as the length, clarity, and/or formatting of the questionnaires(s), or collection mode.

There may also be exogenous factors that affect unit nonresponse rates. One such factor that may affect nonresponse in establishment surveys is the interaction between two or more surveys that arrive around the same time, competing with each other and with the respondents' regular work duties for attention. However, it is rare that survey analysts both learn about competing surveys and can merge data from them to identify and examine the interactions in response/nonresponse between them.

In late 1998, personnel at three federal agencies realized they had two competing surveys with overlap in sample, time, and some content, providing a very unique opportunity to study the nonresponse interactions between them. We agreed to match the response files from the two competing data collections.

The purpose of this collaborative paper is to identify and examine the interrelationships of nonresponse between the two competing surveys sent to the same juvenile facility universe. Data for this paper come from the two surveys: 1) the Department of Justice (DOJ) mailout test of a new questionnaire in 497 juvenile facilities, conducted by the Census Bureau from October, 1998 to February, 1999, and 2) the Department of Health and Human Services (HHS) census of roughly 3500 juvenile facilities (including those in the DOJ test) from June 15, 1998 to the DOJ test closing date.

In this paper, we compare the unit response/nonresponse rates, patterns, and outcomes for these two competing data collections in the same target universe. First, we examine the unit response rates, patterns and outcomes for each data collection separately. Second, we analyze the interaction of response/nonresponse outcomes between the two data collections. Our first purpose is to assess whether the overlap in time of two competing data collections affected the unit response/nonresponse indicators of either of them. The second is to evaluate whether differences found between them were due to the time overlap, to differences in respondents' propensity to respond to surveys generally, to both factors, and/or to other factors.

Friday, 10:30-12:00**Persuading Reluctant Recipients in Telephone Surveys***Wil Dijkstra and Johannes H Smit, Free University, Amsterdam, the Netherlands*

A major cause of unit nonresponse is the reluctance to cooperate. Especially in telephone surveys, persuasion attempts by the interviewer are the main tools to reduce nonresponse caused by refusals. To investigate the effectiveness of such attempts, a total of 2155 introductions to a telephone survey about tv commercials and advertisements in papers and magazines were recorded and transcribed. All utterances of interviewer and recipient were coded. Successful interviewer actions appeared to be:

- reducing the length of the interview (e.g. after the initial statement that the interview should last 15 minutes, reducing it to 10 minutes)
- personal appeal (*I would appreciate it very much if you would participate*)
- maintaining interaction.

Ineffective interviewer actions included:

- tailoring (persuasion attempts by using information provided by the recipient)
- *small* requests, like *let's try just the first few questions*
- repeating the recipient's declination (*You aren't interested?*)

Persuasion attempts based on the compliance principles *social validation* (like *most people enjoy the interview*) and *authority* (stressing the university as the research agency) did not appear to be successful.

Additional analyses showed that the behaviour of the recipient during the introductory sentences of the interviewer, e.g. humming, saying *good evening*, interrupting the interviewer) could predict the probability that the recipient eventually refuses, very well.

It is concluded that the ineffectiveness of persuasion strategies like *tailoring* or attempts based on compliance principles, suggests that theories of persuasion derived from marketing research, are not necessarily applicable to persuade recipients to participate in an interview.

Friday, 10:30-12:00**Nonresponse in Longitudinal Surveys***James M. Lepkowski and Mick P. Couper, University of Michigan*

Nonresponse in panel surveys poses unique problems and opportunities for the study of and the compensation of nonresponse than can be obtained in one time cross-sectional surveys. Missing units in a wave, or panel nonresponse, can be viewed as conditional on baseline or prior wave participation. Baseline or prior wave measurement provides extensive auxiliary data that can be used to investigate determinants of nonresponse and to compensate for a missing wave. A theoretical model for understanding and exploring nonresponse in panel surveys is presented, and it is examined using data from two panel surveys. Panel

nonresponse at wave t can be partitioned into several types: (1) sample loss when survey operations fail to locate (L_t) sample subjects; (2) subjects who are located cannot be contacted (C_t); and (3) successfully located and contacted subjects do not cooperate (I_t). A general model of panel nonresponse at wave t , R_t , that uses this typology gives logit of responding at a wave t as $\log(\Pr\{R_t\}) = \log(\Pr\{I_t|C_t,L_t\}) + \log(\Pr\{C_t|L_t\}) + \log(\Pr\{L_t\})$, where $\log(\Pr\{I_t|C_t,L_t\}) = X_{@t} + Z_{@t-1}$, $\log(\Pr\{C_t|L_t\}) = Y_t + W_{@t-1}$, and $\log(\Pr\{L_t\}) = U_{@t} + V_{@t-1}$. Here, t , t , and t denote parameter vectors for characteristics at wave t , and t , t , and t denote lagged effects of the same or different parameters at wave $t-1$. Empirical findings from two panel sample surveys, the National Election Survey and the American's Changing Lives Survey, test the theoretical framework. Both panel survey data sets allow exploration of demographic, geographic, and social-psychological determinants from a prior wave as lagged predictors, and variation in survey designs allow contrasts of operational features of the panel survey data collection. Theoretical propositions are examined in the context of empirical findings to identify gaps in knowledge and to propose a research agenda to explore determinants of panel nonresponse.

Friday, 10:30-12:00 Incentives

Socio-Economic Bias in Mailed Surveys: Some Attempts at Interpretation

John Goyder, Keith Warriner, Susan Coutts, and Maryanna Basic, University of Waterloo, Canada

Surveys often show a "middle class bias" of over-selection from the upper socio-economic strata, but not always, as Groves and Couper showed in their recent monograph. The present research explores sources of variation in the amount of SES bias in a mailed questionnaire, beginning with an experiment using prepaid vs. postpaid \$5 cash incentives. The data are from the Kitchener-Waterloo Metropolitan Area Survey 1998. Contrary to the hypothesis prompting the experiment, the postpaid format did not increase lower SES representation compared with the prepaid format.

Among respondents in the sample, measurement was attempted for the norm of reciprocity. Factor analysis differentiated this norm from self-maximization and altruism. Again contrary to expectations, the reciprocation measure was un-related to SES. Instead, those with higher SES most embraced a self-maximization orientation.

The lower strata do, however, show low sense of power on the affect control dimensions. This prompts the tentative conclusion that when low SES representation occurs in a survey the cause is more self-disqualification due to a sense of inability to contribute useful information to the research than class hostility or non-exchange orientation.

Using Express Delivery and Pre-Paid Incentives to Increase Response Rates on a Random Digit Dial Survey

David Cantor, Patricia Cunningham, and Pamela Giambo, Westat, Inc., USA
Genevieve Kenney, The Urban Institute, USA

One of the most striking features of random digit dial (RDD) telephone surveys is the number of people that hang up during the introduction (Collins, et. al. 1988). This is indicative of the minimal communication that actually occurs in the first few moments of the call. This poses special problems for the survey designer when trying to increase response rates. Attempts to tailor the introduction to increase the legitimacy of the study are not particularly effective because respondents are simply not listening at this point in the interview (Groves, et al., 1979; Dillman et. al., 1976).

Prenotification and pre-paid incentives may reduce the communication problem. Theoretically, prenotification can convince the respondent that the call is legitimate. Incentives should increase the value of participation to the respondent. Previous research on both incentives and prenotification in the context of RDD surveys, however, is not encouraging. While research has shown that prenotification does increase cooperation rates, the results are somewhat mixed (Singer, et al., 1997; Camburn, et. al., 1995; Brick, et. al., 1997). The problem is thought to be that many people do not open mail they do not recognize or the letters simply do not arrive at the appropriate address (Traugott, et. al., 1997). Similarly, promised incentives have not been found to be effective in this context. Many respondents simply do not believe the offer of money at this stage of the process (Strouse and Hall, 1997; CMOR, 1996; Cantor, et. al., 1997; Singer, et. al., 1997).

The purpose of this paper is to report results of a series of experiments that attempted to address each of these issues by varying: 1) the type of pre-notification used (first class mail vs. express mail) and 2) whether or not monetary incentives were sent to respondents prior to contact. The use of express delivery was used to increase the chances that the mail would be opened and read by potential respondents. Pre-paid incentives were used to convince respondents that the study appreciated their participation in the study. Experiments will be reported for use of these methods at both the initial conversion stage, as well as when trying to convert initial refusers.

The results of these experiments indicate that, in fact, nonresponse on RDD surveys is relatively soft. Both express delivery and pre-paid incentives have significant effects. For example, the combination of these treatments increased cooperation rates at refusal conversion above those achieved at the initial contact. The two "treatments" were found to have independent effects. This is consistent with the hypothesis that express delivery increased the number of persons that opened the mail, while the incentive increased the respondent's motivation to participate.

Other, more preliminary analyses of these data, suggest there is also an interaction between the composition of the sample and the effectiveness of these two treatments. In addition to reporting the above results, the proposed paper will further explore this interaction by estimating multivariate models predicting response rates using both the experimental treatments and sample characteristics (geography, interviewer effects, stage of calling).

Money and Motive: Results of an Incentive Experiment in the Survey of Income and Program Participation

Denise Abreu, Elizabeth Martin, and Franklin Winters, U.S. Bureau of the Census

Nonresponse is a significant problem for longitudinal surveys such as the Survey of Income and Program Participation (SIPP) because it reduces sample representativeness and may bias the estimates. In SIPP, households in poverty have higher attrition rates than non-poverty households (Waite, Huggins, and Mack, 1997). Nonresponse bias is of special concern for the 1996 SIPP Panel, which has higher nonresponse rates than previous panels. The household non-interview rate as of the 7th interview (or wave) is approximately 27%.

The current practice in SIPP is to revisit nonrespondents once more after their initial nonresponse. About a third of nonresponding households are converted to interviews in the next wave. An incentive experiment was conducted to improve these results and reduce attrition by directly targeting nonrespondents.

Households were randomly assigned to receive a prepaid incentive of \$20, \$40, or no incentive. Consistent with current practice, all groups received an advance letter prior to the interviewer's visit. The letter received by the incentive groups provided information about the incentive, and included a debit card. The letter received by the no incentive group was the usual letter sent to nonrespondents. All letters were sent via priority mail, to ensure that respondents received the incentives (priority mail is not usually used to follow up nonresponding cases).

The experiment included all Wave 7 and 8 Type A noninterviews (N=2,900), which include refusals as well as cases of no-one home, temporary absence, or language barrier. Four sample selection strata were formed by cross-classifying the poverty category (high poverty stratum / low poverty stratum) by noninterview status (refusals / other noninterviews). Within each stratum, three randomly selected subsamples of almost equal size were assigned to one of the three treatment conditions.

The results of the experiment show that the overall conversion rate in the next wave increased significantly when an incentive was offered. Both the \$20 and \$40 incentives significantly improved conversion rates; in addition, use of priority mail alone appeared to improve conversion rates, compared to the usual procedure. The effect of incentives was concentrated among refusals in the prior wave, with no significant effect for other noninterviews.

Conversion rates in both nonpoverty and poverty households were improved by incentives. In poverty households, both the \$20 and \$40 incentives resulted in significantly higher conversion rates than no incentive, while in nonpoverty households only the \$40 incentive led to a significant improvement.

Interviewers' notes for the experimental cases were coded and analyzed to examine motivational factors which may influence respondents' reactions to incentives, and survey cooperation more generally. Results show that burden concerns expressed in a prior interview were associated with an announced decision to stop participating in the survey, which led to lower interview rates in a subsequent wave, independent of the effects of the incentive experiment. Analysis suggests that receipt of a prior incentive (in Wave 1) interacted in complex ways with perceived burden and the decision to stop participating in the survey.

Friday, 10:30-12:00
Interviewer Effects 2**Evaluation of Interviewer Behavior During the Doorstep**

Geert Loosveldt, Ann Carton, and Jaak Billiet, K.U. Leuven, Belgium

Contacting the respondent and persuading him or her to participate in a survey interview are important tasks of the interviewer in face-to-face interviews. Recent research about non-response, the role of the interviewer during the initial contact and the importance of the initial doorstep interaction between respondents and interviewers as an intervening variable preceding the response decision is stressed. Several aspects of interviewer behavior related to the task of contacting and persuading can be evaluated. The first important aspect is an element of outcome evaluation: the contact rate and the refusal rate. It is recognized that some interviewers have higher refusal rates than others and that interviewer effects remain strong for individual and household refusals after controlling for characteristics of households and individuals. In this paper we do not try to explain these differences between interviewers, but these differences are related to the quality of data collected during the interview. The first research question is: do the more successful interviewers during doorstep contacts also obtain better data from the respondent during the interview? We hypothesize that the task of the interviewers at the door is different from the task of the interviewer during the interview itself and that the data quality obtained by successful interviewers is not necessarily higher. The behavior of the interviewer during the interview is much more rule guided by rules and structured by the questionnaire and the interviewer training. Different skills are used: for example, tailoring during the doorstep contact and reading all the questions exactly as written in the questionnaire during the interview. Although interviewer behavior in contact and doorstep strategies is largely unscripted, in most research projects the interviewer must follow some elementary instructions. These are instructions about how (by telephone or face-to-face) and when (contacts at different times of the day) to contact the respondents. The second evaluation question is related to these instructions: do the interviewers follow these instructions and do interviewers acting according to these instructions obtain data of higher quality? We assume that following the instructions of the contact strategy is an indicator of good interviewer behavior during the interview. As a consequence, we expect higher data quality for interviewers following the contact and doorstep instructions.

In testing these two hypotheses - related to the two evaluation questions- we will control for respondent characteristics that are correlated with data quality. Several indicators of data quality will be used, including item non-response and the amount of information obtained from open ended questions. For the analysis the Flemish data are used from the second (1995) wave of the Belgian Election Study. The sample is representative of the Flemish population aged 18-75 years. In this research project, information from a simple contact form is available. This information is sufficient to answer our research questions.

Interviewer Characteristics and Item Nonresponse in Telephone Surveys

Craig Helmstetter and Patricia A. Gwartney, University of Oregon, USA

We test hypotheses for telephone survey item nonresponse in relation to interviewer and respondent characteristics, as well as interviewer-respondent interactions. Item nonresponse refers to survey respondents' failure to provide substantive replies to interview question, that is, responses coded as "don't know," "refuse" and "no answer." Item nonresponse is problematic because it can be a source of bias in survey results and because it becomes missing data in analysis, or researchers must assign some reasonable estimate of values to the missing data to maintain a respectable sample size. The item nonresponse problem is best solved in advance, by minimizing it. As one step in doing so, we test the characteristics of both interviewers and respondents, and their interactions, which theory and previous empirical literature suggests will correlate with item nonresponse.

We begin by reviewing the literature on the effects of interviewer characteristics on survey item response and nonresponse and the literature on respondent characteristics and item nonresponse. We pose hypotheses for nonresponse patterns in relation to interviewer characteristics, respondent characteristics, and their interactions.

We then identify a number of difficult, demanding and sensitive survey questions in six surveys (one national, one regional, and four campus studies) conducted by the Oregon Survey Research Laboratory (OSRL), assuming that item nonresponse will be greatest for these types of questions. Using bivariate and multivariate analysis (multiple logistic regression), we examine item nonresponse in relation to seven interviewer characteristics: sex, race, citizenship, language/voice accent, age, and two uncommonly precise measures of interviewer experience. The respondent characteristics we include in the analysis are sex, race, age, education, and, where relevant and available, citizenship, labor force status, family status, and income. We also assess interviewer-respondent interactions for hypothesized relationships with item nonresponse, with particular attention to sensitive questions (e.g., concerning sexuality), where interviewer sex and respondent sex may be especially relevant to item nonresponse.

In general, we find that respondent characteristics are more closely related to item nonresponse than interviewer characteristics. In particular, respondent age is most consistently related to item nonresponse (older respondents more likely to not respond). Contrary to previous empirical literature on item nonresponse, we find that when controlling for other attributes, respondent education is not strongly related to item nonresponse. Neither interviewer characteristics nor interviewer-respondent interaction terms have consistent relationships with item nonresponse in the data we analyze.

In the concluding summary discussion, we speculate on the implications of our findings for question wording and interviewer training regimens necessary to minimize survey item nonresponse.

The Respondent, the Interviewer and the Questions as Sources of Item Nonresponse

Jan Pickery and Geert Loosveldt, K.U. Leuven, Belgium

In this paper we analyze the item nonresponse to several questions in a face-to-face survey and assess the impact of respondent and interviewer characteristics on it. We distinguish different kinds of questions (threatening or not, easy - complex) and different kinds of nonresponse (missing value - "don't know" or "no opinion" answer, with or without the anticipated possibility in the questionnaire). We examine which item nonresponse is subject to interviewer effects, using the appropriate multilevel model. Multilevel models offer the best possibilities to analyze interviewer effects on survey data. Multilevel analysis takes into account the hierarchical structure of survey data. The main purpose of the paper is to discern respondent, interviewer and question related sources of item nonresponse. Data come from the Belgian Election Studies.

A question commonly considered as being threatening is the income question. In our survey the respondent's income is asked for with an open and a closed question. Both are combined into one variable with three categories (answer to open question, answer to closed question, no meaningful answer), which is analyzed with a multilevel multinomial logit model. Questions of the political part with considerable item nonresponse were those regarding political party preference (most preferred, second most preferred, and least preferred). The question "Which political party do you prefer second most?" faced the largest nonresponse. We analyze it with a multilevel binomial logit model.

The possibility of item nonresponse for the two previous questions was not provided for in the questionnaire. A more complex task of the questionnaire with anticipated possibility of a "don't know" answer was rating six parties on eight different 11-point scales (non-Catholic - Catholic, free enterprise - government regulation, immigrants same rights - fewer rights,...). An explicit "don't know" filter was included in the question, but it was not mentioned on the card with the alternatives given to the respondent. We dichotomize the number of "don't know's" to these 88 questions and analyze it with a multilevel logit model. The respondents were also asked the extent of approval with 19 topics that were being discussed during the general elections. Also for this easier question a "no opinion" filter was included in the question but not on the card with response alternatives. Again the number of "no opinion's" is dichotomized and analyzed with a logit model.

Afterwards the respondents had to select three topics out of the 19, which were important to them when voting for a political party and evaluate the last government's policy regarding those topics. For these last three questions the "no opinion" alternative was supplied in the questionnaire, but it was not mentioned in the question, nor on the card for the respondent. For these questions the availability of this response alternative lies between the first and the last two questions. Again we dichotomize the number of "no opinion" answers and use a multilevel logit model.

Comparing the results of these different analyses we try to detect factors that have an impact on the presence of interviewer effects on item nonresponse.

Friday, 10:30-12:00

Software for Imputation

Using Prediction-Oriented Software for Estimation in the Presence of Nonresponse

James R. Knaub, Jr., U.S. Energy Information Administration

Survey sampling and inference may be accomplished by solely design-based procedures, or solely model-based procedures, or by model-assisted, design-based procedures. Depending upon circumstances, there are advantages to each of these methods. Further, in the case of solely model-based procedures, which employ basic econometrics models, often the Principle of Randomization is not completely ignored. Though the main proponents of modeling, Brewer, Royall, et.al., have not ignored randomization, Cochran did not ignore modeling either. Randomization is often useful, but there are cases, particularly in (highly skewed) establishment surveys, when, either in terms of timeliness, monetary or other resources, and/or in terms of nonsampling error, it may not be practical to sample from among the 'smallest' members of the population. Models based on simple linear regression have been produced which are often useful. Variance estimation has been explored (Royall and Cumberland (1981)) and the case of multiple regression has also been explored (Knaub (1996 and 1997)).

Imputation and the estimation of the impact of imputation on variance are transparent and integral to this process. That is, a value not collected during sampling will be treated as if it were not to have been part of the sample from the beginning. This article describes a general approach that may be used to organize estimation in a flexible manner. This approach was employed in another paper (Knaub (1999b)), for the Joint Statistical Meetings, for purposes of performing more general model-based inference and small area estimation. Here, however, the emphasis is on imputation and the corresponding variance estimation. Choice of regression weight may be impacted by the distinction between these uses. (Knaub (1999a) is an article that covers both of these uses.)

Readily available software may be used to change, add to or delete regressors, and results may be easily reorganized to present various aggregations. Standard regression software may be used to produce a predicted value and the corresponding estimated variance of the prediction error for each member of a population (given the data element of interest is collected from some other members of the population, and regressor data are available). If we want to present an estimated total, and an estimated variance of that total, which involve two or more such 'missing' values, one can not simply add the variances of the prediction errors, because contributions to the variance which are attributable to the coefficients are not directly additive. Note that the concept of model variance (Royall (1970)) as a measure of uncertainty applies equally well to the uncertainty in a reported total after imputation has been applied to a census, as it does to a sample.

The number and distribution of 'missing' data points can make a difference in the optimal assignment of the regression weight. This is influenced by the type of survey (establishment or household), and whether the goal is imputation or sampling. (Note that the difference between imputation and model-based sampling is only a matter of the number and distribution of 'missing' data points.) One must sometimes also consider whether it is useful to study a large 'area,' or better to consider strata.

In a sense, this method can sometimes convert non-ignorable nonresponse to ignorable nonresponse by allowing categorization of data one way for modeling, and another way for publication of results. A simple change to the variance estimation is made to handle this. An example application will be presented.

Teaching Applied Multivariate Analysis in the Context of Missing Data: A Comparative Evaluation of Current Software Remedies

R D Wiggan, Kevin Lynch, Simon Gleave, and J Bynner, City University, London, UK

The paper illustrates how postgraduate students in advanced social research can be introduced to the complexities of applied multivariate analysis in the context of unit and item nonresponse. Conventional methodologies (multiple regression, principal components analysis, factor analysis and clustering techniques) are introduced to the student using data from one of the British Cohort Studies, the National Child Development Study (NCDS). The study follows 17,000 births in one week in March 1958 until the present day. To date there have been five sweeps of data collection and attrition accounts for about a third of the original cohort. Examples are provided to compare and contrast new software products designed specifically to handle missing data, notably NORM, CAT, MIX and PAN (all developed by Schafer, J.)

Flexible Multivariate Imputation by MICE

Stef van Buuren and Karin Oudshoorn, the Netherlands

Several approaches for (multiple) imputation of multivariate data have been proposed recently. Schafer (1997) presents a methodology to describe the data by an encompassing multivariate model, derives theoretically sound posterior distributions under this model, and draws imputations from these by Gibbs sampling and other methods. The transcan function (Alzola & Harrell, 1999) imputes each incomplete variable by cubic spline regression given all other variables. This method is essentially a modernized version of Buck's regression technique (c.f. Little & Rubin, 1978). Though conceptually easy and flexible, the transcan algorithm lacks a theoretical rationale, so it is not known whether the generated imputations are proper in the sense of Rubin (1987).

Multivariate Imputation by Chained Equations (MICE) is an attempt to combine the most attractive aspects of both approaches. The MICE user specifies a conditional distribution for the missing data in each incomplete variable, for example in the form of a linear or (polytomous) logistic regression of the incomplete column given a set of predictors. Predictors themselves can be incomplete. It is assumed that a multivariate distribution exists from which these conditional distributions can be derived, and that iterative Gibbs sampling from the conditionals can generate draws from it. The distributions are not necessarily multivariate normal, but in the special case that all conditionals are chosen as normal linear regressions, then the method produces results similar to Schafer's technique for multivariate normal data. A small number of iterations, say 5 to 10, generally suffices. Although it is hard to establish convergence in the general case, simulation studies suggest that the coverage properties in some important practical

cases are quite good (Brand, 1999).

The algorithm is implemented as an S-plus function. For each incomplete variable the user can choose a set of predictors that will be used for imputation. This is useful for imputing large data sets containing hundreds of variables. Passive imputation is a built-in feature that takes care that transformed data are always in sync with their original values. This can be used, for example, to impute categorical variables along with their dummies (needed for imputing other variables). In addition, the user can alter the visitation scheme of the Gibbs sampler, or plug-in his or her own imputation method. Features like these make it easy to include complex imputation constraints in a practical but principled way.

Friday, 10:30-12:00

Panel: Efforts to Reduce Nonresponse in Commercial Surveys

Friday, 1:30-3:00

Incentives in Survey Research

Eleanor Singer, University of Michigan

A large number of experiments in recent years have manipulated incentives in telephone and face-to-face surveys, complementing the earlier, and even more voluminous, research literature on the effects of incentives in mail surveys. These experiments are based loosely on various forms of exchange theory (though many are quite atheoretical). The present paper synthesizes findings about the effects of incentives from both interviewer-mediated and mail surveys in order to increase their usefulness for practicing survey researchers and to stimulate more targeted research by survey methodologists. The paper (1) reviews what is known about the *intended* effects of incentives on (a) response rates and (b) response quality in both types of surveys; and (2) reviews what is known about such *unintended* consequences of incentives as concerns about equity, transformation of motives, and the development of expectation effects. In addition, it examines a number of special topics: (3) Do incentives operate primarily by modifying interviewer expectations or do they operate directly on respondents? (4) How effective are lotteries? (5) What is the effect of incentives in panel surveys; (6) Do incentives affect sample composition? (7) Are prepaid incentives cost effective? The findings are set in the context of what is known about people's motivation for participating in surveys.

Friday, 1:30-3:00

Replication Methods for Variance Estimation in Complex Surveys with Imputed Data

Jun Shao, University of Wisconsin

This article reviews various replication methods developed in recent years for variance estimation in complex surveys with imputed nonrespondents. We consider regression imputation methods (random or nonrandom), which include many commonly used imputation methods such as mean, ratio, and random hot deck imputation. The emphasis is on when, how, and why a replication method works. Technical details are omitted and can be found in the cited references.

Friday, 1:30-3:00

Reducing Nonresponse in Establishment Surveys

Understanding and Reducing Nonresponse in Establishment Surveys

Young Chun, U.S. Bureau of Labor Statistics

The supply side indicators measured by the surveys of household such as the Current Population Survey present a valid, complete picture of the U.S. economy when they are used along with the demand side indicators provided by the surveys of establishments such as the Current Employment Statistics. A declining response rate in establishment surveys (Christianson and Tortora, 1995) and economic censuses (Ambler and Mesenbrough, 1992) has been of primary concern to the government statistical bureaus as nonresponse errors may seriously affect data quality, leading to inaccurate assessment and forecasting of the economy while the effort to reduce nonresponse errors may significantly increase the government budget.

However, the dearth of research on nonresponse in establishment surveys becomes clear when searching for relevant literature despite the central importance establishment surveys have had in guiding and formulating national economic policy since the beginning of this century (Federal Committee on Statistical Methodology, 1988). Grove's recent work (1998) on nonresponse focuses on household surveys only. Dillman's (1978) comprehensive bibliography of available research on survey designs (1930s - 1970s) includes only one article on establishment surveys (Sergeant, 1958). Cox et. al (1996) is a notable exception where the two chapters discuss approaches to improving response to business mail surveys, or imputing business survey data.

Efforts to reduce nonresponse errors in establishment surveys require an understanding of the process of survey compliance from a business survey perspective. Without this understanding approaches to follow up nonrespondents are ineffectively prescribed, or post-survey nonresponse bias adjustment procedures may be ill-founded. In this paper, we introduce a multi-level theoretical framework that considers a set of micro-to-macro factors that influence the extent to which a business survey informant complies with a survey request in an organizational context. The theoretical concepts relevant to household survey nonresponses (Groves, Cialdini, and Couper, 1992) have only limited applicability to understanding the compliance process in establishment surveys. The informant in business surveys behaves under authority within the organizational system and reports on behalf of an organization,

and his reporting task depends on the use of records that moderates the degree of uncertainty the informant perceives. Therefore, we argue that organizational theory of communication is a paradigm to shift our thinking about (non)response behavior in establishment surveys. Research in the organizational decision making (Taylor, 1911; Likert, 1967; March and Simon, 1957; Katz and Kahn, 1966; Weick, 1969) and concepts specific to surveys of facts (Martin, 1993) bear on theorizing the informant's compliance with establishment surveys. Organizational determinants are specially examined to assess the effects on compliance of employment size, industry type, presence of corporate policies to release sensitive business data, organization of information in central or satellite offices.

The utility of such a theory of participation in establishment survey and reducing nonresponse errors is illustrated with data from a set of BLS experiments we have embedded in actual data collection since 1994 and in-depth interviews with interviewers. The 15 years of an annual survey of Hours at Work data are analyzed to understand the reasons for not complying with a survey request (e.g., ineligible, no contact, contacted and unwilling, contacted and not able, etc.) The establishment surveys we discuss in this paper have a business as the unit of analysis, and an informant's task primarily depends on the retrieval of records from an organization's information system. Some examples are drawn from research on other government establishment surveys.

Understanding the Factors Related to Nonresponse in Mail Establishment Surveys

Ashley Bowers, University of North Carolina Survey Research Unit, USA

Jeremy Morton, Research Triangle Institute, USA

Reducing nonresponse error relies on an understanding of the factors producing nonresponse. In interview surveys, we can identify and focus our efforts on different types of nonresponse, for example, noncontacts and refusals. However, in mail surveys, we are often unable to identify and address these different types of nonresponse. We may not know whether the potential respondent did not receive the questionnaire or received the questionnaire but did not participate for some reason. This is a particularly important problem in establishment surveys, where gatekeepers and an inability to personalize mailings make it likely that a number of nonrespondents never received the questionnaire.

Another difficulty in understanding mail survey nonresponse is that we usually do not know why potential respondents who received the questionnaire did not participate. Understanding cooperation in establishment surveys is particularly complex, since a person is responding for an establishment, where there may be both organizational influences on participation, such as the ability to provide requested data, and personal influences on participation, like personal interest in the survey topic.

We wanted to examine some of these complexities using the Community Health Measures Survey.

From July, 1998 through November, 1998, the University of North Carolina Survey Research Unit fielded this mail survey. We mailed a questionnaire to all local health departments in the United States that serve a jurisdiction of at least 100,000 residents (N=496). The survey collected information on general measures of public health practice, immunization status, and environmental health from local health department directors. The overall response rate, defined as the number of questionnaires returned with at least one of the three sections complete divided by the number of questionnaires originally mailed out, was 71%.

We have several pieces of information about respondents and nonrespondents for this survey which will inform our analysis. First, we have information on whether there was a potential gatekeeper and we plan to see whether response rates differ by the presence of this potential gatekeeper.

Second, we want to examine if response rates differ by size of jurisdiction. Health departments in larger metropolitan areas may be better equipped to provide data and thus more likely to participate.

Third, we want to look at access to e-mail. E-mail access may reflect greater technological sophistication and also greater likelihood of contact with outsiders seeking information. Health departments with e-mail access may be quite used to information requests and therefore more willing and able to provide survey data.

Lastly, we provided space for respondent comments on the back of the questionnaire. There were many comments which we think give us important insight into why some directors participated and others did not. We will code these comments to summarize and present them.

In summary, this analysis will investigate nonresponse in an organizational survey in light of various factors, such as the presence of a gatekeeper and size of jurisdiction. We will consider our results in terms of the nonresponse error associated with survey estimates and how we might reduce nonresponse error in future mail establishment surveys.

Controlling Nonresponse in the Current Employment Statistics Survey

Richard J. Rosen and Christopher D. Manning, U.S. Bureau of Labor Statistic

Timely collection of data is a top priority in all surveys. As a result, most surveys have procedures in place to control/minimize non-response. The Current Employment Statistics (CES) survey is no different from other survey agencies in this respect. CES is a monthly panel survey of about 380,000 business establishments conducted by the Bureau of Labor Statistics (BLS). CES provides one of the earliest estimates of employment, hours, and earnings at the national, State, and metropolitan area levels. CES data, widely viewed as a major economic indicator, are published each month after only two and a half weeks of collection. This restricted collection period places a huge burden on the collection method and magnifies importance of the procedures used to maximize response.

This paper describes what activities are performed in the CES program to maximize response while minimizing workload, costs, and respondent burden in an automated self-response collection environment. CES spends significant time and resources re-contacting non-respondents in an attempt to encourage them to report. These efforts involve setting a fixed 'cutoff' date for the receipt of data,

and following-up with those sample units that do not respond by the predetermined date. CES has developed a model for determining when to re-contact a non-respondent, and uses a mixed mode of contact for nonresponse prompting. The paper will discuss the mode, timing, and content of the various re-contact methods. We will measure the effectiveness of our re-contact efforts, including a comparison by mode between prompting non-respondents by phone, FAX, and e-mail. Each mode has its own advantages and disadvantages, which we will discuss as well.

Friday, 1:30-3:00

Imputation 1

An Evaluation of Imputation Techniques When the Missing-Data is Nonignorable

Susanne Raessler, University of Erlangen-Nuernberg, Germany

In multivariate datasets missing values due to item nonresponse may occur on any or all variables. Since nonrespondents differ systematically from respondents deletion of incomplete cases would lead to substantial bias as far as inference is intended to the population of all cases rather than the population of cases with no missing data. Therefore a variety of techniques to fill in missing data with plausible values have been developed. More and more statistical software packages provide techniques to impute missing values by values created from the observed data and make application for practitioners simpler.

Ad hoc approaches as well as model-based ones of missing-data procedures widely used in statistical practice rely at least implicitly on the assumption of ignorability of the missing-data mechanism. Thus a simulation study is performed to investigate the power of several imputation techniques especially when the data are not missing at random. The joint distribution of all variables, which should be adequately preserved by the imputation techniques, and, furthermore, possible biases of common estimates of means, variances and covariances based on the imputed datasets are the focus of this simulation study. Multivariate normal datasets as well as real datasets with continuous and categorical variables from German surveys are used. Imputation techniques like, for example, regression imputation and different hot deck procedures are investigated. Moreover, model-based methods using Markov chain Monte Carlo techniques are computed and compared with the simple case deletion method.

Due to the needs of German market research companies & acute; customers the study is restricted to single imputation procedures. Furthermore, a single imputation Monte Carlo solution using latent structure analyses and, if available, data from different sources shall be proposed to reduce the rate of nonresponse.

A Comparison of Imputation Methods Considered for the CES Redesign

Shail Butani, Paul Grden, and Larry Huff, U.S. Bureau of Labor Statistics

The Bureau of Labor Statistics' Current Employment Statistics (CES) Survey of monthly estimates of payroll employment, hours paid and earnings provides the business and financial communities with the earliest available indicator of current U.S. economic health. The program was established in the 1920's before probability sampling methods became the standard for sample surveys and the current design is based on a quota sample. In June 1995 BLS announced plans for a full redesign of the CES program based on a probability sample design. Estimates for all employment (ae), average weekly hours (awh), and average hourly earnings (ahe) are generated each month from the CES survey for varying industrial and regional detail. Estimates are adjusted each year to population employment totals from State Unemployment Insurance (UI) reports, which contain reported employment for all establishments.

Current plans for the redesign call for use of the complete survey data set for estimation. Imputation techniques have been developed to estimate for missing values in the case of both item and unit nonresponse. This paper examines the evaluation of the imputation methods that were considered for the CES redesign.

Data. The CES survey is an establishment survey with approximately 400,000 sample units. For each of these sample units, a longitudinal record was created for every CES data element. The population under study is the set of 173,069 CES longitudinal records that contain reported values for all employees (ae), production workers (pw), total weekly hours (wh), and total payroll (pr) for every month of the March 1994 to March 1995 time period.

In order to evaluate the accuracy of the imputation methods considered, we randomly assigned missing values to some of these reported data elements. Approximately 20% of the study population records were assigned missing values for every data element for every month to reflect anticipated levels of unit nonresponse for the CES redesign. The remaining 80% of the records were assigned item nonresponse patterns in proportion to those found among the CES longitudinal records (400,000 records) from the March 1993 to March 1994 time period. For each imputation method, the error of imputation is the difference between the imputed value and the reported value.

Evaluation Period. For the CES Redesign, an imputation method had to be selected for each of two different time periods:

- a. The initial (benchmark) month,
- b. The subsequent (estimation) months.

In our study, the initial month is March 1994 and the subsequent months are April 1994 to March 1995.

Imputation Methods. In the initial month, the employment from the UI report corresponding to each CES sample unit is used to impute for missing values. This source of employment data is essential in order to impute for unit nonrespondents. Two imputation methods were evaluated for the initial month:

- a. Proportional-to-Cell Average
- b. Hot-Deck Nearest Neighbor

For the subsequent months, four imputation methods were evaluated:

- a. Composite Imputation Method ($\lambda=1/2$)
- b. Composite Imputation Method ($\lambda=1$)
- c. Proportional-to-Cell Average
- d. Monthly Trend

Evaluation Criteria. We conclude the paper by presenting the results of our evaluation. Imputation methods were evaluated for both monthly levels and over-the-month changes for AHE and AWH. The absolute error between the reported values and the imputed values were measured at both the microdata level and the Industry Aggregate level for each Major Industry Division (MID) and for Total Private Industries. Results were then compared to determine which method to use in the redesign.

The Impact of Regulatory Exemptions on the Statistics of Income Partnership Studies

Paul B. McMahon, Internal Revenue Service, USA

Administrative records form the backbone of many sampling frames for establishment surveys, but the Internal Revenue Service goes beyond this limited function. The IRS uses tax forms as the survey instrument for its Statistics of Income series of reports. As a result, regulatory exemptions that allow "respondents" to withhold information under some conditions affect not only the stratification, but also generates missing information. We examine the impact of one such exemption, on reporting balance sheet data (asset holdings), for the studies of partnerships.

Friday, 1:30-3:00 Coverage Issues

Nonresponse Bias and Noncoverage Bias in Vaccination Birth Certificate Followback Surveys

Elizabeth R. Zell and Trena M. Ezzati-Rice, Centers for Disease Control, USA

Birth Certificate Followback Surveys are a convenient way to sample young children. However, two obvious problems exist with the use of a birth certificate sampling frame. These problems include: 1) locating the sample child from data available on the birth certificate and 2) the exclusion of children who move into the sample area after birth. Three pilot studies for The Childhood Immunization Birth Certificate Followback Survey (CIBCFS) were conducted in 1995 in West Virginia, Baltimore, and Maryland excluding Baltimore. This paper will focus on the methods and results used to evaluate nonresponse bias and noncoverage bias, associated with birth certificate followback surveys designed to measure vaccination levels in children two-years of age.

The CIBCFS used a birth certificate from of children who were 25 to 35 months of age and were born to women who were resident of the state or city at the time of the child's birth. The sample design was a stratified systematic random sample of birth certificates. The sample size for each site consisted of 600 randomly selected birth certificates with an equal number of birth certificates selected from each stratum. For study purposes, half of the sample within each stratum was randomly assigned to a telephone interview and the other half to a mail questionnaire. Vaccination history information was obtained from parent respondents of the identified children from the birth certificate frame. Data were also collected from vaccine providers for all surveyed children with either a written or verbal consent, dependent on the data collection method, to contact their provider(s).

Each CIBCFS pilot study area represented a different survey challenge. West Virginia is a predominately rural area and Baltimore is an urban area. One impact of these differences was demonstrated in the response rates obtained for each survey area, 83% in West Virginia, 73% in Baltimore, and 75% in Maryland excluding Baltimore.

Demographic data was obtained from the birth certificates for all sample children, both respondents and nonrespondents. Using the demographic data and vaccination status from sample respondents, logistic regression models were developed. These models were used to impute a probability of vaccination for each nonrespondent sample child. The demographic characteristics of each survey population played a part in the differences observed in the nonresponse bias for each area. Nonresponse bias varied from 2.0 percentage points in Baltimore to only 0.25 percentage point in West Virginia.

Birth certificates provide an easy-to-use sample frame. However, when the children being surveyed are currently two years of age, this sample frame excludes children who have moved into the survey area since birth. Another on-going immunization survey, the National Immunization Survey (NIS), a telephone survey, provided estimates of vaccination levels for children who are currently 19 to 35 months of age for each CIBCFS survey area. Adjusting the NIS data to reflect the vaccination status of children 24 to 35 months of age, noncoverage bias varied from 1.3 percentage points in Baltimore to only 0.16 percentage point in Maryland excluding Baltimore.

Change of Telephone Coverage Due to Mobile Telephones

Vesa Kuusela and Kai Vikki, Statistics Finland

Telephone coverage is changing rapidly in many countries because of growing popularity mobile phones. In Finland the proportion of households having a conventional phone has decreased all the decade. In 1990 more than 94% of households had a conventional but 1996 only 86%. Simultaneously the number of households having at least one mobile phone has grown from 7% to 42%. Since then, the number of the mobile phones has nearly doubled. Now there are even more mobile phones than conventional ones (presumably the mobile density in Finland is highest in the world). However, the telephone coverage of households is not known, because many households have several mobile phones.

The restructuring of telephone coverage does not happen equally in all population segments, however. Some people, e.g. young men living alone, have almost totally given up conventional phones, whereas, e.g. nearly all households having a stable position in

the society have both phones. In some population segments the mobile phones are also replacing conventional phones. For instance, one person households probably do not acquire a conventional phone anymore. In Finland, more 15% of households have only a mobile phone.

The changes in the telephone ownership have many effects on the surveys and on the survey practice because of a variety of reasons, e.g. changes in the coverage and changes in the selection probabilities. One important reason is that the status of a mobile phone is different from the one of the conventional phone, which is more like a domestic appliance. A mobile phone is more a personal appliance resembling a wrist watch. The owner has the mobile phone with him or her most of the time and the phone number is more for him or her than for the household. Therefore it may be easier to contact people who have a mobile phone, but interviews may not be possible e.g. because the respondent may answer the phone almost anywhere.

The topic of the presentation is a survey conducted in Finland concerning the present telephone coverage. A similar survey was done 1996 focussing more on conventional phones, however. A new survey was done in the beginning of the year 1999, focussing more on mobile phones now. The purpose is to find out whether the previously observed trend continues and what is the actual telephone coverage at the moment. An important issue is to find out to which extent mobile phones have replaced conventional phones and how the structure of telephone coverage has changed in different population segments.

Methods for Compensating for Nonresponse Bias in the National Immunization Survey

Philip J. Smith, CDC

J.N.K. Rao, Carleton University

Danni Daniels, CDC

K.P. Srinath and Michael P. Battaglia, Abt Associates

Monina Klevens, CDC

The National Immunization Survey (NIS) is a large Federally sponsored random-digit-dialing survey designed to yield estimates of vaccination coverage of children in the age range of 19-35 months in each of 50 states and 28 urban areas. In the NIS over 900,000 households with phones are screened each year for eligible children. From this process approximately 34,000 telephone interviews are obtained annually from households that have children in the targeted age group. In the RDD portion of the NIS, an adult respondent is sought who is knowledgeable about the household's eligible children's immunization history. Approximately 85% of the time, this respondent is the mother of the household's eligible children. In addition to providing information about the demographic composition and socioeconomic status of the household, the respondent is asked for consent to allow NIS personnel to contact each eligible child's immunization providers. Among children for whom consent is obtained, a mail survey is conducted to obtain sampled children's immunization data from their providers. This data is used to determine whether they are up-to-date with respect to immunization schedules recommended by the Advisory Committee on Immunization Practices.

This paper provides an overview of the statistical methods used by the NIS to guard against two important threats to the validity of the findings of the survey, namely, (i) overestimation of immunization rates due to failure of the phone survey to obtain information from children living in households that do not have telephones, and (ii) the unpredictable bias that may be expected from using data from responding immunization providers, only.

To compensate for noncoverage of households without telephones, we draw on work by Keeter (1995) that suggests that some households that do not have a telephone at a particular time are actually between spells of telephone service. Our studies show that such spells are common and that households that have spells of "phonelessness" share similar characteristics with households that never have phones. Also, many of these similar characteristics have been shown to be positively correlated with under-immunization. In our paper, we review methods described by Frankel et al. (1998), Brick, Waksberg & Keeter (1996), and Politz & Simmons (1949) that use information on spells of phonelessness to adjust sampling weights appropriately so as to attenuate bias attributable to noncoverage of households without telephones.

To compensate for nonresponse occurring when no shot information is given by a shot provider, we use work by Little (1986), Czajka, et al. (1992), and Eltinge (1996), to form adjustment cells by using response propensities and predictive mean values. Our results show that children for whom no provider information is available have characteristics associated with a lower probability of being up-to-date with respect to recommended childhood immunization schedules. In our paper, we show that compensating for provider nonresponse by forming adjustment cells using response propensity or predictive mean values attenuates potential bias that would be incurred by only using data from children for whom provider immunization data is available.

Friday, 3:30-5:00

Culture and Survey Nonresponse *Timothy Johnson, Diane O'Rourke, Linda Owens, and Jane Burris, University of Illinois, USA*

We explore the degree to which survey nonresponse may be associated with various elements of respondent culture. Using race and/or ethnicity as proxy indicators of respondent culture, empirical evidence from a variety of sources, including cross-sectional and longitudinal surveys, and assessments of Census undercounts, are reviewed. These data generally suggest that unit nonresponse is greater among minority households, compared to nonHispanic white households, in the United States. We subsequently consider relevant theories developed by anthropologists, sociologists, cross-cultural psychologists and communications researchers that may be useful in interpreting these findings. One approach we consider is "unpackaging" culture by discovering the specific values, norms, behaviors and orientations through which it operates. In this regard, several cultural "syndromes" are considered for their relevancy to survey nonresponse, including individualistic vs. collectivistic orientations, power distance, masculine vs. feminine orientations, uncertainty avoidance, and emphasis on vertical vs. horizontal relationships. Variable communication styles are similarly evaluated, including context requirements,

reliance on nonverbal behaviors, and self disclosure patterns. Patterns of social participation influenced by culture, such as minority oppression and opposition, social distance, and helping behavior are also reviewed. We conclude by proposing a conceptual model to account for cultural influences on survey nonresponse.

Friday, 3:30-5:00

Imputation for Wave Nonresponse - Existing Methods and a Time Series Approach

Danny Pfeffermann and Gad Nathan, Hebrew University, Jerusalem

This paper reviews methods for imputing missing values in longitudinal surveys. The methods considered include simple and weighted group means, nearest neighbors, regression models with exogenous and endogenous regressors and time series models. The latter models are used to account for the time series relationships between individual measurements and higher level group effects like households. The models combine standard multi-level models operating at given points in time with time series state-space models for the random group effects and individual measurements. The unknown parameters of the various models are estimated by maximum likelihood estimation methods.

The different methods are illustrated and compared in terms of bias and variance by use of a simulation study. The study distinguishes between different nonresponse mechanisms (missing completely at random, missing at random and informative missing), and different sampling patterns (continuous sampling, rotating panel surveys). The use of multiple imputations is also considered.

Friday, 3:30-5:00

Use of Administrative Records

An Example of the Use of Registers to Measure Nonresponse Bias in Swedish Surveys

Jan Hörngren, Statistics Sweden

Nonresponse is a major problem in surveys. Nonresponse always results in errors in survey estimates. Nonresponse causes two types of errors:

- i. Sample errors. Loss of observations leads to an increased variance.
- ii. Bias. Nonresponse usually introduces selection bias (nonresponse bias).

Increase in variance (i) is generally a minor problem in comparison with nonresponse bias, since nonresponse bias leads to erroneous estimates and the confidence statements will be invalid.

The three main strategies against nonresponse (bias) are:

1. Taken measures with the purpose to reduce the nonresponse to insignificant levels so that remaining nonresponse causes little or no bias
2. The use of special designs for data collection and estimation, i.e., subsampling of nonrespondents according to Hansen and Hurwitz.
3. Construction of estimators that "adjust" for nonresponse bias. These estimators are based on model assumptions about the nonresponse mechanism and about relations between variables.

If we look at Strategy 1 the main question is: What nonresponse rate causes little or no bias? What happens with the bias in Strategy 2 if we fail to obtain response from elements in the subsample? And if we look at Strategy 3 the main question is: Are the model assumptions we make realistic?

To answer these questions we need knowledge about the nonresponse bias concerning the variables of study. In this paper we will study the possibility to measure the nonresponse bias by using registers. In Statistics Sweden's survey *-Activity After Graduation-* the main aim is to study the population employment situation. In Statistics Sweden's *Register of Employment (RE)* we have information on the employment situation for the total population in Sweden. "RE-employed" is closely related to "employed" as defined in the survey (Activity After Graduation). In a postsurvey procedure it is possible to match survey data with RE-data on an individual level. We have the possibility of estimating the number of RE-employed based on the total sample in the postsurvey data. We also have observational values for nonrespondents. The following estimates can be calculated:

\hat{Y}_{NRB} = The estimate of the number of RE-employed based on the total sample (s)

\hat{Y}_{NRB} = The estimate of the number of RE-employed according to the respondents (r) in the survey

$\hat{B} = 100 \left(\frac{\hat{Y}_{NRB} - \hat{Y}_{NRB}}{\hat{Y}_{NRB}} \right)$ = The nonresponse bias in percentage of \hat{Y}_{NRB}

$|\hat{B}|$ gives a measurement on the nonresponse bias in the estimates of the number of RE-employed. Since RE-employed and employed according to the survey are rather closely related, the results are clear indications of the degree of nonresponse bias in the survey.

Reconciling 1996 CPS and SIPP Money Income with Administrative Records

Marc I. Roemer, U.S. Bureau of the Census

Many researchers have expressed interest in comparing money income measured in the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) with independent benchmarks such as the National Income and Product Accounts (NIPAs). Such a reconciliation provides a summary measure of overall response error, and allows a judgement about the quality of the CPS and SIPP income estimates. This paper updates the work of John Coder and Lydia Scoon-Rogers who reconciled 1990 income, and revises their methodology. Estimates of 1996 aggregate income from the March 1997 CPS and the 1996 SIPP Panel are presented against compatible benchmarks for 17 categories of income and considered next to previous reconciliations for 1984 and 1990.

Differences in the definition of income and the coverage universe comprise the greatest part of the reconciliation challenge. The NIPAs include imputed income and adjustments that are outside the survey paradigm, and define certain entities such as fiduciaries and nonprofit organizations serving individuals as persons. Income in these categories, as well as that of decedents, the military, the overseas and institutionalized populations, and other types of income must be removed from the NIPA measures to make fair comparisons to the survey estimates. The paper presents methodology for making some of these adjustments to the benchmarks.

Tabulation of IRS and SSA exact-match data with the CPS and SIPP will explain shortfalls and overestimates in the surveys relative to the NIPA earnings and property income estimates. Analysis will determine which type of error, item refusal, miscategorization, or amount misreporting, has the greatest effect on the aggregate. Tabulations of aggregate income by income range will show which part of the income distribution contains the most serious response errors, and further investigation will determine common characteristics of the nonrespondents and misreporters.

The Nature of Nonresponse in a Medicaid Survey: Causes and Consequences

Patricia Gallagher and Floyd Jackson Fowler, Jr., University of Massachusetts Boston, USA

Medicaid programs throughout the United States are surveying members to evaluate the quality of their medical care experiences. These surveys are most often done by mail, sometimes with telephone follow-up of mail nonrespondents. The response rates for mail surveys of this type typically range from 20 to 50 per cent. Telephone efforts add another 5 to 15 percentage points to these rates. In order to better understand the significance of the nonresponse, we conducted a study among Medicaid recipients that involved intensive follow-up using three modes of data collection.

Starting with a sample of 3200 Medicaid members, half adults and half children under 18, a standard mail survey protocol was implemented: an initial mailing, a reminder postcard, and a questionnaire replacement mailing to nonrespondents. Next, an effort was made to conduct telephone interviews with those who did not respond to the mail protocol. Then, an effort was made to interview those who responded to neither the mail nor the telephone protocols by sending an interviewer to the sampled person's home, to collect an interview in person.

Thus, we will be able to compare respondents to each phase of data collection with respect to all the topics covered in the survey: demographic characteristics, measures of health status, experience with respect to access and quality of care, and ratings of the health services received. In addition, we will collect administrative record data. Those data will permit comparisons of respondents to each phase and those who never responded at all (including those who refused and those whom we could never find) with respect to age, gender, ethnicity, and amount and types of health services used.

The analyses will focus on both the significance of nonresponse for Medicaid surveys and the procedural implications of what we learned for collecting better data from Medicaid members. With respect to the former issues, we will describe the ways in which nonresponse at each phase of data collection affects the representativeness of the sample responding and the key descriptive results that the surveys are intended to obtain. In addition, we will look at each phase of data collection to identify the reasons for nonresponse. For example, for the mail component, we will calculate the number of nonrespondents whom we could never locate, those whom we eventually located, but had an initial bad address, those who refused once they were contacted (who usually cannot be identified in a mail survey), those whose first language is not English, leaving those who were willing and able to respond, but did not. The same kind of analysis will be done for the phone and in-person phases of the data collection. A particularly important practical implication is to identify the extent to which surveys could be improved and what the data implications would be of having the Medicaid programs do a better job of updating address and telephone information.

Survey data collection was complete in the fall of 1998; the record-check component will be completed in April, 1999. We believe this is one of the most comprehensive studies of the nature of nonresponse -- its reasons and consequences -- ever done of a Medicaid population.

Friday, 3:30-5:00 Calling Strategies

An Elaborate Calling Strategy - Does it Make Enough Difference?

Andrej Kveder and Vasja Vehovar, University of Ljubljana

This paper addresses the optimal calling strategy in telephone (CATI) surveys. The research issue is thus to define a calling schedule that would provide optimal data quality within a given time, budget and data-quality constraints.

A surprisingly limited research effort is dedicated to this problem, particularly when compared with the widespread daily practice of telephone interviewing. The overview of published research shows that the majority of studies deal only with the first-call analysis, with the specifics of certain calling periods, or with respondent-related characteristics. Very rarely explicit models are used, such as logistic regression or the Markov-chain models. An important difficulty arises here from the conditional dependence of the transition

probabilities, which makes any modelling extremely complicated. The complexity of the problem even poses the question, whether it is worthwhile to refine the research (and the calling strategy itself) much beyond the basic findings about best day-shift times.

The research in this paper is based on the meta-database created from 50 survey projects. The database contains more than 500,000 recorded calls and a history up to ten-calls is also recorded. With this data, the empirical probabilities for specific calling outcomes (conditional on calling history) were calculated and Monte Carlo simulations were run using these transition probabilities. The results confirm those of other research, such as the advantage of weekday's evening and the weekend's day shifts. The data also enable other, more refined analysis, such as the impact of the calling history on the probability of a successful call, or the empirical evaluation of different (but very elaborated) calling strategies.

Of course, the calling procedures can not be treated separately from the timing, costs and errors. The relation to the non-response error is particularly critical and it is also incorporated into the model. The extent and the behaviour of the specific non-response error is extremely important when we define the maximum number of call-attempts for a certain survey.

Analysis of Call Patterns in a Large Random-Digit-Dialing Survey: The National Immunization Survey

J. Michael Dennis, Michael P. Battaglia, Ann-Sofi Rodén, David C. Hoaglin, Candice Saulsberry, and Martin Frankel,, Abt Associates, USA

Nancy A. Mathiowetz, Joint Program in Survey Methodology, USA

Philip J. Smith and Robert A. Wright, Centers for Disease Control, USA

Call patterning in random digit dial (RDD) telephone surveys involves scheduling one or more call attempts according to a predetermined sequence. Ideally, optimal call patterning minimizes the number of call attempts required to have an adult respondent answer the phone and then complete the eligibility screen and, if appropriate, the interview. A CATI call scheduler based on analyses of calling patterns has the potential to reduce the costs of an RDD survey without compromising the contact rate nor the response rate. Optimal call patterns may be identified through analysis of contributing influences such as the timing of call attempts (time of day/day of week intersections), seasonality, and environmental factors associated with a given telephone exchange such as median household income, urbanicity, geography, and others. To identify optimal call patterns, we analyzed the call record files of the National Immunization Survey for 1995-1998. The NIS employs a large RDD telephone survey having as many as 10 million call attempts per year and having 9am-9pm/7-days-a-week coverage for all 50 States and the District of Columbia. Results show the rank ordering of call patterns by probability of case resolution (household/non-residential/out-of-scope), screener completion, and interview completion--separately and all combined. The work also examines the benefits of using environmental and demographic information available at the telephone exchange level to select call patterns appropriate for various types of telephone exchanges (e.g., rural/urban and high/low income).

Friday, 3:30-5:00

Respondent and Interviewer Effects

Respondent- and Interviewer-Related Causes of Unit and Item Nonresponse in Repeated Measurements: A Multilevel Longitudinal Model

Pieter van den Eeden and Sandra Geerlings, Free University of Amsterdam, the Netherlands

Unit-nonresponse (resulting in 'missing cases') and item-nonresponse (resulting in 'missing values') can tremendously attack the quality of data in social survey research. This is especially the case if there are repeated measurements. One of the most striking effects concern the bias of the estimated parameters. Although until now much effort has been paid on mitigating the biasing effects of both types of nonresponse, for example by the recently developed procedures of imputing missing data, unfortunately the study of their causes is weakly developed. Unit-nonresponse stems from inability to contact the sample person, inability of the sample person to provide responses to the survey, or refusal of the interview request. Item-nonresponse can also occur for various reasons: the interviewer fails to ask the question, the respondent refuses to provide an answer to a question, the respondent is not able to provide an adequate answer to a question, or the interviewer fails to record the answer provided. All these causes are related to respondent characteristics and interviewer characteristics.

This paper focuses on respondent- and interviewer-related causes of unit-nonresponse and item-nonresponse in longitudinal studies and aims to offer a general model of their analysis. Since the separate measurements of a single respondent in a longitudinal study can be viewed as being nested under respondents, and respondents can be viewed as being nested under interviewers, the multilevel model of analysis is appropriate. The levels of interest are measurements, individuals and interviewers. We define both unit-nonresponse and item-nonresponse as dependent variables of a dichotomous nature; hence the logit regression type of multilevel analysis was used. As a result, the outcome has to be interpreted in terms of probabilities on the types of nonresponse depending on the independent variables. Since the analysis concerns repeated measurements, the multilevel longitudinal model of analysis has been used. In this type of model the repeated measurements are ordered along a time-scale, with respondent-dependent varying time-intervals. In the analysis of item-nonresponse the missing cases as a result of unit-nonresponse at the varying occasions will not be a problem, since only the covered individuals at a given occasion belong to the database.

The data of illustration are adopted from the LASA (Longitudinal Aging Study Amsterdam) project, consisting of about 3200, 55-85 years old people from a nation-wide representative sample in the Netherlands. From this 662 respondents were incorporated in a follow-up study on the course of depression. They have been interrogated eight times, with a between-measurement distance of approximately six months. In the first and eighth time the respondents are interrogated in a face-to-face situation by interviewers and the other times by a paper-and-pencil procedure. The item-nonresponse analysis concerns 20 items of the Center for Epidemiological Depression Scale (CES-D), being assessed eight times. The CES-D scale has been developed for assessing minor depression. Since the study incorporates both interviews and self-administered mail questionnaires, the database enables us also

to get some insight in the net effect of the interviewing mode in addition to the time-, interviewer- and respondent effects that are mentioned in the above.

Quality of Interview, Total Nonresponse and Item Nonresponse in the Italian Household Budget Survey

Nicoletta Pannuzi and Donatella Grassi, Statistics Italy

Since 1996 Istat introduced quality control methodologies in order to monitor the quality of fieldwork in the Household Budget Survey (HBS) and in particular the performance of the non-professional interviewers used for the face-to-face interviews at the municipality level. Therefore every month CATI re-interviews has been carried out on a sub-sample of households included in the HBS to evaluate the correctness of the way the interview is performed. The results obtained with the CATI re-interviews can be used to verify first of all the link between the quality level of the face-to-face interview and the Total Non Response Rate (TNR) and, as a second step, the possible connection between quality standard and Items Non Response Rate. As a first step, for the period 1996-1998, we analyse the link between the percentage of incorrect interviews and TNR under the hypothesis that incorrect interviews hide total non-response. As a second step we try to estimate, using the sub-sample of CATI re-interviews, the effect of the quality of the interview on the INR for the main variables surveyed.

Measuring Nonresponse in a National Longitudinal Survey of Children

Paula C Baker, Ohio State University, USA

This paper examines nonresponse in the Children of the National Longitudinal Survey of Youth. Since 1986, the children of the NLSY79 mothers have been administered a battery of cognitive, socio-emotional, and physiological assessments. This dataset of more than 10,500 children contains extensive information on maternal and infant health, child development, maternal employment history, family income, childcare, education, and a large number of background variables. The introduction of CAPI in 1994 to replace the previously paper child questionnaire offers an opportunity to address the implications of moving to a new data collection mode in the context of a longitudinal survey. We propose to examine the role of interviewer, mode of data collection, task burden and respondent characteristics on the degree to which children participate in the survey and complete a range of assessments and sensitive questions. The analysis will document field practices related to the use of mixed mode design and refusal conversion to reduce nonresponse by young respondents and their mothers.

Since its inception, retention of respondents in the main NLSY79 has remained high, particularly among mothers in the cohort. Of the children available for interview in each survey round, nearly 90% are typically administered the age-appropriate assessments. The nature of the sampling linkages between children and mothers mean that some children are unavailable for assessment because their residence and the status of their mothers. The causes and distribution of noncoverage in the main NLS Youth/79 surveys have been studied and are only briefly discussed here.

The focus of the paper is overall child nonresponse, selective nonresponse for the child assessments, and item nonresponse in other substantive areas of the child survey. The analyses assume that socio-demographic characteristics and attitudes of both interviewer and respondent as well as the nature of the assessment instruments and environment contribute to the level of reporting. To date, two forms of the primary dependent variable have been examined: (1) whether a child was assessed (interview-level) and (2) whether particular tests were completed (assessment-level). Analyses based on the data from pre-CAPI years indicate that the principal causes of assessment noncompletion are interviewer error and premature termination due to child nonresponse, parent refusal or other interference. Since the dataset contains measures of interviewer characteristics as well as child and mother background traits, it is possible to identify the conditions under which each test was administered and to analyze the general assessment environment. Assumptions about the effects of interviewers who have returned to the same household over the years will be tested. This process can be supplemented by incorporating into analyses the child's attitudes toward testing, physical problems that may have affected test performance, any noticeable interference, the nature of the testing site, and the mode of instrument administration. The dataset allows one to examine the individual and joint effects of child, mother, interviewer, task, and mode of interview on reporting.

Preliminary analyses indicate that certain child characteristics are related to noncompletion at both the interview and assessment level. Overall child interview rates differ somewhat by race. The level of assessment completion varies across age and race categories. A mixed effect model for measuring effects is anticipated. The model will ultimately be extended to incorporate both individual and interviewer specific covariates. An initial series of logit models relating respondent characteristics to assessment completion included categorical measures of race, sex, maternal education, number of siblings, father absence from the household, childbearing age of mother, and whether the child had a significant health problem. Hispanic, male, and urban children were less likely to have received a score on certain cognitive assessments for the very young, while children of mothers who were not high school graduates were less likely to receive a verbal memory score. Father absence and urban region of residence affected the likelihood of completing the achievement tests. Race and maternal education were also determinants of receiving a score on the assessment of receptive vocabulary.

Analyses on the effects of interviewer characteristics such as age, education, race, and interviewing experience are as yet inconclusive. Interviewer ratings of child's attitude, rapport, perseverance, cooperation, and motivation will be used as a proxy for interviewer personality set toward the child. The degree to which certain groups of children prematurely terminated a particular assessment or testing session will be examined.

Finally, a comparison of interview mode will be made to identify the impact of introducing CAPI to administer previously paper and pencil assessments. Patterns of response will be examined to identify differential response and to establish whether there is a change in the patterning of the valid versus missing responses as a result of the difference in mode of administration. Several child and home background items have shown substantial increases in levels of response in CAPI due to enforced skip logic and question filtering. Initial reviews indicate that respondents and interviewers report increased speed of administration and heightened interest, particularly by children, in the CAPI survey. Such perceptions of reduced respondent burden and interest in the CAPI mode may affect the willingness of respondents to participate and complete a lengthy battery of assessments.

The NLSY79 Child data offer an opportunity to examine multiple sources of nonresponse in a large sample, using a comprehensive set of variables. This type of examination should benefit future waves of longitudinal child research including the NLSY79 and such counterparts as the PSID Child Development Supplement, the British National Child Development Study, and the Canadian National Longitudinal Study of Children and Youth.

Saturday, 8:30-10:00

Impact of Missing Values From Converted Refusals on Nonsampling Error

*Robert Mason and Virginia Lesser, Oregon State University, USA
Michael W. Traugott, University of Michigan, USA*

Efforts to complete interviews with hard-to-reach respondents are a common strategy employed in sample surveys. Even when improved response rates are reported, the problem of error related to item nonresponse remains. This study evaluates the magnitude of nonsampling error and bias associated with item nonresponse. It focuses on the hypothesis that, while repeated callbacks and refusal conversions bring people into the sample who otherwise would be lost, they also may produce higher levels of item nonresponse or increase nonsampling error. The hypothesis is tested on data from a statewide RDD telephone survey of 1,483 adults. The study concerned the selection of a new U. S. senator chosen at a 1996 special vote-by-mail election in Oregon. An extensive converted refusal attempt was successful in adding 377 cases to the sample, representing nearly half of all the refusals. A logistic regression model for binary responses was fit to the data. Observed and predicted estimates for candidate choice are presented for converted- and non-refusal respondents. We employed three approaches for handling missing data. One approach reports the results in which all missing values are dropped from the analysis. A second approach reports results from a "hot deck" imputation of missing data. The third approach is based on a multiple imputation adjustment. The pursuit of additional respondents produced higher levels of item nonresponse (24% vs. 11%) but no discernable bias in voter choice or significant differences in nonsampling error. Results are evaluated in terms of the extra cost required to obtain additional respondents.

Saturday, 8:30-10:00

Variance Estimation from Survey Data Under Single Value Imputation *Hyunshik Lee, Westat,*

*Inc., USA
Eric Rancourt and Carl E. S@ rmdal, Statistics Canada*

This paper reviews recent contributions to the theory of variance estimation in surveys when single value imputation is used for missing values. Single value imputation, in contrast to multiple imputation implies that a single imputed value is created to take the place of a missing value. A number of contributions to this topic have appeared in the recent literature.

The topic is important since survey nonresponse often reaches high levels, and resources are insufficient to renew contact with respondents or to obtain by other means the desired but missing values. Imputation is then usually resorted to. It is a common practice to use single imputation methods for this purpose and use ordinary variance estimators as if imputed values were observed. However, this approach could lead to a severe underestimation of the true variance. It is hoped to rectify the situation to the extent possible by providing a review on the topic and useful recommendations.

The paper is developed around three aspects from which this variance estimation problem can be examined:

- (1) the approach taken to the variance estimation;
- (2) the imputation method(s) used to complete the data set;
- (3) the sampling design and the prototype estimator used for point estimation.

After the theoretical review of various methods, some empirical results and discussion are presented.

Finally, some useful recommendations are given and some outstanding issues on the topic are discussed. One of such issues is the development of software.

Saturday, 8:30-10:00

Interaction Between Unit and Item Nonresponse

Item Nonresponse as a Predictor of Unit Nonresponse in a Panel Survey

Geert Loosveldt, Jan Pickery, and Jaak Billiet, K.U. Leuven, Belgium

In face-to-face interviews, both unit non-response and item non-response can be considered as a negative reaction of the respondent. Item non-response is a negative reaction to a question of the questionnaire asked by the interviewer. Unit non-response is a negative reaction of a sample person to the specific request to participate in an interview. Given this basic common element, we hypothesize that the same factors are responsible for unit and item non-response and that both kinds of non-response are related to each other. To investigate the relationship between item non-response and unit non-response a panel survey can be used. With panel data the amount and pattern of item non-response realized during the first wave of a panel can be used to predict the unit non-response in a second wave of a panel. The amount of item non-response refers to the number of unanswered questions. We expect that respondents characterized by a high degree of item non-response in an interview will be more likely to refuse to the request of the interviewer to participate in a subsequent interview. The pattern of item non-response refers to which questions are sensitive for item non-response. Relevant characteristics of the questions are : embarrassing or threatening questions, difficult questions and substantive important questions (e.g. political preference in an election survey). We expect that item non response on threatening or difficult questions which are strongly related to the substantive topic of the questionnaire will be a good predictor for unit non-response. Therefore, we assume that the interview is a negative or unpleasant experience for respondent with item non-response for these questions and that respondent with such a negative experience are

more likely than others to refuse the second interview.

In testing these hypothesis, we will control for other respondent characteristics that are correlated with item non-response and also can effect unit non-response. Some of these respondent characteristics can be related to the cognitive aspect of answering question (e.g. age, education).

For our analysis the Flemish data are used from the first (1991) and the second (1995) wave of the Belgian General Election Study. The sample is representative of the Flemish population aged 18-75 years.

"Don't Ask Me Nothin' About Nothin', I Just Might Tell You the Truth" .The Interaction Between Unit Nonresponse and Item Nonresponse"

Nick Moon, NOP Social and Political, UK

Dr Heather Laurie and Jonathan Burton, University of Essex, UK

Organisations conducting surveys usually concentrate far more of their efforts on minimising unit non-response than on minimising item non-response. This is partly because the potential damage to survey quality caused by non-response is far greater than that caused by item non-response.

Also, although unit non-response is more prevalent, it is also generally easier to rectify. Reducing unit non-response is achieved in the long term by interviewer training, and in the short term by various response maximisation. If, despite this, the final response rate is deemed unsatisfactory, further steps can be taken to increase it. In the case of item non-response, the main means of reducing it is again interviewer training. Post-survey, there is virtually no chance of converting item non-response into response.

There is, however, a third component, which is less discussed, and this is the interaction between unit and item non-response. There is no doubt that there is some correlation - people who are more reluctant to take part in surveys will also be more reluctant to answer individual questions, and vice versa. This paper examines that interaction, and its implications. In one direction - that of item non-response tending to lead to unit non-response - the interaction can only be measured on a longitudinal study, and the paper presents some evidence from the British Household Panel Study.

Interviewers fill in a short questionnaire about the conduct of the interview, including a measure of how co-operative each respondent was during the interview. This question not surprisingly correlates with item non-response, and is also a good predictor of future unit non-response. Similarly, respondents who refuse to answer the key income questions are more likely to refuse the whole survey in a subsequent wave.

Response rates on the survey are exceptionally high, with a wave on wave re-interview rate of around 97%. With so little scope to reduce unit non-response, our efforts turned to reducing item non-response, although despite concentrating on this at briefings, we have not been able to reduce this significantly. This then raised the question of whether there is an interaction in a different direction that of low unit non-response leading to higher item non-response. In theoretical terms, one can explain this by arguing that a very high response rate means that people who would not normally take part in surveys do take part in the panel survey. Because they are reluctant to take part in surveys at all, they are more prone to refuse certain questions. In most surveys they would not have taken part at all, and so would not have appeared as item non-response.

The paper looks at this issue in two principal ways. Firstly, wave by wave response patterns are used to examine the propensity for people who have refused at least one wave but taken part in others to refuse certain key question. Secondly, BHPS interviewers were sent a questionnaire to ask how hard it was to persuade each respondent to take part, and this was compared with propensity for item non-response.

Unit- and Item-Nonresponse in the Bern/Munich Lifestylepanel

Schotten K and Spegel H, Ludwig-Maximilians-Universität, München, Germany

Niemann S, Abt. Gesundheitsforschung, Universität Bern, Germany

Weitkunat R, Ludwig-Maximilians-Universität, München, Germany

Crispin A, Öffentliche Gesundheit, München, Germany

Janssen CH, Institut für Medizinische Informatik und Systemforschung, Germany

Abel T, Abt. Gesundheitsforschung, Universität Bern, Germany

Our presentation will describe the methods, results and conclusions of the unit and item nonresponse analysis of a study on health-related lifestyles.

Patterns of health-related behaviors of adults between 55 and 65 years of age were analysed in two consecutive surveys. From November 1995 to February 1996, 791 computer-assisted telephone interviews were conducted in Munich. The questionnaire included about 250 questions and covered the following aspects: sociodemography, socioeconomy, anthropometry, health-related behaviors, subjective health status, sense of coherence, health locus of control, health related orientations and life events. Additionally, the variables sex of the interviewer, date of the interview and length of the interview were analysed.

Item-nonresponse behavior was to be investigated for the questions concerning household property, monthly income, and satisfaction with sexuality. Gender-specific nonresponse could not be observed. The screening of numerous other variables with respect to nonresponse resulted in a higher readiness to respond in interviewees with high-school/college or university degree regarding the questions on property and income.

A second survey with the persons being interviewed in the first survey were re-interviewed from November 1996 to February 1997. The aim of this second survey was to investigate the question, whether the identified health-related patterns of behavior were stable or changed within the interval of one year. The question whether reinterviews took place randomly or dependent on

systematical influence is the content of the second part of our study.

Predictors of unit nonresponse were analysed via multiple logistic regressions. The risk of not participating in the second survey was 1.4fold higher in interviewees assessing their health as being rather poor. This result confirm the findings of other studies. Additionally, rare regular meals (1.7fold risk of not taking part) as well as discontent with housing conditions (2.2fold risk of not taking part) were associated with not taking part in the second survey. In the event of all three risks for not participating, the unit-nonresponse rate was above 50 percent.

Saturday, 8:30-10:00

Special Populations and Nonresponse 1

Reducing Nonresponse in Surveys of Vulnerable Populations

Dretha Phillips and John Tarnai, Washington State University, USA

Surveys of incarcerated, abused, welfare, and adolescent populations pose special challenges for researchers. Features that promote high response rates or, in their absence, defensible accounting for nonresponse are more than offset by factors that inhibit response, yielding unacceptably high nonresponse rates for these vulnerable populations. A few examples should make the point here.

Among the advantages offered by surveys of incarcerated and welfare populations, in particular, are sampling frames with administrative data on the subjects and survey topics with high saliency for respondents. However, sample information often is incomplete or out-of-date. Literacy levels often preclude mail or other self-administered modes, and telephone interviews are not an option for the substantial proportion of the sample without working telephone numbers or reliable access to a telephone. Primary among the disadvantages, perhaps especially for abused and adolescent populations, is that responding to a survey may be perceived by the subjects as a risky enterprise. Adolescents present a double nonresponse problem: Prior permission to interview the adolescent must be obtained from parents or guardians, and then consent must be obtained from the targeted respondent for the interview. Such a two-phased procedure for reaching survey respondents--a not uncommon protocol with vulnerable populations--produces two sets of response rates, which are multiplicative in generating a final overall response rate. Nonresponse in either of these phases will exacerbate the final response rate and increase nonresponse error.

Further compounding nonresponse in vulnerable populations is the issue of screening samples, which typically is how vulnerable populations are identified. Again, this is a two-phased process in which the first phase involves screening for eligible participants and the second phase entails approaching those eligible persons for an interview. How we estimate the proportions of eligibles and ineligibles from the non-contacted and refused proportions in the first phase can have significant effects on nonresponse error. A final example offered here is that human subjects review boards can exacerbate nonresponse in vulnerable by imposing particular research protocols for these subjects' protection.

The purpose of the proposed monograph paper is to provide a state-of-the-art assessment of techniques that are and may be employed to reduce nonresponse in surveys of vulnerable populations. Toward that end, the paper will be organized along these lines. First, extant research on the four vulnerable populations identified here will be reviewed and synthesized, taking care to specify nonresponse contingencies that they have in common and that distinguish them from other survey populations. This review will encompass not only literature published in professional journals, but also unpublished reports from the archives of research organizations such as the Social and Economic Sciences Research Center, the Institute for Social Research, Abt Associates, Westat, and Rand Corporation. Second, techniques or elements of the survey process will be identified that are effective in reducing nonresponse, as well as nonresponse contingencies that appear immutable, in the four populations. Third, a future research agenda will be proposed to evaluate methods of increasing response in surveys of vulnerable populations.

Unit and Item Non-Response across Multiple Methods: Surveys of AARP's Members

Ada-Helen Bayer, Teresa A. Keenan, Katharyn Marks, American Association of Retired Persons, USA

Decreasing unit non-response and obtaining high quality data -- particularly at manageable costs in quick turn-around situations -- continue to be important issues in survey research. To explore these issues, we conducted experiments with AARP members using the same questionnaire with seven mail, two telephone, and one mixed-mode methods.

We present analysis of unit and item non-response effects within and across survey modes. Within the mail mode, we examined seven variations or methods: (a) random sample single contact, (b) random sample single contact in a machine scannable format, (c) four random sample multiple contact versions, and (d) a single contact using a random sample drawn from a consumer panel. The four multiple contact mail versions contained variations in the type of pre-notification mailing (letter or postcard) and the mail class of the first survey mailing (priority or first class). Within the telephone mode, we assessed the unit and item non-response levels with random digit dial and randomly drawn list-assisted samples. We explored one method using both mail and telephone: respondents received a mail questionnaire but were instructed to respond using an 800 telephone number. With the exception of the samples for the consumer panel mailing and the random dial telephone interviews, all samples were randomly drawn from the membership database. Respondents in the random digit dial telephone and mail panel methods were screened for organization membership. In addition, for the enhanced mail and scannable mail methods with samples drawn from the AARP membership database, we compared responders to non-responders on key demographic variables.

We present results that show significantly different unit and item non-response effects across some of these methods. For example, the consumer panel mail method had a higher response rate than the other methods and, within this sample, responders were younger than non-responders. The lowest response rate was to the mixed mode method, which also had the highest item non-response rates on all demographic questions except income. On most questions in a series of Likert scale items, responders to the second mailing of the enhanced mail method displayed higher missing levels than those responding to the first mailing in that

method and than responders in the other methods.

Persons with Disabilities: Nonresponse, Proxy Response, and Assisted Response

GE Hendershot, L Colpe, and PC Hunt, National Center for Health Statistics, USA

In *Nonresponse in Household Interview Surveys* (New York: Wiley and Sons, 1997), RM Groves and MP Couper draw on social psychology to develop a conceptual framework for participation in household interview surveys. They conceptualize interviews as social interactions between interviewers and householders with two phases: contact and cooperation. They then conceptualize relationships of variables affecting contact (e.g., respondents' at-home patterns) and cooperation (e.g., as householders' psychological predispositions). Then they derive and test hypotheses with data from six national household surveys. The results of their analyses form the basis of recommendations for survey design procedures to reduce nonresponse.

The Groves-Couper model assumes that householders and selected respondents are able to move, see, hear, attend, understand, and talk within the activity performance range expected by survey designers, and it relies on judgements of interviewers to determine if the respondents are able to provide an interview. It does not, therefore, include the large and growing proportions of householders and respondents who are limited in their performance of the basic human activities assumed in interview situations. This omission reflects the survey research field, in which disability has received little attention. The work reported here re-conceptualizes the Groves-Couper model for householders and respondents with disabilities. Based on this revised conceptual framework, hypotheses are derived and tested using data from the U.S. National Health Interview Survey (NHIS).

From 1994 to 1997 the NHIS did a special study of disability. During the home interview for the basic NHIS questionnaire in 1994 and 1995, another questionnaire asked about disabilities of all sample persons. Follow-up home interviews in 1994-97 obtained additional information about all persons with a disability. The follow-up data classify interview outcomes as nonresponse, proxy response, self response with assistance, and self response without assistance. Using these data, the authors are investigating the effects of the type and severity of disability at first interview on outcome of the second interview, with statistical controls for other characteristics related to outcomes.

Preliminary analyses show that adults with disabilities eligible for the follow-up interview are more likely to have characteristics associated with higher rates of contact and cooperation: for instance, they are more likely to be elderly and less likely to be working. This may explain why, despite a self-response rule in the follow-up, the overall nonresponse rate was only slightly higher (6.6%) than it was in the initial interview (5.9%). Preliminary comparisons of types of disability (sensory/communication, mobility, learning disability, diagnosed mental disorders, and mental health symptoms) show the highest nonresponse for other mental health problems (OR=1.7). However, among respondents, persons with learning disabilities had the highest rates of assisted response (OR=1.7) and proxy response (OR=2.3) than other types. Other analyses to be reported will include severity of disability and characteristics of proxies and assistants.

Saturday, 8:30-10:00

Imputation 2

Strategies for Handling Nonresponse in Highly Multivariate Surveys

Thomas R. Belin and Juwon Song, University of California, Los Angeles, USA

It is common for surveys to solicit information on dozens if not hundreds of variables. Although substantial progress has been made in recent years in handling missing items in multivariate survey applications (e.g., Schafer 1997), technical problems arise when the number of variables is large relative to the number of subjects. For example, even with only 20 variables, a multivariate normal model with a general covariance structure includes 210 variance and covariance parameters, which is apt to be excessive for modest-sized surveys.

The simplest and most common approach for dealing with large numbers of variables amidst modest numbers of cases is to parse the problem into a series of smaller problems, each with fewer variables. We first discuss issues surrounding more and less sensible ways to parse a data set into smaller pieces by considering the implied conditional independence assumptions being made.

Because there is inevitably an ad hoc component in any effort to select smaller sets of variables, we consider other approaches that do not require explicit variable selection by the investigator. Schafer (1997) offers one strategy for dealing with the possibility of poorly estimated parameters in a highly multivariate setting, namely to introduce a ridge prior that smooths the estimated covariance matrix toward a covariance structure consistent with independence across variables. Belin and Song (1999) offer an alternative approach that draws on factor analysis ideas to reduce the dimension of the problem. That is, the data might be viewed as being consistent with a smaller number of uncorrelated factors comprised of variables that are correlated within factors. This presentation will contrast the merits of alternative approaches and will illustrate their properties both in simulation studies and in applied surveys with incomplete data.

Imputing Multivariate Nonresponse to Preserve a Given Correlation Matrix

Nanak Chand and Charles H. Alexander, U.S. Bureau of the Census

This paper develops a method for imputing missing income for multiple-source consumer units (CU) in the Consumer Expenditure Survey (CE), the imputation being designed to reproduce a given correlation matrix of the various sources of income. The CE collects information on expenditures and income from persons fourteen years and older in a CU. A CU is a family or a set of unrelated individuals who live together or share expenses. Each person in a CU may have one or more components of income such

as wage or salary, self-employment, and social security.

The general method of the paper is applicable to other surveys involving imputation of multivariate structures where one of the objectives is to preserve the joint relationships in the analyses which use units with imputed values in addition to the completely observed cases. The imputed values take into account the variability of the observed values of the income variables conditional on other variables. The procedure involves generating random variables with replacement from appropriate conditional distributions, and the imputation for a source of income is based on the same random variables as for any other source.

One important use of income variables is to study the correlation of income with expenditures and other variables, thus the process involves imputing missing value of income for any person with missing amounts so that the imputed values within any CU are consistent with one another and with the characteristics of the person or the CU. The problem is complicated because a given CU may have several earners and each of them may have one more types of income.

The imputation process consists of two parts, modeling the relationship of income with other variables and using the models to produce imputed values. In contrast to the unmanageable procedures resulting from the number of variables which could possibly give information about the likely amounts from a particular type of income for a CU taking into account the reasonable interaction effects and nonlinear relationships, this paper proposes a simple model for general patterns of missing data for varying CU sizes and number of sources. While the imputation process does not bring information from outside the underlying data set, it makes it possible to use observed information about all cases, even those with missing income. The proposed model takes into account variables related both to the earner and the CU.

Production of multiple imputed values to meet our objectives is based on a rectangular decomposition of the underlying correlation matrix. This provides a general and consistent imputation process resulting in individual imputed values which are less subject to extreme fluctuations. In addition to the main procedure of the paper which meets the stated objectives for all sample sizes, we also present an asymptotically equivalent method satisfying the requirements only for large samples. In conclusion, this research results in a simple and practical method for producing complete data sets meeting the desired criteria in a complex environment of nonresponse involving multiple persons, different components of income, and varying CU sizes.

Flexible Multiple Imputation by Chained Equations of the AVO-95 Survey

Karin Oudshoorn, Stef van Buuren, Jan van Rijckevorsel, TNO Prevention and Health, Department of Statistics, the Netherlands

The National Services and Amenities Utilization Survey (in Dutch: Aanvullend Voorzieningen Onderzoek, abbreviated by AVO-95), conducted by the Dutch Social and Cultural Planning Office in 1995, includes 6421 households with information about housing, education and services (Spit [1996]). The aim of the survey is to get insight in the use of a large number of social and cultural services by the Dutch population. The non-response on these items is substantial. For example, from 49.7% of the 2792 households with a mortgage the monthly mortgage interest is missing. From 3427 of the households that own a house, 11.1% did not report the current selling price of the house.

To impute a part of the AVO-95 survey we use a new approach of (multiple) imputation (Rubin [1987]) where for each incomplete variable, a separate imputation model by means of set of predictor variables is used (van Buuren et al [1999], Brand [1999]). The algorithm, which is based on Gibbs sampling, imputes each incomplete column in an iterative fashion.

The implemented method differs from other available methods (e.g. Schafer [1997]) in that the conditional models can be specified directly, thus without the need to choose an encompassing multivariate model for the entire dataset. Instead, for each variable one chooses a (possible different) set of predictor variables and a univariate imputation method (e.g. normal linear regression, logistic regression, polytomous regression). The method is flexible and very well suited for large datasets, like the AVO-95 survey, with many variables (in total more than 50 predictor- and imputed variables). The algorithm is implemented in S-Plus and named MICE, which stands for "Multivariate Imputation by Chained Equations".

We will focus on the process of deriving the imputation models for the variables mortgage, the current selling price of the house and some predictor variables with missing values also. This phase appears to be the most critical and complex phase in Multiple Imputation. Next, we describe the methodology used and practical issues encountered, to obtain the imputations. Furthermore, we give some characteristics of the obtained imputed values. Finally, we will examine the added value of the multiple imputed dataset with respect to listwise deletion and the, with an hot-deck method, imputed dataset as used now by the Social and Cultural Planning Office.

Saturday, 8:30-10:00

Panel: Tradeoffs in Evaluating the Impacts of Nonresponse

Saturday, 10:30-12:00

Unit And Item Nonresponse In Business Surveys

*Diane Willimack and Elizabeth Nichols, U.S. Bureau of the Census
Seymour Sudman, University of Illinois, USA*

As with household surveys, the presence of unit and item nonresponse presents a major problem in business surveys and censuses. The problem may be even more severe for some business surveys that have item nonresponse rates much higher than those found in household surveys. The problem has serious consequences for the bias or variance of the statistics being estimated, such as revenues or value of production, particularly if the adjustments for missing data account for a large proportion of the estimates.

In this paper, we review current literature on unit and item nonresponse in business surveys. Even though business unit and item nonresponse are quantifiable, the literature is sparse in pinpointing the causes of nonresponse. In addition, we present qualitative data from a series of in-depth personal interviews with data providers from thirty large companies. These data shed light on potential and unlikely sources of unit and item nonresponse for large businesses. We synthesize these data with findings from existing research on nonresponse in business surveys. Comparing this analysis with models of household nonresponse, we develop a model of factors which affect unit and item nonresponse in business surveys.

Saturday, 10:30-12:00

Modeling Non-Ignorable Attrition and Measurement Error in Transportation Panel Surveys

*David Brownstone and Thomas F. Golob, University of California, Irvine, USA
Camilla Kazimi, San Diego State University, USA*

Modern panel surveys frequently suffer from high and non-ignorable attrition, and transportation surveys suffer from poor travel time estimates. The initial sampling process for most transportation surveys is also non-ignorable since rare travel modes are oversampled (and mode choice is the key dependent variable). This paper examines new Multiple Imputation methods for adjusting forecasts and model estimates to account for these problems in a new panel survey of 1500 commuters in San Diego, California. These data are collected to evaluate charging solo commuters to use an existing 8-mile underutilized freeway carpool lane. We illustrate the impact of attrition and measurement error on a standard conditional logit model of commuters' mode choice (solo drive in free lanes, pay to solo drive in the carpool lanes, or carpool for free in carpool lanes). Although the attrition rate between waves is 40% and non-ignorable, the quantitative impact on the results is negligible. However, measurement error in travel time does have an important impact on the key results from our model. Finally, failure to account for the measurement error process using Multiple Imputations yields a downward bias of at least 50% in the standard errors of the logit coefficient estimates.

Saturday, 10:30-12:00

Nonresponse and Measurement Errors as Sources of Bias

Analysis of the Expressions of the Nonresponse Bias for Some Well-Known Estimators

Sixten Lundstrom, Statistics Sweden

Lundström and Särndal (1999) suggests in a calibration approach a method for constructing weights that reduce the sampling error and the nonresponse bias. They also derive a general expression for the nonresponse bias of this estimator and show that the bias is zero in three situations, namely (i) when the response probabilities are equal for each element, (ii) when the population residuals are all zero or (iii) when the inverse of the response probabilities are linear functions of the auxiliary vector. These findings can be valuable when selecting auxiliary variables. Lundström and Särndal (1999) also show that, due to the specification of the auxiliary information, most "conventional" estimators are obtained as special cases. In this paper the explicit bias expressions for the expansion estimator, the poststratified estimator, the ratio and the classical regression estimator, as well as separate versions of the latter two, are analyzed. The purpose is to increase the knowledge on how to select auxiliary information in order to reduce the nonresponse bias. By using artificially generated data sets representing various populations and nonresponse mechanisms the nonresponse bias for these estimators are studied.

Treating Missing at Random Data in Covariance Structure Modeling

Gino Verleye, University Gent, Belgium

This performance (Ph.D.) study analyses the performance of 5 missing data solutions to handle MCAR and MAR data problems in structural equation modeling (SEM) : listwise and pairwise deletion, EM estimation of the covariance matrix, proper multiple imputation and direct parameter estimation.

This factorial design uses 5 factors :

- * the type of SEM model : a confirmatory factor analysis model versus a full model
- * normal versus skewed data,
- * 3 levels of missing data (5%, 15% and 25%),
- * 2 types of missing data : MAR and MCAR,
- * the five missing data strategies.

The evaluation is done by means of multiple criteria (both numerical and graphical) and useful suggestions are given for the researcher with MAR or MCAR (item non-response typing) missing data problems.

This analysis, compares methods that, to this moment, have not been compared. The idea to apply multiple imputation with the multiple group approach in SEM is also a new feature.

A Comparison of Nonresponse Bias and Measurement Error Bias for Health Characteristics Collected by Two Modes of Interview

Paul P. Biemer, Research Triangle Institute, USA

This paper reports on a recently completed study conducted by the authors which compared the biases due to nonresponse and measurement error for two modes of interview: CATI and face to face PAPI. The study was conducted using data from the 1994

National Health Interview Survey (NHIS) and an RDD survey, conducted in parallel with the NHIS using the NHIS questionnaire. The NHIS face to face interviews were collected by the Bureau of the Census while the RDD CATI interviews were collected by RTI. Over 10,000 interviews were conducted as part of the evaluation whose purpose was to estimate and compare both nonresponse and measurement error biases for the face to face and CATI modes of data collection.

The evaluation was conducted in two states - Texas and California. In each state, interviews were conducted with two random sample of households: one sample received the face to face mode and the other received the CATI mode. The CATI samples were selected by RDD methods while the face to face samples were the usual NHIS sample in Texas and California. In addition, a subsample of the respondents in each mode treatment was reinterviewed to obtain replicate measurements of the key data items in the original interview. Finally, nonrespondents from the NHIS face to face sample were contacted for interviews in order to obtain data on nonrespondents for the nonresponse analysis.

Estimates of the classification error rates for the telephone and face to face interview modes were obtained using latent class analysis. The key assumptions in the measurement bias analysis are the following:

- a. classification error rates for the CATI mode in Texas and California are equal,
- b. classification error rates for the reinterviews are equal to those for the RDD survey
- c. classification error rates for face to face interviewing differ from those for telephone interviewing.

Note assumptions (a) and (b) are plausible since both the original interviews and the reinterviews for Texas and California were conducted by the same RTI CATI interviewers. Under these assumptions, the parameters of the misclassification error model are estimatable using maximum likelihood estimation. For most of the items in the analysis, the data were sufficient to test assumptions (a) - (c) as well as the fit of the final models.

Using the estimates of the classification error from the latent class analysis, estimates of the measurement error bias were obtained. Further, differences between true characteristics of respondents and nonrespondents were estimated using data from respondents and nonrespondents adjusted for measurement bias. In this manner, estimates of the nonresponse and measurement biases for the NHIS were obtained. These bias estimates were then subtracted from the NHIS estimates in order to obtain estimates for the target population adjusted for nonresponse and measurement error biases.

Similarly, the classification error estimates obtained for the telephone mode were used to adjust the RDD/CATI estimates for measurement bias. Then, by subtracting the unbiased estimate of the telephone population obtained from the NHIS analysis from the measurement bias corrected RDD/CATI estimates, an estimate of the RDD/CATI nonresponse bias was obtained. In this manner, we obtained reliable estimates of the nonresponse and measurement error biases for RDD/CATI and NHIS estimates for a wide range of health related variables.

A few of the findings of from this evaluation include the following:

- * Response rates for the RDD survey for California and Texas were substantially below those for the NHIS, as expected. In California, the response rate was 57 percent for RDD and 82 percent for the NHIS HP2000S. In Texas, the response rate was 64 percent for RDD and 80 percent for the NHIS HP2000S. Overall, the response rate was 60 percent for RDD and 81 percent for the NHIS HP2000S.
- * For most of the 19 variables in the analysis data set, the differences between RDD and NHIS estimates were within sampling error. However, for a few important variables (viz. Number of Smoke Detectors, Personal Opinion of Health, questions on cigarette smoking, and questions about doctors visits) significant differences were observed.
- * Poor response reliability is a problem for a number of variables in the analysis data set. In general, reliability tends to be better for the RDD survey than for NHIS. However, it is difficult to predict which mode will be more reliable for specific questions and question items.
- * For 23 of 41 data items, the total bias (accounting for nonresponse and mode) in estimates of means, totals, and proportions is greater for the NHIS than for the RDD survey.
- * For both surveys, measurement error bias is usually the larger source of bias. In the latent class analysis, significant measurement error effects were found for 10 of the 14 variables in the analysis data set. It appears from these data that the NHIS is subject to greater measurement bias than the RDD survey.
- * As expected, nonresponse bias tends to be greater for the RDD survey than for the NHIS, but not as much as the differences in the response rates would suggest. For more than 40 percent of the items, the NHIS nonresponse bias exceeded the RDD nonresponse bias.

In addition to the findings in this paper regarding the relative magnitudes of nonresponse and measurement biases for health characteristics, the paper will also provide the technical details of the latent class analysis used to produce the bias estimates.

Saturday, 10:30-12:00

Interviewer Effects 3

Anatomy of Nonresponse

Charlie Palit, Neli Esipova, Nancy Davenport, and Amy Fusek, University of Wisconsin, USA

The refusal event and its nature as perceived by the interviewer were recorded. The authors use this information plus information on where in the interview path the event occurred to classify and code these events. This classification scheme is used to examine the refusal conversion performance of Interviewers as a function of the original refusal category. Both interviewer characteristics and by the nature of the previous refusal event are used as independent variables in this analysis, in short male/female interviewer/informant interactions are studied to determine their influence in the refusal conversion process. This exercise produces some interesting results.

Further, the authors hypothesize that looking at refusals by such operational categories will facilitate the development of effective refusal avoidance and refusal conversion strategies. As a first step they attempt to use this tool to reduce first time refusals by preparing special scripts for the interviewers to use when a refusal event appears to be imminent. These scripts are tailored to the category of refusal events expected and programmed into the CATI questionnaire. Multiple scripts are provided for most categories of refusal events.

Interviewer Attitudes about Privacy and Confidentiality

Thomas S. Mayer, U.S. Bureau of the Census

Although the U.S. Bureau of the Census has an admirable record for maintaining confidentiality of census and survey data, investigations have found that some people still hold a negative perception of the Census Bureau regarding issues of privacy and confidentiality (e.g., Bush, 1986; Dillman and Reynolds, 1990; Fay, Bates, and Moore, 1991; Ira O. Glick & Associates, Inc., 1978; Kerwin and Edwards, 1996; McDonald, 1985; Moore, 1982; Singer, Mathiowetz, and Couper, 1993). These studies have found that people often feel that questions asked by the Census Bureau are an invasion of their privacy, that they don't trust the Census Bureau to keep data confidential, and thus do not respond to Census Bureau surveys.

Perhaps equally disturbing is the indication that some Census Bureau field staff do not have a clear understanding of the Census Bureau's confidentiality policies and sometimes do not believe the confidentiality statements themselves (e.g., Baca, 1983; Groves and Couper, 1998; Lavin, 1989; Rothwell, 1969). Singer and Kohnke-Aguirre (1979) reported that interviewers who expect sensitive questions to be difficult to ask obtain higher nonresponse for those items. It follows logically then, that interviewers' attitudes of disbelief regarding confidentiality may render them less effective as converters of respondents' attitudes of disbelief, and thus lead to unit nonresponse. In addition, attitudes held by field staff regarding an item's potential to invade a respondent's privacy may affect data quality and lead to item nonresponse.

This research project seeks to understand current (interviewers') attitudes regarding privacy and confidentiality of census data, how these attitudes compare to those of previous assessments (e.g., Lavin, 1989), and if there are any noticeable trends in these attitudes. To accomplish this goal a questionnaire will be administered to a large sample of field staff. Questions will include those from previous questionnaires as well as new items exploring issues such as the sensitivity of questions, the invasion of privacy, the knowledge of legal issues concerning Census Bureau surveys, ways in which confidentiality can be breached, interviewers' perceptions of public attitudes, reasons for nonresponse, suggestions for improvements, etc. Furthermore, the project begins to explore the relationship between Field Representatives' attitudes regarding privacy/confidentiality issues and their individual response rates. Of particular interest is whether attitudes of disbelief held by the field staff regarding Census Bureau confidentiality policies and/or the lack of understanding of these policies affects the attitudes and beliefs of the respondents and therefore contributes to higher levels of nonresponse. Results indicating that negative interviewer attitudes lead to nonresponse would merit a follow-up examination of the current Census Bureau privacy/confidentiality training practices. This investigation could lead to the proposal of improved training procedures as a possible method for reducing nonresponse.

How the Initial Contact Can be Determinant for the Final Response Rate in Face to Face Surveys

Ann Carton and Geert Loosveldt, KULeuven, Belgium

Recently, from a theoretical viewpoint on the one hand combined with a search for empirical evidence on the other hand, there has been done a lot of work in exploring the two steps procedure of survey participation: respondents must be contacted and once contacted they must be willing to cooperate. Apart from influential factors associated with the social environment (e.g. survey taking climate), with the household (e.g. socio-demographics, knowledge of the topic, survey experience), with the survey design (e.g. length of the survey, topic) and with the interviewers (e.g. experience, expectations), some social scientists (among others Campanelli et al., 1997; Groves and Couper, 1998; Morton-Williams, 1993) are more and more convinced that the crucial factor of survey participation lies in the initial (and further) contact(s) between the interviewer and the respondent. In fact, the aforementioned determinators come together during the doorstep interaction between the different actors of the face to face interview.

Research from four large scale face to face surveys in Belgium (Health Survey of 1994, N=1,631 respondents; National Election Study of 1995, N=2,099; Survey about the knowledge of the federal structures in Belgium: zero-assessment in 1995, N=710 and effectiveness in 1996, N=987) makes clear that the first three contacts are determinant for the final response rate. On average, about 85% of the total amount of completed interviews is realised during one of the first three contacts (1st contact: 25%, 2nd contact: 40% and 3rd contact: 20%). Therefore, the calling strategy which the interviewers use during the initial (and subsequent) contact should be as efficient and effective as possible. It's clear that a productive calling strategy can act as a motivating factor during the fieldwork. A calling strategy is efficient and effective when during the initial contact (1) an interview can be conducted or (2) when an appointment can be made which during the second contact results in a completed interview. So, the strategy used at a subsequent contact is conditional on the information of the previous contact(s).

The objective of the analysis of this paper is twofold. Firstly, to get insight in the determinant factors of an efficient and effective calling strategy. Apart from controlling for known relevant factors such as (non)respondents' age, gender and neighbourhood, information available from the contact description forms will be introduced into the different models. For each of the surveys the interviewers had to complete a standard form for all their respondents and nonrespondents. The form contains data about day and time of each contact, whether the contact was in person or by telephone, whether the contacted person was the respondent or

another member of the household and the outcome of the contact (e.g. completed interview, appointment, refusal, not at home). Secondly, to detect whether some interviewers use a more efficient and effective calling strategy than others. To analyse this kind of interviewer effect multilevel regression analysis will be used. The differences between the interviewers will be explored by introducing several interviewer characteristics into the models (e.g. experience, expectations).

Saturday, 10:30-12:00

National Experiences on Nonresponse 1

Separating out the Stages of the Survey Participation Process

Johanna Laiho and Peter Lynn, National Centre for Social Research, UK

Analysis of the factors associated with survey non-response typically treats the outcome of the attempts to gain a response as a unidimensional polytomous variable. The outcome is commonly described using a single binomial or multinomial regression model. In recent years, increasing attention has been paid to separating out the contact and co-operation stages of the response process. Theories have been developed to describe each of these two processes and survey outcome data has been used to validate the theory. However, the complex temporal and conditional nature of the survey participation process has not been explicitly taken into account. In this article, we recommend explicit identification of the necessary stages of the process, and sequential modelling of each stage, thereby acknowledging that each stage is conditional upon the outcome of the previous one. We posit that this approach should be capable of yielding information about the participation process that is not apparent when the conditionality is not taken into account. We illustrate these ideas with data from the British Crime Survey and present the results of fitting models using a range of auxiliary variables.

Nonresponse in Institutional Surveys to Estimate Statewide Annual Incidents of Homelessness

Mack C. Shelley, II, Paula W. Dail, and Scott T. Fitzgerald, Iowa State University, USA

Disagreement exists regarding the optimal methodology for estimating the incidence of homelessness--one or more fixed points in time or annual and/or continuous counting--and the relative virtues of capture-recapture (i.e., count-recount) methods versus sampling in space and time, the use of key informants (key person surveys), or random sample surveys. Methods used to estimate the incidence of homelessness span partial counts, extrapolations from partial counts, windshield street surveys, and area probability designs. The fluid and episodic movement of persons in and out of the condition of homelessness creates additional difficulties for measuring homelessness. Agencies, reporting techniques, and record-keeping abilities change over time.

Homelessness may be underreported by some agencies and overreported by others. Using an operational definition of homelessness provided in Section 103 of the McKinney Homeless Assistance Act (1987), supplemented by U.S. Department of Education 1989 guidelines regarding homeless children, we estimate the number of homeless individuals and the incidence of homelessness in the state of Iowa.

Questionnaires were distributed to all public schools in Iowa and to all known shelters, community action program agencies (CAPs), county general relief offices, transitional housing programs, and Department of Human Services (DHS) offices, as well as to miscellaneous programs such as medical outreach services. Response rates were 53.8% overall, 55.2% for schools, 57.3% for shelters, 34.7% for general relief agencies, 70.2% for DHS offices, 3.7% for CAPs, 18.8% for transitional housing programs, and 66.7% for miscellaneous other agencies.

Low, midrange, and high unduplicated estimates were constructed based on assumptions regarding the probability of duplication of each record. These values were 4,828, 4,983, and 5,291, respectively. Approximate 95% confidence intervals were derived for the estimated unduplicated reported numbers of homeless, assuming independence among the schools, shelters, and other agencies strata.

Reciprocals of response rates provided initial adjustments of institutional nonreport. For shelters, these adjustments were refined further by using shelter-bed capacity information. To estimate the incidents (episodes) of homelessness, the nonshelter agency data were inflated for time. Data provided by shelters and other agencies covered a one-month period, while schools provided annual data. An annualized estimate of the incidents of homelessness was constructed using a time-inflation factor of 12 for the data from shelters and from other agencies. Time-adjusted estimates of the statewide incidents of homelessness were 38,950 (low), 59,558 (midrange), and 83,502 (high). Approximate 95% confidence intervals are constructed for the adjusted estimate of the number of incidents of homelessness.

Local-area estimates of the incidents of homelessness were constructed for large metro, small metro, and rural county strata, by allocating the statewide total proportionately first across the three county-types and then among counties within each stratum. These calculations were conducted separately for two different definitional categories of homelessness: sheltered and unsheltered homeless, and doubled-up and transitional housing/other.

Determinants of Underreporting of Induced Abortions in Japan

Hiroshi Kojima, Nat Inst of Population & Social Security Research, Japan

This is an attempt to clarify the determinants of underreporting of induced abortions in Japan, based on the analysis of two successive cross-sectional national fertility surveys conducted in 1987 and 1992 by the Institute of Population Problems (currently, the National Institute of Population and Social Security Research) in Tokyo. The incidence of induced abortions is on the average about twice as high in the data from the 1987 survey than in the data from the 1992 survey. It is true even when the same marriage cohorts are compared in the data from two surveys although the data from the 1992 survey should reveal higher level due to the accumulation thorough time. The underreporting in the 1992 survey seems to be facilitated by the simpler questionnaire for pregnancy history.

A preliminary analysis has been conducted by comparing the results from OLS regressions for determinants of the number of induced abortions, applying almost the same model to the two data sets. The analysis has been limited to first-married wives living with first-married husbands. It shows that the factors encouraging the underreporting of induced abortions include arranged marriage, the wife's higher or lower age at marriage (below 20 and 25 or higher), rural residence, residence in Kyushu District, the wife's part-time employment, the wife's junior or technical college education, the wife being the eldest daughter without brothers, the husband being the only child or from three-child family.

A more elaborate analyses will be attempted including the multinomial logit analysis by pregnancy order. But other methods might as well be explored including the conditional imputation based on the statistical matching of women in the same marriage cohort from the two data sets.

Saturday, 1:30-3:00

The Effects of Repeated Call-Backs and Reallocation on NonResponse Bias and Survey Accuracy

Peter Lynn and Paul Clarke, National Centre for Social Research, UK
Jean Martin, Office for National Statistics, UK
Patrick Sturgis, London School of Economics, UK

This paper is concerned with the effects and effectiveness of attempts to make contact with, and gain the co-operation of, sample members on face-to-face household interview surveys. The focus is particularly on the use of repeated callbacks beyond a standard minimum and reallocation ("extended interviewer attempts") as tools to maximise contact and co-operation rates.

We discuss the impacts that these extended interviewer attempts have upon survey nonresponse bias. We also describe how these impacts can be related to the costs of different elements of extended interviewer attempts in order to assess the relative cost effectiveness of different field strategies.

Specifically, we argue that for a meaningful analysis of response elicitation effort, the contact process and co-operation process must be separately identified, but considered simultaneously as they cannot be assumed to be independent. (Virtually all previous studies have either concentrated on just one of these two processes, or have combined proxy measures of both into a single measure of "response resistance". We extend this and look at interactions and relationships between the two.) Consequently, we develop measures both of the effort needed to make contact and of the effort needed to achieve co-operation once contact is made. We then model both measures using a set of auxiliary variables as covariates. The modelling is guided by the theory presented by Groves and Couper (1998). We aim to demonstrate that this approach provides more information than previous approaches and may have practical implications for field procedures and/or adjustment techniques.

We use data from three major UK household surveys: The Family Resources Survey (FRS), The Health Survey for England (HSE) and The British Social Attitudes Survey (BSAS). These surveys were deliberately chosen to include a range of subject matters, response rates, and levels of conversion effort.

Saturday, 1:30-3:00

Using Matched Substitutes to Adjust for Nonignorable Nonresponse through Multiple Imputation

Donald Rubin, Harvard University, USA
Elaine Zanutto, University of Pennsylvania, USA

Nonignorable nonresponse can be a potentially serious problem in surveys. For example, it can arise when there are hidden clusters in the population that share a common, unobserved, characteristic that is related to both the outcome and the response probability of each member of the cluster. In this case, standard nonresponse adjustment methods fail to correct the bias due to nonresponse. To address this problem, we propose to impute the missing data using responses from substitute survey units who are contacted when originally selected units do not respond. To improve on existing methods that use substitutes' values to replace nonrespondents' missing values, we use substitutes, respondents, and covariate values that are available from nonrespondents to build a model to multiply-impute the missing data. The three key elements of this strategy are (1) the matched substitutes, which provide information about respondents who are similar to the nonrespondents with respect to "field" covariates; (2) the multiple imputation model, which adjusts for observable differences between nonrespondents and their substitutes on "modeling" covariates; and (3) the multiple imputations themselves, which facilitate point and variance estimation, even for complicated estimands and complex surveys. The proposed method can produce estimates with substantially less bias than those obtained with either the traditional use of substitutes or other methods of nonresponse adjustment that do not use substitutes (e.g., weighting class adjustment or standard multiple imputation methods that are based on respondent data and nonrespondent covariates).

Saturday, 1:30-3:00

Imputation Algorithms

Categorical Constraints Guided Imputation for Nonresponse in Surveys

Eric Rancourt and Tzen-Ping Liu, Statistics Canada

In surveys and censuses, it is recognized that nonresponse is present even with the best prevention programs. There are various techniques to deal with nonresponse, but in this paper we are interested in imputation (for item nonresponse). Among imputation

techniques, a wide range of methods exist for situations where the characteristics imputed are numerical or continuous, but there are fewer methods when the variables imputed are categorical. Yet, imputation of categorical variables may have a greater impact on final estimates, especially when they are in turn used to create imputation classes for continuous variables or when they form domains of interest.

Unless specifically taken into account, imputation does not usually preserve the distribution of categorical variables and particularly if nonresponse is not missing completely at random. For instance, if hot-deck imputation techniques are used, then one can expect to have the distributions preserved, but it will not necessarily be the case for a given sample. To solve this, ratio adjustments (sometimes called pro-rating) is often applied to the data set after imputation, thus satisfying additivity constraints or benchmarking totals. However, this approach fails to work for categories that remain empty after imputation and it may over adjust some imputed data. There are also more sophisticated methods to model the probability of being into a given category, but even with those approaches, the constraints may not be satisfied by the data after imputation.

In many surveys such as some of those carried out at Statistics Canada, there is a need of adjusting the data after imputation. In this paper, we describe an iterative imputation algorithm, which performs categorical imputation and simultaneously calibrates the data set according to auxiliary information available for all units in the sample. The method is based on a constrained selection of records to insert into categories, thereby imputing the categorical variable(s). To insure that all nonrespondents are imputed, a forcing procedure also takes place after the insert step.

The approach also applies to situations where an already-imputed data file does not satisfy specified constraints. In this case, categories having large differences in their distributions between auxiliary totals and imputed totals are identified. Then, records are exchanged between categories, using a search algorithm to satisfy the constraints imposed by the auxiliary totals. The resulting data file after imputation is balanced. That is, has an architecture, which preserves the categorical associations (distribution) of the auxiliary variables used in the process.

The algorithm will be presented and described along with an example.

A Generic Implementation of the New Imputation Methodology

Michael Bankier, Martin Lachance, and Paul Poirier, Statistics Canada

A New Imputation Methodology (NIM) was introduced in the 1996 Canadian Census to carry out the hot deck Edit and Imputation (E&I) for the demographic variables. For the first time, qualitative and quantitative variables could be imputed simultaneously. The NIM used a data-driven approach with single-donor imputation to determine the best imputation action. For the 2001 Canadian Census, the NIM will be extended to more variables in accordance with the long-term objective to progressively move all census variables to the NIM for the 2006 Canadian Census. This expansion of the NIM use provides the opportunity to make a generic implementation of the methodology. The generic implementation will have new features under consideration, such as the evaluation of many quantitative variables in a single edit rule, the use of derived variables and also, the capability of treating non-linear edit rules. The flexibility of the generic implementation will allow using the NIM in a wider variety of surveys.

Theory and Methods for Longitudinal Imputation

Z. Patak,, P. Lavallée, and B. Lapierre, Statistics Canada

Longitudinal imputation (i.e., imputation for longitudinal surveys) presents some difficulties that are not present in cross-sectional imputation. With the latter, the focus is on assigning values to the missing variables to obtain reliable point estimates. Cross-sectional imputation does not usually place too much emphasis on the imputed values, but more on their effect on cross-sectional estimates. With longitudinal imputation, imputed values are important because they are linked to the continuity of the longitudinal units through time.

Imputing by only considering data from the current wave could possibly lead to a large number of unrealistic changes through time. On the other hand, only carrying forward historical data may lead to underestimation of change through time. It is therefore recognised that longitudinal imputation needs to consider the past, the present and the future data (if available) to come up with the best-imputed values.

In the paper we will limit ourselves to handling the case of longitudinal item imputation where the missing values are missing completely at random. We will identify several methods found in literature and evaluate them using a Monte Carlo simulation. Both continuous and categorical data will be used to assess the performance of the selected estimators. We shall focus our attention on methods that make use of auxiliary information from the past, the present, and the future waves (if available).

All methods will be evaluated under the assumption of ignorable non-response using data from a Statistics Canada longitudinal survey, SLID (Survey of Labour and Income Dynamics).

Saturday, 1:30-3:00

Experiments in Mail Surveys

Stemming the Tide of Declining Response Rates Among Members of a Dues-paying Organization: A Case Study

Jeanette O. Janota, Herbert M. Baum, and Sarah C. Slater, American Speech-Language-Hearing Association, USA

Low response rates are costly to researchers in many ways, and much research has been conducted to identify practical ways to increase them. Surveying special populations such as members of professional organizations offers special challenges. On the one hand, members might reasonably be expected to be motivated to respond to questionnaires on topics related to that organization's activities, priorities, and allocation of resources. Alternatively, some may object to having their dues spent on incentives, a

technique successful in increasing response rates.

There are few published results from member surveys, particularly ones that provide methodological details. This paper will report on response rates over time to a survey mailed to members of a non-profit, professional organization with nearly 100,000 members and recent efforts to reverse declining response rates.

The Omnibus Survey has been conducted regularly by the American Speech-Language-Hearing Association (ASHA) since 1982. Since 1991, response rates have declined by about 2% per year to a current low of 56%. Two methodological experiments were built into the 1997 survey to measure impact on response rates: one on outgoing postage and one on incentives.

Half received surveys with metered first-class postage and half with first-class stamps. Slightly more surveys were returned by individuals who received the metered envelopes (50.4%) than stamps (49.6%), but the difference was not statistically significant.

The second experiment involved an incentive: respondents would be entered in a drawing for one free membership (\$174 value). The incentive was not sufficient to reverse the declining response rates, and there was no cost savings in reduced follow-up mailings because the speed with which surveys were returned did not increase.

A follow-up survey was conducted from a sample of 1,000 nonrespondents asking how many surveys they had received during the previous year, why they had not completed the Omnibus Survey, and what would motivate them to respond to future ASHA surveys. A 15.4% response rate was achieved from this postcard survey which included no identification numbers and cost \$5.42 per return.

One of the methods shown to increase response rates is the use of incentives. However, using an incentive with any monetary value is problematic when the sample is selected from an organization whose members pay dues. The choice of incentives requires great creativity for such populations. The use of a sweepstakes drawing for a year's membership fees did not increase the response rate from previous years.

The 1999 Omnibus Survey introduced several new measures to increase response rates. Split ballot experiments included mailings (comparing prenotification with receipt of a third survey form) and incentives (a specially designed lapel pin vs. a checklist of reports from the 1999 data that informants could elect to receive as they become available). It is hoped that these incentives will be judged by members to be valuable--and cost effective.

Effect on Response Rate of Handwritten Versus Facsimile Signatures in Mail Survey Cover Letters

Jeannine James, American Research Company, George Mason University, USA
Richard Bolstein, George Mason University, USA

Prior studies testing the effect of hand-signed cover letters on response rate have yielded inconsistent results, possibly because those studies simultaneously tested other personalization techniques.

The present study tests the sole effect of a handwritten signature versus facsimile signature in an otherwise personalized two-wave mailing, that is, with an outer envelope individually typed and a cover letter containing the recipients name, address, and personal salutation. The results indicate that a hand-signed cover letter may be effective in increasing response rate among recipients who are predisposed to respond anyway, since the rate increased significantly in this survey among recipients who had responded to a previous survey by the same sponsor. However, the hand-signed cover letter had no effect on nonrespondents to a previous survey, indicating it may not be effective among hard-core nonrespondents.

A cost-benefit analysis suggests that hand-signed cover letters may be a cost effective way of increasing response rate among those predisposed to respond, especially in comparison to the use of monetary incentives.

Empirical Contributions to a General Survey Response Model

Gustav Haraldsen, Heidi Kristin Reppen, and Eiliv Lund, Statistics Norway

The overall purpose of the paper is to better specify how questionnaire design and data collection procedures contribute to the response rate, and how a low vs high response rate affect the response quality of different questions.

The analysis will be based on a theoretical model which says that the response to a questionnaire is dependent on a) the interest we are able to trigger in the respondent and b) how well the respondent feel that she is competent to answer the questions. We believe that the interest-factor is mainly affected by a user-friendly survey design and incentives, whereas the competence-factor is reckoned to be more dependent on how well the questions are phrased. This model will be explained in fuller detail in the paper. One of the theses we will try to illustrate empirically in our paper, is that even if a low response rate may result in a disproportionate coverage of different population groups, the response quality of questions with a high competence threshold may not gain very much from a higher response rate.

The data for our analysis is drawn from a huge postal survey on cancer related behaviour among women aged 30 to 69 years old. The survey was carried out by Statistics Norway for the Institute of Community Medicine at the University of Tromsø in the period 1991-1997. More than 180 000 women were randomly selected to participate in the survey. The questionnaire included questions on social status, previous diseases, reproductive factors, cigarette smoking, contraception use, dietary habits, alcohol consumption and sun bathing habits.

The sample was drawn from the Norwegian Central Person Registry. In the registry, every person is identified with a unique number. This number can also be linked to several other registers. In addition to standard demographic variables, it is therefore possible to compare distributions in the gross and net sample according to variables that are more relevant to the subject of the questionnaire. Examples of such variables are the woman's age by her first child birth and the number of childbirths.

During the data collection period, several experiments with changing presentation and different ways of reminding the participants to answer, were carried out. The main experiments were:

Factors such as the title and length of the questionnaire were manipulated in five versions that were tested among women aged 30 - 49 years old. Each version was tested in a subsample of 1000 women.

Five versions, which differed in length and design, were tested among women aged 50 to 69. In addition, the effect of a question that revealed that the respondents would be contacted a second time, was tested in this age group. Each version was tested in a subsample of 1000 women. Different kinds of reminders and administrative strategies to collect remaining questionnaires were tested in four subsamples, each with 2500 women in the age group 45-69. An experiment with real stamps instead of a pre-printed franking stamp was also carried out.

Saturday, 1:30-3:00

National Experiences on Nonresponse 2

Nonresponse and Data Yield: Experiences from Austria and France

KW Axhausen, ETH Zuerich, Switzerland

In many surveys, for example of crime, illness, travel, purchasing etc, the amount of information requested from the respondent is not known apriori to the survey designer. In situations like this, the design of the survey will aim to assure both the largest possible response, but also a significant data yield, i.e. reports of the survey object, for example episodes of illness, number of victimizations, number of trips undertaken, objects bought. The design process has therefore two main objectives and it not clear that one can achieve both at the same time.

This paper will report the analysis of two sets of experiments carried out in 1996/97 in Austria and France, which aimed to provide insight into these trade-offs. The surveys were concerned with the amount of long-distance travel undertaken by the population. The experiments in Austria involved eight postal surveys, which varied the temporal orientation of the survey, the reporting period and the amount of detail requested about each movement, while leaving the other elements of the survey unchanged. The protocol involved a detailed telephone interview of those not responding in writing, which captured the essential items from the written survey.

The eight experimental surveys in France varied different elements: the sample frame (fresh random sample versus an existing panel of respondents), the data retrieval method (CATI versus postal) and the number of contacts (reporting twelve weeks either in one contact or in three monthly intervals).

The paper will report the disaggregate modelling of the response behaviour and of the data yield, as a function of the experimental variables and of the socio-demographics of the respondents. The results indicate that there are substantial interactions and that certain variables increase response, but lower the data yield.

Methods Used to Address Missing Data Issues in the Design and Analysis of the Belgian Health Interview Survey 1997

Tomasz Burzykowski and Geert Molenberghs, Limburgs Universitair Centrum, Belgium

Jean Tafforeau, Herman Van Oyen, and Stefaan Demarest, Scientific Institute of Public Health, Belgium

Leila Bellamammer, National Institute of Statistics, Belgium

The Belgian Health Interview Survey was conducted in 1997. It was the first such survey in the Belgian history. The target sample was 10,000 individual interviews, spread over the calendar year. A stratified, multi-stage, clustered sampling scheme was used. Primary sampling units were municipalities, secondary: households, tertiary: individuals. Missing data issues received attention starting from the design phase of the survey. To limit the possibility that drop-out of households invited to take part in the survey would prevent achievement of the required sample size, oversampling of households was planned. Systematic trends in the drop-out were addressed by replacing non-interviewed households not in a random fashion, but by households matched on a set of variables related to household size, location etc. To reduce impact of drop-out of individuals on survey results, proxy rules for non-participating individuals were established.

The multi-stage nature of the sampling scheme implied that there were three levels, at which missing data might occur: household level, individual level, and item level. The following taxonomy was used to describe drop-out of the invited households and individuals:

- non-participation (NP): no interview was obtained;
- non-availability (NA): no interview was obtained due to contacting problems;
- non-response (NR): no interview was obtained due to an explicit refusal.

To reach the required sample size of 10,000 individual interviews, 35,023 households were sampled. Among 11,568 households which were invited to participate in the survey, in 6904 (59.7%) cases the attempt to obtain an interview failed (NP): in 3358 (29.0%) cases due to problems with contacting the household (NA) and in 3546 (30.7%) cases due to explicit refusal of the household (NR). A more detailed descriptive analysis of the household-level NP, NA and NR was conducted, using the beta-binomial model (extended to allow for missing data) and logistic regression. It allowed to identify covariates that were associated with a higher probability of a household drop-out.

Individual-level NP was infrequent. Among 10,339 individuals from the households that participated in the survey that were selected for an interview, individual NP was only noted in 785 (7.6%) cases. In consequence, individual-level NP was not the primary objective of missing data modelling.

Clearly, at the item level only NR occurs. Its frequency was also generally low: it did not exceed 11% for any of the items. In the main analysis of the survey, an available-case strategy was used. Effect of ignoring the item-level NR on analysis results was investigated using multiple imputation for selected important variables (continuous, categorical). It was found that, at least in the analyzed examples, the effect was negligible.

The investigation of missing data in the Belgian Health Interview Survey 1997 also underlined importance of collecting information that might allow to study influence of the missing data on survey results.

Nonresponse Error and Emergent Challenges - Census Enumeration in Caribbean Micro-States

Godfrey Bernard, University of the West Indies

Non-response is known to be one component of non-sampling error that impacts upon the magnitude of errors associated with large-scale surveys whether they assume the form of a census or sample survey. In its various manifestations, non-sampling error continues to threaten the quality of results emanating from censuses and sample surveys. In the context of census enumeration, it ought to be the only source of error insofar as sampling error is not of any theoretical significance. In contributing toward efforts geared toward reducing total error associated with census enumeration in Caribbean Micro-States, the paper examines the characteristic features of unit and item non-response. Unit non-response has implications for the under-enumeration of populations and sub-populations. In this regard, they can produce biased results that misrepresent the magnitude of population characteristics. For any given item, non-response may bias results upward or downward. Thus, there is a need to discern whether or not, regular patterns emerge between the socio-demographic characteristics of respondents and non-response associated with any given item. The latter has implications for the imputation of missing data on given items.

In the Caribbean, a Regional Census Coordinating Committee administered the 1990/1991 Round of Censuses. A total of 16 countries and territories participated in this project. However, this paper relates to the census taking experiences in five of these countries, each of which, is considered to be a Micro-State insofar as it is independent, has a land area of less than 700 square kilometres and accommodates a population size not exceeding 300,000. More specifically, the paper draws upon the 1990/1991 census data of Grenada, St. Vincent and the Grenadines, St. Lucia, Dominica and Barbados. It strives to explore relationships that could provide a basis for effectively treating with problems associated with the two main sources of non-response error in different modules. It is expected that the findings will be instrumental in informing processes that can be tested in the context of the other eleven countries. However, noteworthy departures could be indicative of country specific nuances and thus, have implications for replicating the inquiry in the other countries.

To date, region specific inquiries relating to analyses of non-sampling errors are not commonplace. This is true for census enumeration and sample surveys, and is even more critical in the case of the former given the magnitude of the undertaking in relation to sample surveys. In the case of a census, non-sampling error might be so great that it outstrips total error associated with efficient samples. Having been sufficiently privileged to experience efforts toward census planning and data analysis, I am convinced that there is a critical need to initiate scholarly inquiries that pursue such methodological concerns. It is expected that such a pursuit will augur well in enhancing the quality of census administration and preparation, data collection and subsequent data analysis. This is especially warranted insofar as a next round or regional censuses are to be undertaken during the 2000-2002 period. Moreover, there are lessons to be learnt that can also enhance processes akin to administration, preparation, data collection and statistical analyses as they relate to sample surveys.

Saturday, 3:30-5:00

The Influence of Auxiliary, Symbolic, Numeric, and Verbal Languages On Navigational Compliance in Self-Administered Questionnaires

Cleo D. Redline, U.S. Bureau of the Census

Don A. Dillman, Washington State University, USA

Failure to navigate through a self-administered questionnaire correctly constitutes a major source of item nonresponse. In this paper, we propose a theory of navigational compliance that is based upon how respondents process the visual information contained in questionnaires. This information appears in auxiliary, symbolic, numeric and verbal languages, which are woven together on each page in a way that is theorized to influence respondent perception, comprehension, and even judgment of what is to be done when answering a questionnaire. Results from a preliminary experiment are reported in which these languages were systematically manipulated in two distinctively different ways in an effort to test their possible influence on a particularly challenging aspect of navigational compliance, that of skip-pattern compliance. One of these methods was labeled the prevention method. It combined several individual language manipulations, including placing reminders in advance of the question containing a skip instruction and relocating the verbal skip instructions closer to the answer box. Also, the verbal skip instructions were expressed in larger, bolder fonts against a high contrast background. The second method, labeled the detection method, combined several language manipulations as well. These included placing skip directions in a larger bolder font, adding an arrow connection from the non-skip check boxes to the next question, and placing a reminder at the beginning of the next question about who should answer it. Both the prevention and detection methods were found to reduce skip errors of commission (not skipping to the next question when directed to do so) by about half compared to a traditional skip-pattern procedure to be used for the 2000 Census. But neither method reduced errors of omission (skipping questions when not directed to do so). These results provide limited support for the proposed theory of navigational compliance, but further development and testing is needed in pursuit of the goal of effectively guiding respondents through questionnaires without fail.

Saturday, 3:30-5:00

A Congenial Overview and Investigation of Imputation Inferences Under Uncongeniality

Xiao-Li Meng, The University of Chicago

Uncongeniality, defined in Meng (1994, *Statistical Science*), is a theoretical concept that formulates situations where a user's procedure for analyzing a partially imputed data set cannot be embedded into the imputation model used by the imputer. This occurs typically because the user and the imputer have different objectives, have different amounts of information/data, and/or adopt different modes of inferences. This paper provides a "user-friendly" overview and some on-going investigation of this practically important and theoretically challenging real-life constraint. A regression problem is detailed to demonstrate the impact of the uncongeniality on repeated-imputation inferences, namely, inferences using Rubin's (1987, Wiley) combining rules with properly created multiple imputation. The emphasis here is on finding statistically most efficient and practically most effective methods when facing the separation of the imputer and general users with limited information exchange between them, rather than on methods that are guided by idealizations that are only sensible in a congenial environment.

Saturday, 3:30-5:00 Trends in Response Rates

Are They as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century

Charlotte Steeh, Georgia State University, USA

Nicole Kirgis, University of Michigan, USA

Brian Cannon and Jeff DeWitt, Georgia State University, USA

The first systematic examination of changes in nonresponse rates in national surveys spanned a period of twenty years. By 1999 another twenty years will have passed since the publication of this article, and it seems appropriate that we should return to these same surveys to assess their trends at the end of the twentieth century. Has the refusal rate continued to rise or has the use of incentives and improved survey methods stemmed the tide?

As of 1979 it appeared that the National Election Studies and the Surveys of Consumer Attitudes at least had successfully overcome the problem that the greater mobility of adults in the United States presented to survey researchers. Interviewers persistently tracked potential respondents so that the refusal rate climbed while the rate due to non-contact with a household or eligible respondent correspondingly declined. The development of new telephone and computer technology, however, suggests another interesting research question. How have technological innovations affected survey nonresponse? Are refusals still the dominant type or have we now entered a period when noncontacts constitute the greater obstacle. We already know that the shift to telephone interviewing by the Surveys of Consumer Attitudes bumped the refusal rate to a significantly higher level.

These are the main questions that my proposed paper will address with data from the National Election Studies and the Surveys of Consumer Attitudes maintained by the field office at the Institute for Social Research. The fact that the administrative data for each of these surveys is now stored electronically means that, in addition to looking separately at an overall refusal rate and an overall non-interview rate, I will be able to make finer distinctions. Is the refusal rate, for example, due to instant refusals that don't even give the interviewer a chance to select a random respondent or is it more often the case that the refusal comes from the person chosen to be the respondent? Does mode of administration determine the type of refusal interviewers will encounter?

The main emphasis in the paper, however, will be on extending the trends in refusal and noninterview rates over the forty years from 1959 - 1999 and in examining afresh the possible correlates of the changes that do occur. The analysis will not only take into account the timing of technological innovations but also the timing of changes in survey methods, such as the use of respondent and interviewer incentives, the diversification of interviewing staffs, and improvements in sampling procedures.

The findings of the proposed study should help us decide what we need to do in the future to improve participation in surveys by giving us a clear picture of where we have been over nearly a half century.

Response Rates in RDD Telephone Surveys: Are they Increasing, Decreasing, or Remaining the Same?

Brenda G. Cox, James T. Massey, Daniel J. O'Connor, and Kathryn Chandler, Mathematica Policy Research, USA

Conventional wisdom suggests that obtaining response in telephone surveys is becoming difficult. In describing current problems, interviewers mention the increasing use of answering machines and caller ID as well as the frequency with which households receive sales calls. This perception of declining response rates provided the impetus for our investigation of whether the 1990's had witnessed a decline in response rates for random digit dialed (RDD) telephone surveys. Response rate results were compiled for public and privately sponsored RDD surveys conducted since 1990. To allow comparisons across surveys, each survey's response rate was recalculated using the definition provided by the Council of American Survey Research Organizations (CASRO). In this paper we summarize the results of that investigation, including the wide variety of definitions the surveys used in defining response rates, the variation in response rates across surveys, and potential correlates of response rate differences.

Nonresponse in Household Surveys: New Measures and New Insights

B.K. Atrostic and Geraldine Burt, U.S. Bureau of the Census

The Census Bureau and the Bureau of Labor Statistics recently founded an interagency group, the Interagency Household Survey Nonresponse Group (IHSNG), to study nonresponse in several of the large, continuing household surveys conducted by the Census Bureau. The IHSNG group prepared an initial report in 1998, and will also prepare the proposed paper. One of the findings of the initial report was that while all of the surveys examined had experienced increasing nonresponse rates during the 1990s, there did not seem to be a single, consistent explanation for these increases across surveys. It also appeared that the household nonresponse rates did not always measure the same thing across surveys (because of differences in survey designs and procedures) and that the measures were not always calculated exactly the same way across surveys. The report noted the

importance of examining differences in survey design and implementation when examining and seeking to explain trends in nonresponse rates across surveys. In the first part of the paper, IHSNG will update its analysis of trends in the Census Bureau's traditional measure of household nonresponse – average annual nonresponse rates. It also will further analyze and seek to explain an alternative measure introduced in the group's initial paper – initial contact nonresponse rates.

In addition to examining the traditional measures and how they are calculated, the 1998 IHSNG report also made a series of recommendations for innovations in measuring and monitoring nonresponse in these surveys. These recommendations would both improve measurement of nonresponse sources and trends in a particular survey and aid in making cross-survey comparisons. The first recommendation was to define consistent sets of core nonresponse concepts and statistics that can currently be calculated for each survey. The second, related, recommendation, was to define an expanded set of consistent and appropriate concepts and statistics for each survey. The minimum set of statistics would include annual and initial contact nonresponse rates for the sample as a whole. In addition, it would include statistics appropriate for the goals of the survey, including rates for key population subgroups, for key data items, and for geographic areas, such as region, primary sampling unit, or state. The set could also include related statistics, such as cooperation rates. The second part of the paper will report progress and results of launching the group's recommended innovations in monitoring nonresponse rates. It will present and report on nonresponse measures that are consistently defined across surveys, and will identify the appropriateness of these measures for each survey and for comparisons across surveys.

Saturday, 3:30-5:00

Imputation 3

Imputing Missing Data From Volunteered Comments

Nedra S. Whitehead, Centers for Disease Control and Prevention, USA

Mail surveys are less expensive than telephone or face-to-face interviews, but are more prone to both unit and item nonresponse. Imputation of missing values is one way to address missing data. Imputation may be done using the respondent's answers to other questions, answers from similar respondents, or other information. If captured, a respondent's volunteered comments, whether written or spoken, can also be useful for imputing missing values. The Pregnancy Risk Assessment Monitoring System (PRAMS) is an ongoing survey of a state population-based sample of women who delivered a live born infant 2-6 months previously. PRAMS is currently conducted in eighteen states. Data are collected by mailed questionnaires with phone follow-up for the non-responders. In addition to question responses, PRAMS captures respondents written or spoken comments and makes the comments available to analysts.

We used written comments associated with questions on tobacco and alcohol use to impute missing values for these questions. Values were imputed for 1.5% of 74,317 observations. Most of these were women who did not check any answer, but stated they did not smoke or did not drink. Before imputing these values, 4.6% of observations were missing information on tobacco use and 6.3% were missing information on alcohol use. After imputation, the percent missing was 4.4% for tobacco use and 5.4% for alcohol use. The estimated prevalence of smoking during the three months before pregnancy and smoking during the last three months of pregnancy decreased 0.5 percentage points after recoding. The estimated prevalence of drinking alcohol during the last trimester also decreased 0.5 percentage points, but the decrease in the estimated prevalence of drinking alcohol before pregnancy was larger, 2.0 percentage points. Recoding also affected the estimated association between smoking before pregnancy and experiencing stressful life events in the year before delivery. The relative risk was decreased by as much as 0.2 and, in general, became closer to the null hypothesis, 1, after recoding.

Even a small proportion of missing values can affect study results. Imputing data from respondents comments provides one way to reduce bias from missing data.

Strategies for Evaluating and Reducing the Item Imputation Error for the 2000 Census

Yves Thibaudeau and William E. Winkler, U.S. Bureau of the Census

We attempt to quantify the uncertainty associated with non-response to four categorical items that will be requested at Census 2000. The items are: tenure, race, Hispanic origin, and sex of the householder. We propose a hierarchical log-linear model to describe the probabilistic process inducing the interactions between the four items, and also between neighbors in terms of the same items. The model channels information from three levels of geography: the household, the neighbor, and the neighborhood (tract).

Based on the model we generate the predictive distribution of several cross-tabulations over the four items using the returns from the 1998 dress rehearsal of Census 2000 in Sacramento, Calif.. To generate the predictive we use Schafer's adaptation (1997) of bayesian iterative proportional fitting (BIPF). BIPF algorithms were first proposed by Gelman and Rubin (1991). BIPF yields approximations of the posterior distribution of the inclusion probabilities associated with the model, and of the predictive distribution of the cross-tabulated Census population counts. The predictive distribution allows us to evaluate the performance of the item imputation methodology for Census 2000, the sequential hot-deck, relative to the assumptions of the model. The predictive also allows us to give an error assessment of our alternative to the sequential hot-deck. Our proposed alternative is a randomized assignment based on the MLE of the inclusion probabilities.

Use of Highly Influential Covariate and Multiple Waves in Reducing Nonresponse Impact in Surveys

R. Sowmya Rao, Mark E. Glickman, and Robert J. Glynn, Boston University and Harvard Medical School, USA

Using an influential covariate available on all sampled individuals along with an intensive follow-up of non-respondents is a useful approach to reduce the impact of non-response. We examined the performance of several different non-response models to account for such covariates and multiple survey waves. The problem was motivated by a survey of birth control use in 106,463

women who began taking the teratogenic drug, Accutane. Non-respondents to a first mailing were followed with two more mailings and then a telephone call. The first mailing had a response rate of 59.7%; the overall non-response rate was 15.2%. Age was available on all women and was strongly associated with both response and outcome: the youngest and oldest women had higher non-response rates and lower birth control use. We considered both multiple imputation and weighting strategies using information on age and time to response. Our models assumed that non-respondents were: (i) like all respondents, (ii) like all respondents controlling for age, (iii) like late respondents, (iv) like late respondents controlling for age, and (v) projected from a model which includes trend and an interaction between age and time to response. The results were sensitive to the assumptions made about the missing data and suggested that the complete case analysis gave biased estimates. We used simulation studies to examine the potential bias in this setting and compared mean squared errors of our models under different assumptions about the missing data mechanism. Results describe the extent to which availability of covariate information and use of multiple waves.

Saturday, 3:30-5:00

Theories of Survey Participation

Communicative Competence as Indicator for Survey Participation

P.Ph. Mohler, M. Wiedenbeck, J. A. Harkness, and A. Koch, ZUMA, Germany

This paper concentrates on persons' characteristics as factors of survey non-response. Research on personal characteristics concentrated for a long time on "objective" factors like respondents' age, gender, education, income etc. It is only recent that more effort is put on attitudinal data. The power of attitudes as factors for survey participation could be shown by the "Outreach Evaluation Survey" (1990, US Census Office) and the "Survey of Census Participation" (1990, NORC - Chicago).

Societal attitudes and reported behavior should also be of specific relevance for participation. Because survey research claims to benefit the "res publica", attitudes connected with civic duties should be correlated with survey participation. First results from a 1996/97 "Detroit Area Study" (ISR, Ann Arbor) indicate that indeed "civic obligations" and "community activity" are associated with participation in a mail survey.

Surveys are communication, whatsoever may be the mode of the respective survey. If respondents feel not at ease with the communicative situation "interview", it is likely that they will avoid to enter the stage and take part in the act. Analysing panels of the 1994/95 and 1996/97 panels of the German General Social Survey (ALLBUS) and the German International Social Survey Programme (ISSP) indicate that "willingness to give answers" (Antwortbereitschaft) in the first face-to-face wave of the panel is a strong indicator for non-response in the second, mail survey wave (one year later). These analyses replicate also the above mentioned two American studies in a different setting and a different country. They also take into account the research on civic obligations and participation.

The two German panel studies take advantage of the comprehensive information on different attitudinal as well demographic areas in the first wave (ALLBUS/ISSP 1994 and 1996). However, the interpretation of the results has its limits, because the non-responses in wave two are a "positive selection" (they already took part in wave one).

In order to close the gap between people who at least responded once and those who did not, data from a special non-response study from as early as 1986 were re-analysed. Two randomly selected sub-samples of all addresses (5278) used in the 1986 ALLBUS face-to-face study were followed up by phone. The first sub-sample consisted of 541 people who took part in the face-to-face study, the second consisted of 1518 of those who did not respond in the face-to-face part. 670 of these participated in the short phone survey and their responses can be compared to the 407 respondents of the ALLBUS who took also part in the phone survey.

First results indicate, that communicative aspects play also a role in survey non-response for those who were not willing to participate in a long face-to-face study. Furthermore, the design of the study allows a replication of the 1994/95 and 1996/97 analyses by predicting participation in the phone follow-up from information contained in the face-to-face interview. The results confirm those of the later (1994/95 & 1996/97) studies.

In addition to the theoretical gain of these analyses in explaining non-response, one can point to a more practical outcome too. Further in depth studies on communicative competence can now be designed to construct a small set of "communicative competence indicators". These, in turn, could be used by contacters and interviewers as key information for achieving higher response rates.

Theoretical Explanations for Nonresponse in Telephone Surveys: Findings From a Survey of Nonrespondents

Jennifer A. Parsons, Timothy P. Johnson, and Amy DeGrush, University of Illinois at Chicago, USA

Declining response rates in telephone surveys, particularly random-digit-dial (RDD) surveys, have been documented consistently over the past several years. While the consequences of low response rates are understood, there has been more of an emphasis on increasing cooperation and response rates than there has on developing an understanding of why it has become so difficult to persuade the general population to consent to interviews by telephone. In particular, theoretical explanations for increased nonresponse in RDD surveys have seldom been investigated. The present study aims to broaden our understanding of the factors contributing to nonresponse in RDD telephone surveys. We conducted a nonresponse survey of 108 persons who either refused or were otherwise not available to participate in a statewide RDD omnibus survey to test existing theories of nonresponse. Omnibus poll nonrespondents were offered a \$25 incentive for cooperating with the follow-up survey. The questionnaire included several key demographic items, measures of perceived requests for and intrusiveness of telephone survey requests, and indicators of respondent availability. Moreover, using constructs proposed by Groves and Cooper (1998), we developed indicators to assess the effects of social isolation and social exchange processes on willingness to participate in a telephone interview. In addition, a random subset of 114 of the persons who cooperated to the statewide omnibus survey were administered the same block of questions,

allowing for comparisons between those who cooperated to and refused the same survey request. Major findings indicate that omnibus poll nonrespondents were more likely to reside in urban areas, have listed telephone numbers, and feel that surveys are an invasion of privacy.

A Video for Training Interviewers in Doorstep Techniques to Maximise Survey Response

Graham Farrant, Aelfthryth Gittings, and Peter Lynn, National Centre for Social Research, UK

The presentation will consist of the screening of a training video for interviewers that has been produced by the (UK) National Centre for Social Research (formerly known as SPCR).

The origins of the video lie in two investigations carried out by the National Centre's Survey Methods Centre, into the nature of the initial interaction between interviewer and potential respondent in face to face surveys. The studies examined how the behaviour of survey interviewers affects response rates, and identified the first moments of the interaction as the most critical. Both studies found that the most experienced interviewers tended to have the lowest refusal rates. They seemed better prepared, more aware that potential respondents were often slightly anxious and distracted, and more adaptable to the cues they encountered. They tended more often to abandon any predetermined script, replacing it by more appropriate and more subtle (frequently non-verbal) messages, in an attempt to deflect reluctance or suspicion which might otherwise have jeopardised co-operation. It was clear that these skills - though often acquired over long periods - could be taught, but that they would be difficult to teach *via* conventional training methods. There was thus a perceived need for more creative approaches to interviewer training in 'doorstep skills'. The video is based on those studies, and draws on the considerable body of knowledge among the National Centre's most experienced interviewers and field trainers. The video is aimed at interviewers and trainers of interviewers who conduct social research interviews. It is designed to be suitable for people at all levels of experience and to be suitable either for use in formal training or briefing sessions or for individual home study. The video is divided into 14 short sections, each dealing with a different topic, and comes with a booklet for trainers.

Saturday, 3:30-5:00

Adjustment for Nonresponse

Applying Response Propensity Models to Record-Based Health Surveys

Catharine W. Burt, National Center for Health Statistics, USA

National event-based health statistics suffer from clustering effects of responses. Because events are abstracted from medical records from a sample of providers, the resulting distribution of visit and patient characteristics are dependent on responding physicians having the same characteristics as nonresponding physicians. This paper examines the relationship between response propensity and provider characteristics for part of the ambulatory care component of the National Health Care Survey. The National Ambulatory Medical Care Survey (NAMCS) measures the distributions of patient and visit characteristics for patient encounters with office-based physicians. The response rate for the NAMCS is 70 percent. With such a level of nonresponse it is possible that survey estimates may contain nonresponse bias despite the corrective factors applied at the estimation stage. Unit response may vary across provider characteristics. In studying variables related to survey participation, it is important to gather as much information about the sampling unit as possible. Typically, sampling frame variables are the only source available for such analysis. For this study, additional variables were collected from various sources including geographic contextual variables, survey administrative variables, and other variables obtained via auxiliary data collection efforts. A discussion of the possible effects of unit nonresponse on the validity of survey results is also included.

Data from the 1997 NAMCS were matched with county data from the Area Resource File (ARF), physician data from the American Medical Association and American Osteopathic Association, and data from a special survey of nonresponding physicians to examine differential unit nonresponse and resulting bias in survey estimates. The NAMCS is a physician-based survey where information on a representative sample of patient and visit characteristics are abstracted onto a one-page encounter form for each sampled visit. A sample of 2,500 physicians were approached for induction into the 1997 NAMCS resulting in approximately 25,000 encounter forms. A one-page mail-out/back survey was sent to approximately 600 nonresponding NAMCS physicians to collect practice characteristics such as size and ownership. Provider characteristics included type of physician specialty, geographic region, MSA status, practice ownership, and practice size. Geographic contextual variables included unemployment rate, percent below poverty, HMO penetration, population density, per capita income, mortality rate, rate of physicians to population, percent of population that is white, percent of population that is black, and percent of population 65 years and older among others. Survey administrative variables included Census Bureau regional office, field representative age, gender, and job longevity. Cases for most of the analyses were weighted by their initial sampling weight. Logistic regression was used to model unit response and chi-square analysis was used to compare characteristics with response status using SUDAAN. Survey bias estimates were made for various survey items.

Variables in the model that were found to be associated with survey cooperation were geographic region and gender of the physician. Other variables that are related to cooperation include whether the physician is in solo practice or not. The effects of this differential nonresponse on measurement bias was greater for some variables than for others. Variables that were included in the nonresponse adjustment factor for the sampling weights had less observed bias than those variables that were not included in the adjustment factor.

Obtaining variables on sampling units from multiple sources is helpful for studying the effects of survey nonresponse. Strategies for increasing response and adjusting for nonresponse are discussed.

Methodology to Assess and Compensate for Nonresponse in Complex Sample Surveys

Meena Khare and Trena M. Ezzati-Rice, National Center for Health Statistics, USA

Despite intensive efforts made to reduce nonresponse, unit and item nonresponse occur virtually in all population based surveys and affect the quality of the survey and survey estimates. Unit nonresponse generally occurs due to noncontact or refusal to participate, and item nonresponse occurs due to unwillingness or inability to answer specific questions or refusal to specific examination measurements. The third National Health and Nutrition Examination Survey (NHANES III, 1988-1994) is a multipurpose periodic national survey conducted by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC). NHANES III, based on a complex sample design, is designed to provide national statistics on the health and nutritional status of the U.S. civilian noninstitutionalized population. It collects a wealth of information including socio-demographic, economic, and medical history data through household interviews, and health and nutrition data through standardized physical examinations conducted in especially equipped Mobile Examination Centers (MEC). To assess the extent of nonresponse, weighted and unweighted interview and examination rates were computed. To evaluate potential for nonresponse bias, socio-demographic, and health-related characteristics of the respondents and nonrespondents were compared from selected interview questions and examination measurements. Unit nonresponse in sample surveys is generally compensated by adjusting sampling weights, and item nonresponse with imputation. In the NHANES III, of the 39,695 selected survey participants aged two months and older, 86% were interviewed in their home and 78% were examined in the MEC. Examination component nonresponse ranged from 2 to 23% among examined persons with large variations by age, gender, and race-ethnicity. This paper provides a general guideline on how to assess the extent of unit and item nonresponse and presents a summary of alternative methods to account for nonresponse in complex surveys.

Nonresponse in Nested Longitudinal Surveys: A WEE Approach

Christian Kastner, Institut fuer Statistik, Germany

Regression models are a standard tool in analyzing the relationship between one or more dependent variables (response variables) and a set of independent variables (covariates). In longitudinal surveys the classical assumption of statistics, the independence of observations, is violated. Therefore, the usual regression methods as linear regression, logistic regression and so on are not suitable. Marginal regression models are an extension of the generalized linear model (GLM; McCullagh & Nelder, 1989) to correlated observations (see e.g. Fahrmeir & Tutz, 1994). Using marginal regression models the marginal expectation of the response given the covariates---the mean structure---is of primary interest. The association between the observations within a cluster is treated as nuisance.

Inference for these cross-sectional models for longitudinal data can be drawn with the semiparametric generalized estimating equations approach introduced by Liang & Zeger (1986). Qaqish & Liang (1992) extended the generalized estimating equations to nested correlation structures. Such a nested structure might be a longitudinal survey of families or house holds, where the family members are the subunits. In this situation we have correlated observations over time and within each time point the association between family members.

Many studies---especially longitudinal surveys---suffer from missing data. Robins, Rotnitzky, & Zhao (1995) proposed to use weighted estimating equations (WEE) in estimating the mean structure, if missing data are present in dependent variables or if there is wave nonresponse. The observed units are weighted with their inverse response probability. Therefore, the conditional response probabilities are modeled. This is a natural procedure having longitudinal data.

When there are additional subunits within each time point, this conditional procedure is no longer useful. If the subunits have some order---as e.g. families---a modified conditional approach for estimating the response probabilities could be used. If there is no explicit order within the subunit, or if the order should not be taken into account, the response probabilities can be modeled using symmetric models as considered by Qu et al. (1987) or Conolly and Liang (1988).

In this paper several weighting methods will be presented. The several weighting methods are compared using simulated data and a data from the German socio-economic panel (GSOEP) demonstrating the practical use of this approach.

Sunday, 8:30-10:00

Nonresponse in Exit Polls: A Comprehensive Analysis

Daniel Merkle, ABC News Polling Unit, USA

Murray Edelman, Voters' News Service, USA

This paper uses data from Voter News Service's election day exit polls to study factors that influence survey nonresponse, and it takes the important next step of looking at the impact of response rates on survey error.

Data from exit polls can make a unique contribution to the study of nonresponse. Unlike most surveys that employ relatively homogenous pools of interviewers, VNS employs a larger, more heterogeneous group of interviewers each election. As with other types of in-person surveys, it is possible to obtain, by observation, some basic demographic information on nonrespondents. Finally, and most importantly, a measure of survey error can be computed for the key variable of interest in exit polls, the vote. This error measure, along with the response rate, can be computed for each interviewer (i.e., each precinct).

After a discussion of the methodology, we begin by exploring the influence of a variety of factors on response rates. First we look at the respondent characteristics of age, race and gender. Next, we explore which interviewer characteristics and election day factors are related to response rates. Then we look at the extent to which the interviewer characteristics of age, race and gender interact with these same respondent characteristics in terms of nonresponse. Finally, we explore the relationship between response rates and survey error.

Sunday, 8:30-10:00

Evaluating Nonresponse Error in Mail Surveys

Dana Lee Moore and John Tarnai, Washington State University, USA

Past research on evaluating response to mail surveys has maintained an almost singular emphasis on improving response rates. This research has been successful in identifying specific factors that can be manipulated to improve response rates, such as personalization of mailings, inclusion of stamped return envelope, cover treatments, the number of follow-up contacts, special postage, and use of incentives. For mail surveys to achieve high response rates, many response inducing techniques need to be applied in conjunction with one another and past research is consistent in showing that the use of multiple follow-up contacts, special postage and incentives are the most effective response inducing factors overall. What this research has not done is to show whether improvements in response rates are also successful at reducing non-response error, i.e. the extent to which respondents differ from nonrespondents on characteristics relevant to the study. Nor has it been examined which non-respondents become respondents and what their characteristics are.

In this paper we propose to: 1) review available research on response improving techniques that have demonstrated success at reducing nonresponse error, and 2) analyze the results of a series of experiments on response inducing techniques in which an effort was made to measure the reduction of nonresponse error. The method for doing this will be a regression model of response, where by response is a function of demographic characteristics and survey factors. The data for this study comes from individuals who had turned in out of state drivers licenses to obtain a state license. This data base provided information on demographic characteristics for individuals selected to participate in the survey thus allowing for a comparison of respondents and non-respondents to the six survey treatments or cohort studies. A third objective is to describe the areas of needed research in mail surveys to reduce nonresponse error.

For the empirical study, four survey factors were held constant across all experimental treatments, namely, saliency (topic), number of mail follow-ups, postage use across follow-up contacts, and survey length. The factors that were manipulated in the experimental treatment groups include: questionnaire cover treatments, use of incentives, pre-notice letter, and personalization. The factors that were not manipulated include age, gender, and urban location. Probit and logistical regression techniques are used to model the effects of the experimental variables, demographic variables, and their interaction on response.

Preliminary analysis suggests that there may be differences in older versus younger people responding to survey elements and that there may also be an interaction between age and the use of incentives that is important to achieving response and the reduction of non-response error. Overall, this analysis provides support for the idea that further tailoring and customization of survey design for targeted population subgroups known to be least likely to respond may help reduce non-response error in mail surveys.

Sunday, 8:30-10:00 Replication Methods

Capturing and Evaluating the Effect of Reweighting for Unit Nonresponse

Phillip S. Kott and Michael Bellow, National Agricultural Statistics Service, USA

One popular way to handle unit nonresponse in a survey is to treat it as another phase of sampling. Typically, the sample is divided into mutually distinct response homogeneity groups, with every unit (element) in a group assumed to be equally likely to respond to the survey. This assumption forms the basis of a quasi-randomization model. With appropriate reweighting, an estimator based on the respondent sample can be shown to be quasi-randomization consistent, i.e., randomization consistent under the quasi-randomization model. If reweighting is done properly, the estimator may also be unbiased under a parametric model for an item value of interest, assuming that the model holds and unit nonresponse within a group occurs at random. The latter can be a weaker assumption than quasi-randomization.

Previous work by Kott and Stukel (SURVEY METHODOLOGY, 1997) and Kott (SSC: PROCEEDINGS OF THE SURVEY METHODS SECTION, 1997) shows how a conventional stratified jackknife can be applied to a reweighted estimator based on a two-phase sample. These arguments can be extended without much difficulty to a delete-a-group jackknife in which the sample is divided at random into 15 mutually distinct replicate groups, which are then treated like clusters in a simple jackknife. These replicate groups are independent of the response homogeneity groups described above.

This paper uses the delete-a-group jackknife not only to estimate mean squared errors under a quasi-randomization model, but also to compare two alternative formulations of the response homogeneity groups. The survey data used are from the National Agricultural Statistics Service's Crops/Stocks (CS) Survey. One method of group formation employs a simple size measure, which is based on control values for cropland and stock storage capacity. The other method employs a more complex clustering algorithm tied to the survey's multivariate selection routine (see Bailey and Kott, ASA: PROCEEDINGS OF THE SRMS, 1997). The delete-a-group jackknife is used to determine whether these group-formation methods lead to significantly different results (if so, one must be based on an inappropriate model). If both methods are assumed to lead to unbiased estimation, then the delete-a-group jackknife can be used to assess whether one method is more efficient (produces smaller mean squared errors) than the other.

The delete-a-group jackknife is employed in an analogous manner to evaluate two contrasting reweighting schemes for the CS Survey. In one scheme, the number of farms in a response homogeneity group estimated using the respondent sample is forced to equal the estimated number of farms computed from the first-phase sample. In the other scheme, the estimated total of a relevant control value based on the respondent sample is forced to equal the estimated total for this value based on the first-phase sample. Interestingly, both schemes assume the same quasi-randomization model but different parametric models.

Simulation Study of "Informative Nonresponse"*Per Nilsson, Statistics Sweden*

In some situations the unit nonresponse is concentrated to certain domains of study in the population, for instance, having a stronger tendency to occur among units with low or high values. Often this situation is referred to as the case of "informed nonresponse". Traditional estimators will then be biased if they are not adjusted for by reweighting or by using imputation procedures.

In a simulation study is shown how bias could be reduced and even be eliminated by using bootstrapping methods in combination with reweighting. Different models are applied to describe how nonresponse occurs in different domains of variation. The simulation study is based on standard distributions such as the normal and the gamma distributions.

Estimating Variance Due to Imputation in ONS Business Surveys*Susan Full, Office for National Statistics, UK*

Most of the ONS short-term business surveys are stratified designs with some strata completely enumerated. For the completely enumerated strata values for non-responders are imputed. The imputed values are then used as if true values and the variance estimates were calculated using the standard formulae. This procedure leads to an underestimate of the true variance. It is therefore important to be able to estimate the variance due to imputation in surveys and also to use these estimates to compare different imputation methods.

The method of imputation used in these surveys is a simple autoregressive imputation using a weighted respondent average of period on period growth to determine the autoregression parameter ("link"). Some robustness is introduced by using two year's link estimates from the same point of the seasonal cycle, and weighting these together to produce the final imputation link.

This paper looks at using resampling methods to estimate the variance due to imputation in a typical ONS business survey.

Sunday, 8:30-10:00**Imputation Theory****Regression Models for Non-Ignorable Nonresponse***Md Hossain, Shinto Eguchi, and Toshiya Sato, Graduate University for Advance Studies, Japan*

In this paper we consider a simple regression model as a model for the outcome variable. The data can be missing either in the outcome variable or in the covariates, but not both simultaneously, and the missing data are nonignorable. The parameter of interest is the mean response for a given set of covariates. The main objective is to check the local sensitivity of the regression estimates for the nonignorability of the missing data. The proposed methodology is based on the idea of local selectivity (Copas and Li (1997)), and we have adapted it to the pattern-mixture model. The proposed methodology is applied to the phenobarbital data discussed by Farwell et al. (1990).

The Influence of Nonresponse Treatment Methods on Parameter Estimates of a Regression Model When the Nonresponse is Nonrandom*Nand Kishor and Lucas Othuan, The University of British Columbia, Vancouver, BC, Canada*

When item nonresponse is random, the problem may be ignorable because the statistical parameter estimates tend not to be seriously affected (Little & Rubin, 1987). However, most field data collected via surveys do not miss randomly, and very little is known about the best way to analyze such data sets (Little, 1992; Roth, 1994). Several nonresponse treatment methods (NRTM) have been proposed, but all these methods assume that data are missing at random (Little & Rubin, 1987).

Although several NRTMs like listwise and pairwise deletion, mean substitution, regression and EM imputation are already included in analyses software, not much is known about the relative impact of these NRTMs on parameter estimates. Roth (1994) noted that a comprehensive evaluation of these methods, specifically the more complex methods (e.g. EM imputation) is seriously lacking. A limited number of studies using randomly missing data suggest that EM imputation produces less biased estimates than listwise deletion (Graham & Donaldson, 1993), and more accurate estimates than mean substitution (Azen, Van Gulder, & Hill, 1989).

However, most researchers have used randomly deleted data, and only a few have incorporated patterns of incomplete data. Data sets where nonresponse is random are rare in real research. Nonrandom nonresponse may occur when respondents have trouble understanding a question and skip it, or when respondents may only partially complete a survey. Such omissions usually form clusters at the middle or at the end of a questionnaire, with the possibility of forming a monotonic or a non-monotonic missing pattern. Moreover, real data are usually non-normal. Fleishman (1978) had noted that many of the psychological variables found in practice are skewed and/or kurtotic to various degrees. Micceri (1989), and Sawilosky and Blair (1992) also observed that many data sets in real situations are skewed and/or kurtotic.

In a simulation study, we have investigated the relative performance of four common NRTMs: imputation via expectation-maximization method (EM), imputation via linear regression (RS), mean substitution (MS), and pairwise omission (PW), with respect to bias in estimates from a multiple regression model (MRM). The design had 72 different conditions resulting from three levels of normality (NM), three levels of sample size (SZ), three levels of proportion of missing data (PM), and two missing patterns (PT). Under each condition, nonresponses were treated by all four NRTMs before fitting a MRM with nine predictors. All decisions about the design of the study were based on a meta-analysis of 177 published papers that used MRM.

We have found that performance of MRTMs with respect to bias in parameter estimates depends on nonnormality, proportion

missing, and missing pattern. With moderate sample size (94, 153), bias in R2 due to nonnormality was similar with all MRTMs. However, with large sample size (265), bias in R2 was due more to percent missing than to nonnormality and similar for all MRTMs but PW. At large sample size, proportion missing introduced bias in R2 when missing data were treated with EM method. The MRTMs differed little in their performance when nonresponse was only ten percent, regardless missing pattern. However, the performance of MRTMs was more differentiated when missing pattern was nonmonotonic. Overall, bias in R2 was the smallest with PW and the largest with EM. Bias in regression coefficients was mainly due to nonnormality, and the MRTMs differed most when missing pattern was nonmonotonic. Overall, PW outperformed the other MRTMs despite the recently suggested superiority of the EM (Azen, et, 1989; Graham & Donaldson, 1993, Little, 1988, Muthen, Kaplan, & Hollis, 1987).

Some Theoretical Results for Person Mean Imputation in the One-Factor Model.

Coen A. Bernaards, Utrecht University, the Netherlands
Klaas Sijtsma, Tilburg University, the Netherlands

In much test and questionnaire research one is confronted with the phenomenon of missing item responses. Missingness can cause many problems in consecutive multivariate analyses of data, for example factor analyses aimed at finding the underlying structure of the data. This paper is concerned with missing item responses in an exploratory maximum likelihood factor analysis setting. Several methods are used for correction or substitution of the missing item scores, such as listwise deletion, pairwise deletion and imputation methods. The methods are applied ad hoc in practical situations since they are readily available in computer programs such as SPSS. However, surprisingly little is known about theoretical implications of the use of such missing data methods for the final factor loadings, probably due to the mathematical complexities that occur even in the most simple case: the three variables, one factor model. This case is the only one for which the likelihood equations can be solved explicitly. Therefore, this case will be used throughout the presentation.

This research extends work by Brown (1983) and Bernaards and Sijtsma (in press). Brown (1983) studied pairwise deletion and listwise deletion in the three variable, one factor model with missings present only on the third variable. He established conditions under which pairwise deletion resulted in smaller asymptotic variance of the factor loading corresponding to the variable with missings than the variance resulting from application of listwise deletion. Bernaards and Sijtsma (in press) performed a simulation in order to study the effects of several missing data methods on factor analysis of rating scales. They found that, in general, imputation methods gave better results than listwise deletion in the sense that the squared difference between the factor loadings of the complete simulated data was closest to the factor loadings based on the data including imputed scores. In the same way, imputation of the mean per person gave the best results of all imputation methods.

Our presentation extends Brown's work to the case of person mean imputation when factor analysis of the data is envisaged. Exact mathematical expressions are derived for factor loadings based on data including imputed scores. For simple situations conclusions can be drawn on which missing data method will give the smallest variance of the factor loading. Next, these results are compared to the bias which results from each of the missing data methods. The variance and the bias supply the researcher with useful tools to decide which missing data method should be used in practice. Extensions to higher dimensions and to other missing data methods are provided. Finally, a real life example shows the result of the developed methodology.

Sunday, 8:30-10:00

Other Issues in Household Nonresponse

Comparison of Random Digit Dialling (RDD) and Area Probability Sampling (APS) on Nonresponse and Within Household Noncoverage

Donna J. Brogan, Maxine M. Denniston, Jonathan M. Liff, Elaine W. Flagg, Ralph J. Coates, and Louise A. Brinton, Emory University, USA

The case-control study is an important method for identification of potential risk factors in cancer epidemiology. Population based controls typically are used, frequently using random digit dialling (RDD). However, concerns persist about RDD's potential bias from higher nonresponse rates and noncoverage. In the context of a breast cancer case-control study of women aged 20-54 years in Atlanta, we compared two independent sampling methods for obtaining a control group: RDD (n = 652) and area probability sampling (APS, n=640). All women were interviewed face-to-face in their home. The two control groups were compared to each other and to 1990 sample Census data (PUMS).

Household (HH) screening determined whether any eligible women (aged 20-54) resided therein. The HH screening response rate was significantly higher for APS (94.9%) than for RDD (89.4%), primarily due to RDD's higher refusal rate. The interview response rates for the two samples were equivalent. The overall survey response rate for APS, compared to RDD, was significantly larger: 76.5% versus 71.2%. The HH contact procedure for screening differed by sample--telephone call for RDD and letter to HH for APS, followed by an interviewer's visit to the HH. The letter contact and personal visit with HHs may have contributed to our higher APS screening rate.

The RDD sample estimated a significantly larger percentage of households (HHs) with no women aged 20-54 years (41.3%) than either APS (34.9%) or PUMS (33.5%). The HH screening respondent may be more reluctant to reveal HH composition over the telephone than face-to-face, especially if the only adult(s) in the HH is(are) female.

Note that failure to answer honestly about HH composition is not the same thing as refusing to provide HH screening information. In fact, failure to disclose age-eligible female residents of the HH actually increases the screening response rate since the screening respondent is able to successfully complete the HH screen by providing false information. Factors which may lead RDD HH samples to refuse the screening process may also operate to deny the existence of age-eligible subjects in the HH, given participation in the RDD screening process.

In weighted and clustered analyses, the RDD, APS and PUMS samples made similar inference on age, number of live births, HH income, and telephone coverage (APS vs. PUMS only). Both APS and RDD samples were more educated and more likely to be married than the PUMS sample, factors typically related to survey response. The RDD sample estimated a smaller percentage of the population to be black. In assessing breast cancer risk factors also available in the PUMS data, the RDD and APS samples were fairly comparable.

We conclude that RDD sampled HHs are less likely to participate in the HH screening process and, if cooperative, some HHs do not disclose that age-eligible women reside therein. RDD surveys which target a specific subpopulation, especially women, need to develop improved methods for screening and enumeration of HHs.

Screening Methods in NLSY97

A. Rupa Datta, National Opinion Research Center, USA

The National Longitudinal Study of Youth, 1997 Cohort (NLSY97) is the newest cohort in the National Longitudinal Studies Program of the U.S. Bureau of Labor Statistics. Initially fielded in 1997, the study design calls for a nationally representative sample of youths born in 1980-1984 and living in a household in the U.S. in 1997. [A similarly defined over-sample of black and hispanic youths is also part of the study design.] To identify and recruit this sample, the National Opinion Research Center (NORC) at the University of Chicago implemented an increasingly popular screening method, often labeled, "Screen and Go."

Traditional screening processes have called for a field effort to collect household information necessary for the screening decision (typically, age, sex, race, and relationships between household members). Statisticians then process the collected data and select the desired sample, which is once again distributed to field staff for main data collection. "Screen and go" collapses this sequence of events, so that interviewers are able in one visit to collect screening data, identify eligible respondents, and solicit cooperation for the main interview. Although "screen and go" has been used for years in studies with relatively simple area-probability designs, the advent of CAPI now permits its use in studies with far more complex sample requirements.

The proposed paper will review the advantages and disadvantages of "screen and go" implementation from the operational, statistical, and final analytic points of view, and throughout the process of study design, sample recruitment, and main data collection. In addition to the screener and Round 1 data, data from a small number of validation studies will also inform the assessment of "screen and go" as a desired data collection strategy. In particular, the paper will attempt to document the effects of "screen and go" in increasing and diminishing non-response among different sample sub-populations.

The longitudinal structure of the NLSY97 lends an additional element to this discussion. While "screen and go" is often used for recruiting cross-sectional samples, its use in identifying the NLSY97 sample implies that nonresponse experienced in the initial round will be perpetuated in the future rounds of the study. Information about nonresponse in the screening and initial round of the data can be used to determine a probability of selection into the NLSY97 sample, which can then be used to estimate (and perhaps correct/control for) the effect of initial nonresponse on later rounds' analytic results.

A New Quality Approach for Telephone Survey: From Interviewers Training to Daily Monitoring System to Prevent and to Measure Nonresponse

Maria Giuseppina Muratore and Luciana Quattrociocchi, Statistics Italy

In the last quarter of 1997 ISTAT carried out a survey on Citizens' Safety and sexual violence. Fifty thousand persons had been telephone interviewed.

Every survey aimed to study private and/or very intimate behaviour require not only technological or informatic innovations, but also a comprehensive change in the whole production process in order to maintain the scientific correctness. This advantages provided by telephone interviews encouraged the researchers to choose this technique.

However the telephone technique did not suffice to gain easy access to households and to guarantee the quality of data. The refusal and non-response rates are very low for traditional multi-purpose surveys that results from face-to-face interviews made by commune interviewers generally known by the households.

Therefore other instruments should be used as well to reduce refusal and non-response rates and to have more effective telephone interviews. For this reason we planned the survey design focusing on two aspects: on one hand studying the development of a telephone interview and the construction of a daily monitoring system, and on the other hand training continuously female interviewers during the whole period of the survey (3 and a half months).

In order to encourage co-operation and increase the response rate, a carefully drafted advance letter, signed from the President of ISTAT, was sent to the whole sample to inform all the members of the household that an ISTAT interviewer was going to call them. The survey monitoring revealed a clear positive effect of the letter on the probability to obtain a complete interview.

This suggested to experiment additional instruments for awakening interest in the survey on the last group of households interviewed. In particular, a group was sent a second introduction letter two weeks after the forwarding of the first one. Another group was contacted by telephone before the actual interview was made, in order to stimulate the response to the survey and also to gather some additional information.

Moreover, the use of a free toll number was valuable and contributed to the success of the survey. Respondents could decide whether to accept to be interviewed or not: this helped to ensure the success of the survey and it is fundamental to establish a co-operative attitude of the interviewed.

This survey method has also strengthened the positive and trustworthy image of ISTAT.

If a great number of respondents would have refused to answer totally or partially to the questions proposed, the results of the survey would have been heavily influenced. Consequently, the training of the interviewers has represented one of the most relevant aspects to be taken into account. Our basic aim was to make the relationship between interviewer and respondent as close, confidential and productive as possible, establishing a quiet and serene atmosphere which would undoubtedly favour a fruitful discussion.

Briefings, debriefings, discussion groups, hall assistance and the daily field monitoring are the instruments which guarantee the interviewers this kind of role, as well as the achievement of aims such as survey quality and scientific exactness. A quality survey has been conducted on interviewers. They have been asked to report differences in training and survey process, comparing ISTAT social survey and usual market surveys. The paper will present results also on this particular quality survey.

An accurate study of the quantitative and qualitative aspects related to telephone survey was arranged by ISTAT to develop a quality control methodology for Computer Assisted Telephone Interviewing operations.

The aim of quality control methodology is to measure and improve some aspects of CATI operations as coverage, instrument design, survey response rates, interviewer performance and data processing.

This quality control monitoring is included in a statistical framework and involves a quantitative approach in order to measure and improve quality in an objective and continuous manner.

As far as the response rate is concerned, the most important reason of household non response could be due to a non contact. To limit this event a CATI automatic mechanism, which consists on re-dialling the number up to five times when the number is busy or not answering, has been implemented.

A "non contact" could be sometimes due to a non eligible household (this is the case of phone numbers belonging to a "secondary" house as those at the seaside or in the country, ...). We analysed the rate of secondary houses by geographical areas using Census data and we found out a strong relationship with our response rates at regional level.

Sometimes the household couldn't be interviewed after the contact because the first respondent refused to answer the preliminary questions necessary to select the right respondent or he/she refused to call the selected person at the phone. In this case the information about the structural characteristics of the household are very limited. Other times the selected person could not be interviewed because of his/her long absence or refuse or illness or interruption of the interview. In these cases when other members of the household were eligible, it was possible to substitute the selected person. The substitution has been done taking into account the individual probability of non response by sex and age, and selecting a new respondent belonging to the class of highest probability. The non response rates by sex and age were obtained in the pilot survey, and updated with the on-going survey data and were used to estimate such a probability. If no substitution was possible the whole household was considered non respondent.

We had to evaluate the structural differences (by sex, age, education level, work condition) between respondents and non respondents and between respondents and whole population. Furthermore a logistic model was applied to analyse the relationship between the response probability and socio-demographic characteristics of sampled individuals.

Adjusting for Nonresponse Among Medicaid Households That Could Not Be Located or Were Located but Did Not Participate in the Minnesota Managed Care Survey

Karen A. Cybulski, Michael Sinclair, Frank Potter, and Anne B. Ciemnecki, Mathematica Policy Research, Inc.

This paper discusses the use of logistic regression modeling techniques using demographic and socioeconomic characteristics and two levels of nonresponse in a survey to appropriately adjust the survey weights to compensate for these nonresponse issues. Mathematica Policy Research conducted the Minnesota Managed Care Survey for The Urban Institute to address the impact of managed care on Medicaid beneficiaries in Minnesota.¹ The state is currently implementing Medicaid managed care in the rural counties in a shift from the traditional Fee-For-Service (FFS) Medicaid program. The Urban Institute's study compared access and satisfaction under managed care in six rural counties in the northeastern part of the state to outcomes under an FFS Medicaid health care plan in 18 rural counties in the northwestern part of the state. We developed the sampling frame from the Medicaid enrollment records. The selection process consisted of a two-stage procedure that selected households first followed by the selection of an adult and a child enrollee from each selected household.

This survey achieved approximately a 70 percent response rate, and two levels of nonresponse occurred. From a total of 6,019 sampled enrollees, a total of 1,431 (23.7 percent) could not be located. The persons without phone numbers resulted from the households without phone numbers, unpublished numbers, movers, etc. Among the 4,588 households that could be located, 489 (10.6 percent) did not participate in the study. To examine the differences in the locatable and unlocatable households, we conducted an analysis using weighted logistic to estimate locatability status from a variety of demographic and socioeconomic variables available from the enrollment records. The results of the modeling process indicated that households with both adults and children enrollees tended to be less likely to be locatable than adult-only or child-only households. Being currently employed, having a racial status of white, or being married also showed higher rates of locatability. In our nonresponse adjustments, we used the inverse of the predicted probabilities from this model as a nonresponse adjustment factor to effectively compensate for these differences in the locatable persons. This model indicated that the nonresponse rates for non-Hispanic whites, Asians, and Native Americans were higher than those for blacks and Hispanics. We believe that the combined effect of these adjustments substantially improved the representative nature of the weighted survey estimates.

Sunday, 10:30-12:00

Diagnostics for the Practical Effects of Nonresponse Adjustments Models *John L. Eltinge,*

U.S. Bureau of Labor Statistics and Texas A&M University, USA

In practical survey applications, development and implementation of nonresponse adjustment methods depend heavily on three related issues.

- (1) The anticipated magnitude of nonresponse bias in a specified parameter estimator and its effect on associated inference methods.
- (2) Related operating characteristics of specific nonresponse adjustment methods.
- (3) The statistical significance and practical significance of the effects observed in (1) and (2).

This paper reviews several classes of diagnostics that have been used to address issues (1) through (3). Special attention is devoted to the strengths and limitations of these diagnostics. In addition, this paper considers the use of power curves to assess both the practical significance and statistical significance of the magnitudes in (1) and (2). The principal ideas are introduced in the context of unit nonresponse, and then are extended to item nonresponse. Some of the abovementioned diagnostics are illustrated with data from a health survey.

Sunday, 10:30-12:00

Question Salience, Question Difficulty and Item Nonresponse in Survey Research

Toshihiko Murata and Patricia A. Gwartney, University of Oregon, USA

Prior research documents the association between survey respondents' cognitive processing and the occurrence of item nonresponse. Researchers have found that survey respondents who have higher sophistication in cognitive processing and a broader knowledge base have lower levels of item nonresponse, particularly, "don't know" and "no opinion." This research extends prior research in this area by focusing upon the qualitative characteristics of survey questions, particularly question salience and difficulty. We expect to find an inverse relationship between question salience and item nonresponse (as salience increases, item nonresponse will decrease) and a direct relationship between various measures of question difficulty and item nonresponse (as difficulty increases, item nonresponse will increase).

We define survey question salience as the closeness, importance, and proximity of question subject matter to respondents. For example, a question regarding the effect of pesticides on the self and family is closer to respondents than a question regarding the effect of pesticides on the general public. Thus the former question has more salience to respondents than the latter. We hypothesize that greater question salience will be associated with lower occurrence of item nonresponse.

We define question difficulty in several dimensions, including conceptual comprehension, provocation, complexity, and specificity. Conceptual comprehension can be greater or lesser according to respondents' prior knowledge of the concept being asked. When conceptual comprehension is lower, the question is more difficult for respondents. Provocative questions are also more difficult. Provocative questions deliberately challenge respondents' viewpoints, while non-provocative questions merely assess respondents' viewpoints. Long and complex questions are more difficult to answer compared to short and simple questions. Vaguely worded questions are more difficult to answer than specific questions. We hypothesize that greater question difficulty will be associated with greater occurrence of item nonresponse.

The data source for this research is the National Risk Perception Survey conducted by Oregon Survey Research Laboratory in Fall 1997 (n=1226, RDD, CATI, with Black and Hispanic oversamples). This survey comprised roughly 250 questions designed to assess adults' perceptions of risks associated with cloning, blood transfusions, eating red meat, asteroids, pesticides, drugs, nuclear power, vaccines, handguns, motor vehicles, tap water, second-hand smoke, radon, and natural disasters.

This survey provides a large range of questions of varying salience and difficulty which allows us to test the hypotheses discussed above. For example, one question module asks about risk perceptions to self and family (more salient) and another question module asks about risk perceptions to the general public (less salient). Some questions require considerable a priori knowledge of risk issues. Another question group asks increasingly difficult questions about parenthood in relation to adoption, abortion, surrogate motherhood, in vitro fertilization, gene therapy and cloning, some of which deliberately challenge respondents. Some questions and probes are quite long and complex, while others are short and simple. Some questions are very specific, while others are (annoyingly) vague.

Our analysis on item nonresponse in relation to question salience and difficulty, controlling for age and education as indicators of cognitive processing, will further understanding of item nonresponse in survey research.

Sunday, 10:30-12:00

Early and Late Respondents

How Different Are the Late Cases?

Arthur B. Kennickell, U.S. Federal Reserve Board

Cases that do not respond to survey interviews often raise a deep uncertainty about the validity of the results obtained from the survey. Typically, practitioners engage in a variety of activities in an attempt to mitigate the possible consequences of nonresponse. These actions most often include a variety of technical adjustments to weights, and more rarely imputation of missing observations. There are also other less technical approaches. One activity that most practitioners appear to believe is important is making strenuous efforts to obtain interviews with at least some classes of cases that appear "difficult." The implicit presumption underlying this action is most likely being that such cases are as close as one can get to representing the cases that are not interviewed at all. To my knowledge, there has been little formal investigation of how the more difficult cases differ from other

interviews.

This paper proposes to look at the characteristics of two types of difficult interviews in the 1998 Survey of Consumer Finances (SCF): those that required many attempts to complete, and those that were obtained relatively late in the field period. As a part of the administration of the SCF, interviewers were required to maintain detailed call records for each case. This information will be used to identify the difficult cases. Experience during the survey field period suggests that these data are very accurate. For the analysis, the main instrument of the SCF provides a rich matrix of data on the finances and demographics of participants in the survey. As time permits, the paper will also look at some of the broader characteristics of nonrespondents and difficult respondents using sample frame data and context data obtained from matching Census tract characteristics to the full sample file.

From a Silent Majority to an Articulate Minority

*Ineke Stoop, Social and Cultural Planning Office of the Netherlands
Frans Louwen, GfK Nederland bv, the Netherlands*

The Social and Cultural Planning Office (SCP) is a Dutch government research agency. For the last 25 years it has contributed to policy making and appraisal in The Netherlands by producing a coherent picture of the state of social and cultural affairs. This requires an ever increasing amount of empirical data, mainly originating from general population surveys. In The Netherlands survey research has its sworn enemies and fervent adherents. Even the adherents, though, point out the high and increasing nonresponse as a major problem. Journalists and politicians have started to question the value of survey research as a response of about 55 percent, even in respectable face to face research by Statistics Netherlands, is no exception. In the SCP 1991 Facilities Survey, a face to face household survey with an auxiliary paper questionnaire for each family member of 6 years and older, response shrank to a mere 43%. Drastic action was required.

For the 1995 wave of the survey SCP contracted a marketing research organization that guaranteed (and realized) a 70% response level. Not only the response rate was much higher than in 1991, but also — to achieve a high response — the maximum number of contacts. Whereas the interviewer visited each address up to three times in 1991 before registering it as nonresponse, the same address was approached up to 15 times (in a single case) in 1995. Approximately 70% of the final 70% response was reached within three approaches, another 25% required up to six visits and the final 5% had to be approached even more times.

Of the 1995 wave early (1-3 approaches) and late respondents (4+ approaches) were compared. The latter group is generally excluded from the net sample as they require more approaches than customary. Thus they probably resemble nonrespondents and the degree of difference between respondents and nonrespondents can be estimated. Early and late respondents do differ in demographic characteristics and in the target variables, facility use. However, after weighting almost no differences remain. This is a reassuring result. Still, the analyses described above give no indication whether the tough, uncooperative nonrespondent who cannot be cajoled into answering the questionnaires differs from respondents, late or early, with the same demographic characteristics.

The answer to this question must come from the 1999 wave of the Facilities Survey. With supreme effort a small random subsample (n=250) of nonrespondents will be coaxed into answering a number of demographic questions, target questions on facility use and questions on response behavior. The results of this additional survey will only be of use if the response among the small group is very large. This will require a lot of persuasion and a lot of money. But, if it succeeds, it will finally be possible to make well-grounded statements about the characteristics, distinguishing features and facility use of the silent minority.

Characteristics of Survey Respondents According to Willingness to Participate

*Lynda F. Voigt, Fred Hutchinson Cancer Research Center, USA
Koepsell TD, University of Washington, USA
Daling JR, Fred Hutchinson Cancer Research Center, USA*

Some surveys include only respondents who readily agree to participate and are accessible, whereas others expend considerable effort to recruit reluctant respondents. This additional effort is only worthwhile if it results in a sample that is more representative of the target population. We compared characteristics of willing, accessible respondents to those of their less accessible or willing counterparts to determine whether they differed with respect to lifestyle, socioeconomic status, health history and demographic characteristics. We used data collected from in-person interviews of 6,151 respondents who were selected by random digit dialing (RDD) and interviewed between 1987 and 1998. We categorized the respondents recruited by RDD into early responders (those who were interviewed within one month of contact, N=3,246), late responders (those who were interviewed more than two months after contact, N=1,332) and initial refusers (those who initially refused but agreed after recontact, N=394). We excluded those who completed the interview between one and two months after initial contact (N=1,179) because some of these interviews may have been delayed as a result of interviewer schedules. We further separated the respondents by gender and adjusted all analyses for respondent age. In addition, we were able to compare medical records of 49 interviewed respondents with 59 refusers that had been randomly selected from the enrollment files of a health maintenance organization (HMO).

Early, late and initially refusing female respondents were similar with respect to annual household income, lifetime number of sexual partners, alcohol use, oral contraceptive and intrauterine device use, and number of births. Late female responders and initial female refusers were less likely to have female relatives with cancer and less educated than early responders. Late female responders were younger than early responders, whereas initial refusers were slightly older. Late female responders were more likely to be non-white, to smoke, and to have had an induced abortion and to have had tests for infertility than early responders. The proportion of women who were retired or housewives was similar in all three groups, but initial refusers were much less likely to be employed in managerial or professional occupations than early responders. Relative to early male responders, late male responders were younger and had a greater number of sexual partners. Initial male refusers were somewhat older and were more likely to be current smokers than were early male responders. All three groups of male responders were similar with respect to race, presence of diabetes, marital status and alcohol use. The only difference between interviewed respondents and refusers from the

health maintenance organization was that refusers were somewhat heavier than those who agreed to be interviewed. Although no differences were found with respect to reproductive history, tobacco use or menopausal hormone use, the number of respondents in each group was small. These findings suggest that additional effort expended recruiting reluctant respondents may result in more accurate estimates of population characteristics that are of interest in survey and epidemiologic research. Unfortunately, no clear pattern emerged to distinguish variables more likely to be biased by ignoring those responders who either initially refuse or respond late.

Sunday, 10:30-12:00

Imputation in Longitudinal Surveys

Weight Adjustment for Nonresponse: A Comparison of Some Methods

Aldo Russo, C. De Vitiis, P.D. Falorsi, and S. Falorsi, University of Rome, Italy

This paper deals with the nonresponse adjustment weights that attempt to compensate for unit nonresponse. A commonly used procedure for obtaining these weights is to divide the total sample into a set of weighting classes based on information known for both respondents and nonrespondents, and then to increase the base weights for the respondents in a weighting class to represent the nonrespondents in that class (Kalton 1983).

In some surveys there is an extensive amount of information available for the nonrespondents. This information may be available from the sampling frame or by matching sampled elements with administrative records. Also, in panel surveys and other surveys involving more than one stage of data collection, extensive information on nonrespondents at later stages is available from their responses at the early stages.

The main aim of this study is on methods for developing weighting adjustments for nonresponse when many auxiliary variables are available for respondents and nonrespondents. In this situation, one decision that arises is how to select which of the auxiliary variables should be used for this purpose and another decision involves how the selected variables should be used.

The main ideas are presented in this work by applying several different adjustment procedures at the sampling survey European Community Household Panel (ECHP). The nonrespondents to a ECHP panel can be separated in two groups those who fail to respond at the initial wave of data collection (initial wave nonrespondents) and those who respond at the initial wave but fail to respond at one or more of the subsequent waves of the panel for which they are eligible. For the latter group extensive information from the initial wave of data collection can be utilized in adjusting for panel nonresponse. The weighting adjustments studied here relate to the panel nonrespondents only. These adjustments modify the weights of panel respondents to compensate for the panel nonrespondents.

In addition to selecting auxiliary variables and studying alternative methods of using those variables weighting adjustments for panel nonresponse; these study includes a comparative evaluation of the procedures. The final section of these articles summarizes the results and draws conclusions about the effectiveness of the alternative weighting methods investigated (Rizzo, Kalton, Brick, 1996).

Measuring Attrition in Longitudinal Surveys

Jay Zagorsky and Patricia Rhoton, The Ohio State University, USA

Attrition in longitudinal surveys presents two distinct problems for researchers. First, if large numbers of respondents drop out or attrite from the survey, the remaining respondents may no longer constitute a valid random sample. Second, even if overall response rates are high, attrition can be a problem if a select group of respondents leave. For example, if poorer respondents tend to leave a survey, data on income and assets will over represent wealth. When analyzing longitudinal surveys researchers want to know how closely the survey matches the current population.

Unfortunately, prior research has not specified a method of numerically measuring how a survey matches the general population over time. Since a standard measure is not available, this paper outlines the theory and general form of an error measure and then uses this measure to calculate how the bias changes over time for two very long running surveys; the National Longitudinal Survey (NLS) of Young Women and the NLS of Mature Women.

By comparing each longitudinal survey against matching national information at different points in time, an exact computation of the amount of bias due to attrition is presented. This measure enables researchers to quantify and compare the effects of attrition from one longitudinal survey to another. Quantifying the effects of attrition in these two surveys is important because both surveys have large numbers of respondents have either died or refused to participate in the interview. The Young Women survey began in 1968 with 5,159 respondents but by 1995 survey only interviewed 3,019, or 58.52%, of the first round respondents. Higher death rates have caused in larger amounts of attrition in the Mature Women survey. This survey began in 1967 with 5,083 respondents but by 1995 interviewed only 2,711, or 53.3%, of all respondents. Given this large amount of nonresponse, researchers need to empirically know if the surveys are still accurate measures of their respective cohorts. The error measure quantifies the ramifications of losing these individuals.

The organization of the paper is as follows. First the paper discusses the statistical problems caused by attrition. Then it describes the Mature and Young Women surveys, provides details on completion and attrition rates, and investigates the characteristics of individuals who have attrited over time. A theoretical model is created to measure the amount of error or bias introduced by attrition. This measure is then calculated for both NLS surveys by comparing them to Current Population Survey (CPS) data. Finally, a conclusion summarizes the paper and discusses directions for future research.

An Evaluation of Survey Attrition in the Medical Expenditure Panel Survey

Steven Cohen and Steven Machlin, Agency for Health Care Policy and Research, USA

In panel designs with multiple waves of data collection, the overall survey response rate is a multiplicative function of the wave specific response rates. The 1996 Medical Expenditure Panel Survey (MEPS) Household Component follows this model, requiring five rounds of data collection with the same panel of sampled households, to acquire data on health care use, expenditures, insurance coverage and sources of payment that cover two consecutive calendar years. Starting with the second year of the survey, an overlapping panel design is implemented to produce survey estimates, where data covering the second year of a panel are combined with data from the first year of a new panel. To inform the specification of nonresponse adjustment strategies in MEPS to correct for survey attrition, this study identifies the characteristics that distinguish survey participants across all waves from those that only participate in initial rounds and then discontinue their survey participation. The investigation also examines the impact of survey attrition of resultant survey estimates by comparing the estimates derived separately from the respective panels. The results provide insights regarding the efficacy of the MEPS nonresponse adjustment strategies by comparing the survey estimates that characterize the second year of the longitudinal panel with those from a new panel for the same time period.

Sunday, 10:30-12:00

Data Collection Issues

Cooperation and Refusal in Interviews

Karl-Heinz Reuband, Heinrich Heine Universitat, Germany

In general it is difficult to interview those who usually do not cooperate in interviews. If they are hard core refusers it is practically impossible - except if one encourages interviewers to contact relatives and acquaintances. They might usually refuse the cooperation in interviews but might react otherwise when social obligations of cooperation exist. Such kind of obligations evolve when the interviewer stands in a close relationship to the potential respondents. The refusal rates in such cases are usually very low. The study is based on a quota sample of the population in Düsseldorf, Germany, with social science students as interviewers and about 500 respondents. In the article we pursue the following questions: how widespread in the population is the experience of interviewing and non-cooperation, what are the social characteristics of those and which are their attitudes and knowledge about political affairs? To what extent does a desirability bias governs the responses?

Item Nonresponse in Mail Surveys of Establishments

Fran Featherston and Carl Ramirez, U.S. General Accounting Office, USA

While many studies investigate the sources of unit nonresponse in self-administered establishment surveys, less is known about the specific characteristics of item and questionnaire design that contribute to item nonresponse. This paper builds on earlier work using a large "item bank" to codify and evaluate key item and questionnaire design characteristics that may be related to the rate at which unit responders fail to answer individual survey items. In addition, characteristics unique to establishment survey requests are incorporated into the resulting model: for example, whether questions request factual organizational data, policy statements attributable to the organization, or personal attitudes and opinions of the individual respondent may affect response. Our preliminary findings indicate that readability (item format, question length, use of skip patterns) is the most important factor in designing high-response items for establishment surveys.

Nonresponse Errors in Environmental Surveys: Obtaining Access to Private Property

Virginia M. Lesser, Oregon State University, USA

With increasing environmental awareness, more emphasis on surveys to assess changing environmental condition is expected. In 1990, the Environmental Protection Agency initiated the Environmental Monitoring and Assessment Program (EMAP) to monitor environmental condition in the United States. The objective of this program is to estimate and monitor changing conditions of a number of ecological resources, such as prairie wetlands, lakes, and streams. It is critical to obtain access to EMAP sampling sites on repeated occasions in order to obtain biological measures of environmental change. However, since many sampling sites are located on private property, landowner permission is needed to gain access. A comparison of landowner permission rates achieved by asking for site access by mail versus telephone is summarized. The extent and impact of reduced access to EMAP sampling sites are presented. The discussion also compares site access in EMAP with other environmental studies requiring landowner permission. These studies include surveys conducted by the US Department of Agriculture, the June Enumerative Survey run by the National Agricultural Statistics Service and the National Resources Conservation Service's National Resource Inventory. Finally, some design- and model-based remedies to adjust for nonresponse in EMAP are listed.