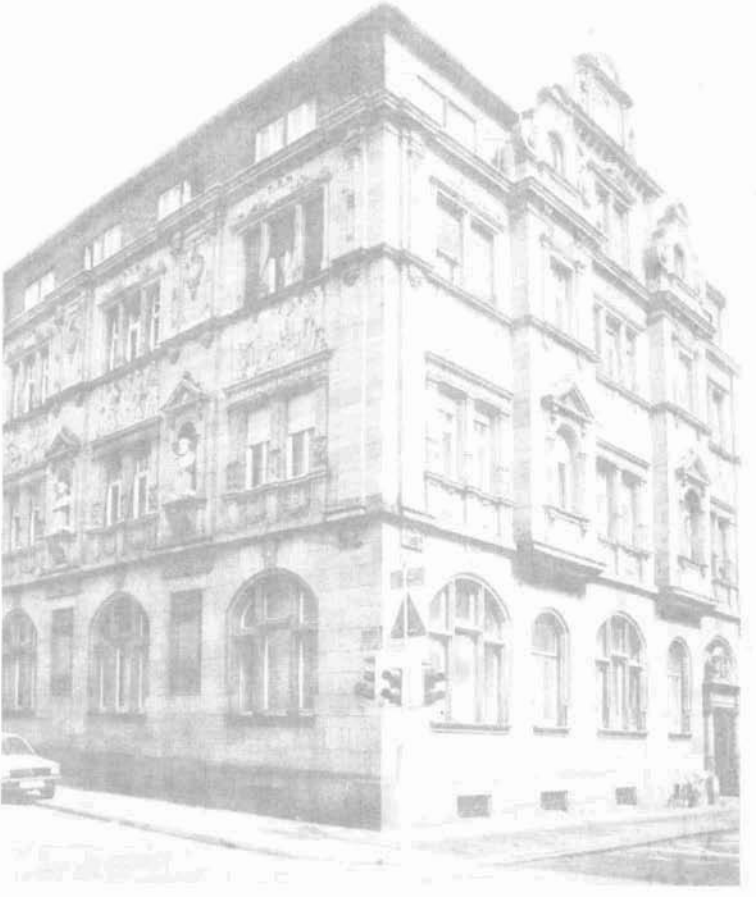
NACHRICHTEN SPEZIAL

ZUMA
Zentrum für Umfragen, Methoden und Analysen

No. 4



NONRESPONSE IN SURVEY RESEARCH

NONRESPONSE IN SURVEY RESEARCH

Proceedings of the Eighth International
Workshop on Household Survey Nonresponse
24-26 September 1997

Edited by

Achim Koch and Rolf Porst

Editors: Achim Koch and Rolf Porst

Publisher: Zentrum für Umfragen, Methoden und Analysen (ZUMA)

ZUMA is a member of the Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS)

ZUMA Board Chair: Prof. Dr. Max Kaase
Director: Prof. Dr. Peter Ph. Mohler
P.O. Box 12 21 55
D - 68072 Mannheim
Germany

Phone: +49-621-1246-0
Fax: +49-621-1246-100

Internet: http://www.social-science-gesis.de/

Printed by Druck & Kopie hanel, Mannheim

ISBN 3-924220-15-8

Contents

Preface and Acknowledgements

This volume, the fourth in the ZUMA-Nachrichten Spezial series on methodological issues in empirical social science research, takes up issues of nonresponse. Nonresponse, that is, the failure to obtain measurements from all targeted members of a survey sample, is a problem which confronts many survey organizations in different parts of the world. The papers in this volume discuss nonresponse from different perspectives: they describe efforts undertaken for individual surveys and procedures employed in different countries to deal with nonresponse, analyses of the role of interviewers, the use of advance letters, incentives, etc. to reduce nonresponse rates, analyses of the correlates and consequences of nonresponse, and descriptions of post-survey statistical adjustments to compensate for nonresponse.

All the contributions are based on presentations made at the '8th International Workshop on Household Survey Nonresponse'. The workshop took place from September 24 - 26, 1997 in Mannheim, Germany, the home base of the workshop host institute, ZUMA. The international workshop series on nonresponse was initiated in 1990 by Robert Groves, University of Michigan, Bob Barnes, Office of National Statistics, UK and Lars Lyberg, Statistics Sweden. Since then, the workshop has been held each year in a different country. The purpose of the workshop is to gather researchers and methodologists from statistical agencies, research institutes and universities interested and active in the nonresponse field, so that exchange of ideas and collaborations across nations and different type of organizations can take place in an informal setting. Former hosts have been Statistics Sweden, US Bureau of the Census, Statistics Netherlands, Office of National Statistics (UK), Statistics Canada, Statistics Finland and Istituto Nazionale di Statistica (Italy).

The Mannheim workshop drew 51 delegates from nine European countries, as well as Canada, Israel and the United States. Twenty-nine papers were presented and discussed, of which twenty-five are included here. We would like to take this opportunity to thank all the participants of the workshop, in particular the presenters and the session chairpersons. We are especially grateful to Lilli Japec, Statistics Sweden, for her assistance in organizing the workshop. We also wish to thank Christa von Briel and Maria Kreppe-Aygün, ZUMA, for their support in planning the workshop as well as their patience in helping produce this volume.

Mannheim, August 1998

Achim Koch, Rolf Porst

Editors

Current Issues in Household Survey Nonresponse at Statistics Canada

LARRY SWAIN AND DAVID DOLSON

Abstract: *A variety of current and recent initiatives and studies related to unit nonresponse for both cross-sectional and longitudinal household surveys at Statistics Canada is presented. Surveys discussed include the Labour Force Survey (LFS), supplementary and longitudinal surveys based on the LFS frame, the General Social Survey, and several household surveys carried out as special surveys. Factors influencing response and initiatives taken to maintain and maximize response are then reviewed. The next sections present adjustment techniques and a summary of the major themes discussed in the paper.*

Keywords: *response rates, maintaining and maximizing response, nonresponse adjustment*

1 Introduction

Despite the best efforts of survey managers and operations staff to maximize response, some degree of nonresponse is virtually certain to occur in household surveys. Nonresponse has two effects on survey data: one contributing to an increase in the sampling variance of estimates as the effective sample size is reduced from that originally sought; the other contributing to bias of estimates if nonrespondents differ from respondents in the characteristics measured. Survey budgets and potential nonresponse bias influence the decisions made about the degree of nonresponse that is tolerable and the depth of research into adjustments to survey data that compensate for nonresponse. Nonresponse is monitored for feedback to survey staff for immediate and future action and is reported to users of the survey data as part of the overall considerations of data quality.

This paper is intended to take stock of recent and current events in household survey nonresponse at Statistics Canada in order to determine underlying themes. It is not intended to be a complete overview of all activities but is to be topical. The emphasis is on unit nonresponse, and not item nonresponse. This paper is an abbreviated version of that presented at the workshop and of the slightly modified version of the workshop paper noted below in the references. In the longer version, a more extensive overview of the various household surveys at Statistics Canada and their response rates is presented, along

with an examination of monitoring and reporting of nonresponse, remarks on Statistics Canada's nonresponse databases and nonresponse standards, and an extensive list of references.

2 Maintaining and maximizing response

2.1 Factors influencing response

Like many national statistical agencies, Statistics Canada conducts a wide variety of household surveys. Some are part of the regular ongoing program while others conducted on an occasional or one-time basis are known as special surveys. The major example in the former group is the Canadian Labour Force Survey (LFS) which achieves response rates of about 95%. Supplemental surveys can be annexed to the LFS. Such surveys include the Family Expenditure Survey with response at about 70% to 80% and the Survey of Consumer Finances (75% to 82%). Longitudinal surveys drawn from the LFS frame have achieved longitudinal response rates in their second wave of 91% to 94%. The General Social Survey, a telephone survey implemented using random digit dialling techniques, achieves response rates generally between 80% and 85%. Most special survey supplements to the LFS have response rates in excess of 85% while under normal circumstances, response rates for other special surveys are 70% or more.

Within available budgets and time frames, survey managers strive to maintain or improve response rates as much as possible. There are, however, many factors which influence response to a particular survey. The survey budget, the importance of the survey to the users and to the respondents, the relative priorities of the allocation of budget to the various survey operations, the content of the survey, the survey population, the method of collection, the season and time of collection, the length and complexity of the survey questionnaire, the degree of follow-up and field quality assurance all play important roles in the response to a survey. Major factors on which Statistics Canada has recently exerted special efforts are presented below and include language of the questionnaire; the cultural background of the respondent; the tracing of respondents who have moved; the use of technology; the occurrence of survey redesign; concerns of privacy; respondent burden; the communications strategy; several factors related to the interviewers themselves such as training, experience, interpersonal skills, rapport building and turnover; various aspects related to collection; and incentives to respondents.

2.2 Language and culture

Canadian society has been changing significantly over the most recent several years. Canada has always been a country to which many immigrants have come. However, more and more immigrants to Canada speak neither official language well on their arrival to the country, making their response to surveys difficult at best unless suitable interviewers are available. Many come from repressive regimes where governments and their agencies are not well trusted and so are often very reluctant respondents. This has given rise to many new challenges to collection staff, particularly in the large cities in which many immigrants settle. Collection managers need to ensure they have interviewers available who can communicate in languages other than French and English. Interviewers need to be equipped to understand the concerns of recent immigrants and to persuade their response to our surveys.

A recent pilot test of a longitudinal survey of immigrants to Canada specifically targeted this population. A sample of immigrants, initially identified upon landing in Canada, was later followed up for administration of the survey questionnaire. Questionnaires and other materials were prepared in a number of other languages like Cantonese, Mandarin, Punjabi and Spanish. Interviewer training included a module on cultural sensitivity. Interviewer assignments were constructed so as to match interviewer language abilities to those of respondents. These actions as well as a careful communications strategy helped make a big difference in the success of the test.

Part of a recent test of redesigned collection procedures for the International Travel Survey (ITS) included the use of questionnaires in Japanese and the use of Japanese speaking interviewers. Results are currently being analysed. In addition to the two official languages, questionnaires for the National Population Health Survey have been translated into Spanish, Portuguese, Chinese, Punjabee and Italian to try to reduce non-interviews because of language problems. When administered on paper, the Labour Force Survey questionnaire was available in several languages; now with conversion of the LFS questionnaire to computer-assisted interviewing, questions are available in paper form in many languages.

2.3 Tracing

Tracing is an integral part of maintaining response for longitudinal surveys and various means are used by the eight Statistics Canada regional offices to find the longitudinal respondents who have moved. These include using the contact information provided earlier by the respondent and, where received, the change of address card sent by a respondent who moved; contacting the former household in cases where only the

respondent has moved; contacting the postmaster in small communities; searching commercial CD-ROMs containing names, addresses, postal codes and telephone numbers; using telephone directories; contacting neighbours or apartment superintendents for a forwarding address; searching for relatives with the same last name and postal code; contacting municipal city hall tax departments to determine the owner of the building the respondent lived in before moving; using city directories, telephone directory assistance and the internet. Interviewers and regional office tracers have demonstrated initiative and thoroughness in trying to find respondents. On occasion, they have used resources such as alumni associations of universities, government employee telephone books, the Association for New Canadians, local pharmacies in small towns, the respondent's last employer, real estate agents, military base commanders, local utility companies and local schools.

More specifically, for the National Population Health Survey (NPHS), the Survey of Labour and Income Dynamics (SLID) and the National Longitudinal Survey of Children and Youth (NLSCY), one or two supplementary contact names were obtained to aid in tracing. For the NPHS, however, one contact was for the longitudinal respondent and the other for the overall household. For SLID, the contact had to be from outside the household. In addition for NPHS, a summer student worked at head office after the second wave to try to find hitherto untraced respondents for the next wave of collection. Also at head office, for each of NPHS, SLID and NLSCY, a coordinated record linkage was done with income tax files to trace address and telephone number changes. For NPHS, an additional linkage with mortality files helped determine those respondents who had died.

For all these longitudinal surveys, in addition to tracing initiated by the interviewers themselves, tracing teams were set up temporarily in Statistics Canada's regional offices in response to the needs of the longitudinal surveys. There are, however, no standard procedures for the tracers to follow. Although tracing techniques are similar from survey to survey, they do differ by regional office. In order to identify the effectiveness of current techniques and resources and to improve those procedures, the current use of resources for tracing is being evaluated and standards for tracing are being developed for Statistics Canada's regional offices. Although efforts are being made to improve tracing, there has been good success to date: for the second wave of the NPHS, the proportion of longitudinal respondents that were unable to be traced was only 1.7%; for the second wave of the NLSCY, preliminary data indicate a similar rate.

Clearly, tracing is important not only for longitudinal surveys but also for other surveys such as those based on aging list frames or surveys in which previously contacted persons or households are contacted a second time. An example of both of these is the Survey of

1995 Graduates conducted in spring 1997. The contact information received from registrars was, on average, two years out of date. Even before sending out the sample to regional offices for data collection, some initial tracing was done at head office to minimize that which would be necessary for interviewers. There was about a twelve percent loss due to tracing failure. Since it is planned to contact respondents again in about two years, additional contact information was collected during the interview. Previous Follow-up of Graduates surveys have had tracing losses of about ten percent.

2.4 Technology

Technology also plays a role in response. When the Labour Force Survey (LFS) was converted from a paper questionnaire to the use of computer-assisted interviewing (CAI) in late 1993, technical problems created a new category of nonresponse. This category replaced one which had earlier represented questionnaires received too late for processing because of postal problems. Technical problems occurred primarily in the transmission of data from the interviewer's laptop computer to the regional office because of situations such as hard disk crash, insufficient memory allocation, excessive heat, power outages or telephone troubles. In most cases specific problems were resolved for the next month but with very tight release deadlines for LFS data, were unable to be done in time for the current month. Although initially reaching highs of 1.7% as a component of nonresponse, technical problems are now well under control as the experience with CAI increased. Familiarity of interviewers with the hardware and software, software updates and a streamlined communication process between interviewers and technical support staff have also contributed to the virtual non-existence of nonresponse resulting from technical problems. Although initially the LFS managers had been concerned about a possible increase in refusal rates because of the presence of a notebook computer at the first interview, no increase was detected.

Other technical innovations are also taking place. The regional offices are initiating a complete revamp of their case management software used for the LFS, its supplements, and other surveys. The current case management system was first put in place in 1993. From a systems development perspective, it needs replacing because the combination of its design limitations and the various patches and updates so far make further development very difficult. From a user perspective, the system needs to be more robust. It needs to provide easy and reliable movement of workloads between interviewers working via CATI and CAPI at possibly widely distant geographic locations. It needs to provide much improved management information especially for supplements to the LFS. Its call scheduling capability may be upgraded as well. These improvements are scheduled for implementation by summer 1999.

Improvements have also been made to call scheduling methods for telephone surveys conducted at the centralized telephone survey facility. This has facilitated a more effective use of interviewing capacity and resulted in higher response rates. This was first done for the Sun Exposure Survey, a recent random digit dialling survey. The status (probably residential, probably business, unknown) of sampled telephone numbers was used for the scheduling of call attempts throughout the first several days of the interviewing period. Residential numbers were called between 16:30 and 21:00, business numbers between 09:00 and 16:00, and unknown status numbers between 09:00 and 18:00 respondent's local time. Calls were made at the time of day when it was felt that contact was most likely to be made. Collection managers and interviewers were very happy with this change. Managers were well able to assess progress much earlier in the collection period because a higher proportion of the sample had been called at least once and because more of the sample had been resolved (i.e., determined to be not working, business, respondent, etc.). Interviewers' time was being used much more effectively.

In the future, there are plans to put in place more sophisticated call scheduling features taking account of the call history of sampled telephone numbers.

Telephone technology available to residential users has made significant advances in recent years. Not only are telephone answering machines becoming more and more common, but also features such as voice mail systems, call display, call blocking, and call forwarding have become available to consumers. Research has shown that the use of answering machines for screening of calls did not appear to be widespread. Later research found that even though about 40% of households with answering machines screened at least most of the time, this did not necessarily translate into a major problem for refusals and those unable to be contacted. Statistics Canada has not formally studied the issue. However, anecdotally, these telephone technologies are not yet presenting a major problem. Collection staff have put in place special procedures to deal with them and good response rates are being maintained. (Many respondents even return messages left by interviewers on their answering machines!)

2.5 Redesign

The phase-in of a new LFS sample resulting from a redesign of the sampling plan resulted in temporary increases in nonresponse. The new design allocated a greater proportion of the sample to urban areas which generally have a higher nonresponse rate than rural areas. In addition, some of the new sample's assignments were given to new interviewers who typically have higher nonresponse rates than more experienced interviewers.

As a last step in the survey redesign, the LFS questionnaire and data processing systems were overhauled. Phase-in of the new questionnaire took place between September 1996 and January 1997; in general the new questionnaire was applied to each new rotation group as it entered the survey for the first time so that by February 1997, all six rotation groups were receiving the new questionnaire. Comparisons during the phase-in period indicated that the national nonresponse rates for the new questionnaire were higher than for the old questionnaire. However, after accounting for the birth rotation group being the group most subject to nonresponse, no discernible differences in response rates could be attributed to the introduction of the questionnaire per se. It had originally been hypothesized that a new question on workers' earnings would have an effect on household nonresponse.

2.6 Privacy

Increasingly, Canadian society has become sensitive regarding personal security, privacy and confidentiality matters. As well, there appears to be much more household survey taking activity in Canada than formerly, not only by Statistics Canada (e.g., major new longitudinal surveys and a much increased volume of surveys for clients on a cost recovery basis) but also by private sector polling firms. Further, telemarketing is an ever present and growing irritant to the population. These factors have combined to make respondents less trusting of telephone inquiries, and make it more difficult for interviewers to convince that they are conducting legitimate survey research.

A recent example illustrating this phenomenon is a test of a survey on Unregulated Child Care Providers, conducted as a supplement to the LFS. In addition to other information, each LFS respondent with children was asked if he/she used the services of an unregulated child care provider and, if so, to provide the name and telephone number so that the care provider could also be contacted. The response rate by parents to the screening questions was an excellent 92%. However, of those, only 64% would give the provider contact information. Finally, only 62% of the providers identified would respond. These lower response rates are likely indicative of respondents' concerns regarding privacy and confidentiality in this situation.

A related issue is that of data sharing. Canada's Statistics Act requires that Statistics Canada protect the confidentiality of its respondents' data. Normally, a master data file with complete detail is prepared for each survey; such files will often contain data that could facilitate identification of individual respondents. In addition, a public use micro data file is also prepared in which appropriate actions have been taken so that individual respondents cannot be identified. Occasionally, respondents are asked to sign an

agreement that their detailed data be shared with specified third parties (usually survey clients) who in turn are then bound by the requirements of the Statistics Act. Typically about ten percent of respondents will refuse to agree. But when interviewers are well prepared and where the survey plans provide for the necessary effort, this can be reduced to one or two percent.

2.7 Respondent burden

The supplementary survey capacity of the LFS provides a microcosm of the increased household survey taking in Canada. This facility has been available for many years now and offers clients access to all the data collected for the LFS itself and use of the LFS infrastructure including its collection staff. This supplemental capacity has been very popular both for internal and external clients who have realized the substantial cost and quality benefits as compared to independent surveys.

The LFS infrastructure is also used in two other manners. Households which have rotated out of the LFS sample can be used as sample for other surveys for which active LFS households are not available on a sufficiently timely basis or which are considered too burdensome or too sensitive to be conducted on households still in the LFS sample. Two-phase methods can be applied for both supplements and for surveys of rotate-outs to target a wide variety of subpopulations. Secondly, the LFS frame can be used for selection of independent samples. In both cases the LFS design, its processing systems, and usually its interviewers are used.

Surveys on rotate-outs realize the benefit of being warm contacts. However, in cases where the sampling unit is the household or a person in the household rather than the dwelling (and the people who happen to be living there), some reduction in response rate is typically observed due to tracing losses related to moving of respondents. This reduction is less severe than what might otherwise be, since telephone numbers are available on these rotate-outs from the LFS collection. Surveys based on independent samples are of course cold contacts and more effort is required to obtain similar levels of response.

This supplementary capacity is now being used much more heavily than before. Years ago, the LFS would have a supplemental survey in most months; now it is every month. In some months, respondents may receive two or three (smaller) supplements. Compared to the past, rotate-outs are being used for surveys more often and sooner after completion of their six month time in the LFS sample. (Rotate-outs used to be given at least about a six month rest before any subsequent use.) These factors are contributing to respondent fatigue and resistance as well as to demands on the interviewing staff.

In addition, there are societal factors that may be contributing to respondents' sensitivity to burden. For example, in more Canadian families than ever, both parents are employed. There are more single parent families. These factors and others have contributed to a common belief that for many Canadians, time has become a scarce commodity, and that consequently it is becoming more and more difficult to convince respondents to use their scarce time to reply to surveys.

As well, the nature of the surveys being done has gradually changed over, say, the last twenty years. Early in this period, supplemental surveys were typically fairly brief and simple - one to two pages. More recently, especially with the advent of computer assisted interviewing (CAI), much longer interviews are being undertaken. Surveys with interviews of 30 minutes to an hour are becoming common. They are more complex not only in terms of the flows and skips in the instruments - facilitated by CAI - but also in terms of the concepts being addressed and the detail of information requested. More sensitive topics are being covered. Two closely related consequences of these factors have a direct impact for nonresponse rates. Both by their length and their complexity, these are more burdensome interviews for respondents and they may be less willing to respond to such demanding surveys. As well, these surveys are much more demanding of our interviewers to understand their requirements and to be able to persuade response to sometimes long and difficult interviews.

A new challenge in recent years has been surveys of children. For example, the Youth Smoking Survey interviewed children aged 10 to 19 years concerning their smoking habits. The National Longitudinal Survey of Children and Youth (NLSCY) has interviewed children as young as ten and has administered tests to children as young as four. In these and other like surveys, special attention is devoted to communications - permission must be obtained from parents to interview their children and interview materials for children must be developed so that they require suitable levels of literacy. For older children, innovative means of eliciting response are needed.

Managers of supplements to the LFS consistently feel that these challenges have been met reasonably well, with only fairly minor deterioration in response rates to the supplements and no deterioration in response rates for the LFS. They also note that it has become much more difficult than in the past to achieve these results. Survey managers feel that a kind of threshold in the use of the supplementary capacity has been reached and that the burden cannot be increased any more without substantial risk of increasing nonresponse and decreasing data quality for the supplements due to respondent resistance, interviewer fatigue and possibly to resulting interviewer turnover. There are tradeoffs associated with possible solutions such as the selection of an independent sample or the creation of a separate interviewer workforce. Further, it is critical that the supplements to the LFS

continue not to have any negative impact for LFS data quality. Although there are other factors, effective communication with respondents and others including basic interviewer training and survey specific training regarding making respondent contact and getting their cooperation has become much more important than ever before, not only for LFS supplements but for all household surveys.

2.8 Communications

Concerns about privacy and respondent burden have led survey managers to be far more thorough with the communications strategies for their surveys. A number of recent project teams have included representatives from Statistics Canada's Communications Division to specifically address these concerns. More attention is being paid to ensuring that public enquiry officers, collection managers, and interviewers are well informed and prepared to respond to questions concerning surveys. Special interest groups related to subject matter of surveys are being more often consulted to solicit their support for our surveys. Advance letters to respondents are being used more often. Not so long ago, survey managers at Statistics Canada rightly tended to believe they could get sufficiently high response rates (say 80% or more) without advance letters.

The Sun Exposure Survey, for the first time at Statistics Canada, used an advance letter in the context of a random digit dialling (RDD) survey. This was done as a test with half the sample assigned to the letter group. Statistics Canada's RDD frame is constructed on the basis of telephone company billing files, and so sampled telephone numbers can be classified as being probably residential, probably business, or of unknown status. The probably residential numbers in the letter group were matched to a commercially available CD-ROM of telephone numbers, names and addresses in Canada. Those having good quality matches with complete name and address, including postal code, were sent an advance letter shortly before the collection period began. All letters were personalized by using the name, and since one respondent per household was to be selected, the recipient was encouraged to show the letter to other members of the household. The response rate for the telephone numbers to whose matching name and address a letter was sent was 75%, about nine percentage points higher than for the corresponding probably residential numbers (ones with good name and address matches) from the no letter group. The difference was due to reduced refusals. This improvement is consistent with that observed by others. Interviewer debriefing suggests that the letter had the effects of both legitimizing the survey in the minds of respondents as well as facilitating better performance by interviewers. (Interviewers knew which telephone numbers corresponded to households which had been sent an advance letter.)

While clearly successful, there are problems to be overcome. A more effective and efficient means of finding good quality and up to date names and addresses is needed. Although its cost was modest, the CD-ROM telephone directory is based upon published telephone directories and hence contains information that is 12 to 18 months out of date. As well, a significant impediment was the amount of manual work required to resolve multiple close matches. A consequence was an 18% non-match rate of telephone numbers to the CD-ROM and a further loss of 11% due to post office returns (PORs) so that at most 71% of the probably residential numbers in the letter sample actually received a letter. As well, it is likely that some of the people sent advance letters no longer had the telephone numbers to which the address matched; not all of these would have been PORs. So whoever actually had such telephone numbers would never have received (or even been sent) an advance letter.

To motivate their continued cooperation, longitudinal respondents of the National Population Health Survey were sent a thank you letter after the first wave of collection and were encouraged to send a change of address card to Statistics Canada should they move. Later, they were sent an advance letter before the second wave appreciating their earlier response and soliciting their participation for the second wave. As well, they were given a brochure containing a summary of the first wave survey results. After the second wave, they will be sent a calendar as a reminder of the upcoming third wave. To encourage response to the National Longitudinal Survey of Children and Youth, activity books for children were provided by interviewers in the first wave of collection. In the second wave, interviewers gave out calendars and showed newspaper clippings of coverage of the first wave to respondents. For the future third wave, newsletters are being considered for both children aged 10 and above and their parents.

For the Survey of Labour and Income Dynamics, a newsletter about the survey, encouragement to respond and change of address card are sent to respondents before collection. After collection, respondents receive a letter of appreciation and a change of address card. A great deal of effort is expended in designing respondent material. Research through the use of focus groups has led to a distinction by region in the content of the material prepared for SLID. Further research addressing the appropriateness, content and need for respondent-specific material is being considered for the Assets and Debt Survey. In fact, the question of what keeps individuals responding needs exploration as the longitudinal surveys age.

As part of Canada's 1996 Census of Population, a new data collection methodology, dubbed "centralized edit", was tested in a selected geographic area of the country. In testing the mail-out of Census questionnaires where possible, it was anticipated that there could be a negative impact on the mail return rate owing to the switch from the traditional

drop-off of questionnaires to mail-out as a mode of delivery. Thus, the inclusion of a mandatory appeal message on the mail-out envelope was examined using an interpenetrating sample to one-half the households in the Census test. It was found that the message, "By law, you must complete and return the enclosed census questionnaire", had the effect of an increase of 2.5 percentage points in the return rate of questionnaires from 80.7% to 83.2%. The increase was similar for both long and short Census questionnaires. As well, with the message, questionnaires tended to be completed and mailed back earlier although this stabilized after about one week.

Statistics Canada's Self-Sufficiency Project (SSP) has obtained excellent response rates by use of a number of strategies well suited to longitudinal surveys. It is a longitudinal study of income assistance (IA) recipients with the goal of assessing the efficacy of the offer of a temporary earnings supplement to help recipients make the transition to economic self-sufficiency. The sample, selected from provincial IA files, was partitioned into control and program groups. Persons in the latter group were offered a temporary earnings supplement if they could find a job within a specified length of time after entry into the study. After a baseline survey, follow-up waves took place after 18, 36 and 54 months. Sample intake took place over several months and, consequently, so did interviewing for each wave.

The 36 month follow-up required interviewing of children, parental consent for this, and the parent interview. Given the longer and more complex interviewing, an advance letter and more in-depth interviewer training were again used. Two other innovations were also used. A series of conference calls involving survey managers and interviewers was initiated to discuss challenges and techniques to obtain parental consent and child participation. This was found to be very helpful. As well, a small incentive of a twelve page activity book was given to each young child in the house at the time of the interview. Finally, taking a lesson from SLID, a newsletter will be sent to each study participant two to three months prior to contact for the 54 month interview.

2.9 Interviewers

Moving now from communications with respondents to that with interviewers, the SSP has benefited from the use of a small group of interviewers working only on this survey. Their training was lengthy and thorough, including special emphasis on obtaining informed consent from study participants. This was particularly important since survey data was to be linked to administrative data from three different sources and then shared with a third party involved in the research. Prior to contact for the baseline survey, at which the informed consent was obtained, study participants received an advance letter

from their provincial Ministry of Social Services explaining why the study was being done and its potential benefits. Finally, survey questionnaires were made available in a set of four common non-official languages. These strategies combined to yield a response rate of 92%.

Although the General Social Survey has been conducted in annual cycles, its data collection is spread out across all months of the year. Thus like the SSP, it has been able to hire a small interviewing staff dedicated to just the one survey. This has helped keep its nonresponse rate low and has likely helped reduce other sources of nonsampling error.

In some cases, collection is done through third parties, usually only when survey requirements are such that there really is no other choice. Staff who are not professional interviewers sometimes need to be asked to conduct (hopefully brief) interviews or distribute questionnaires on behalf of Statistics Canada on top of their regular responsibilities. Immediately, there is a perception of reduced importance of the survey, particularly if supervising staff do not feel a sense of ownership and interest in the research objectives. Without a great deal of special care, results are often less than satisfactory.

A recent pilot test sponsored by the research branch of a federal government department required some collection work to be done by the operational staff of another branch of the same department. Persons selected at this stage would become the sample for a longitudinal survey. Realizing the inherent challenge in this, the survey's project team took a number of communications initiatives to address it. Managers of the collection staff were initially contacted by telephone and sent basic information by mail. On the first day of collection, a Statistics Canada representative briefed the collection staff, their supervisors, and their managers on the purpose of the survey, the procedures, the importance of following them exactly, and other related matters. Each collection officer was provided with a copy of the procedures and an explanatory letter from a senior manager in the branch. Copies were posted at work stations and in common areas. Despite the best intentions, these efforts were still not enough. The sample size ended up being much smaller than expected primarily because procedures were frequently not followed exactly, to the extent that collection did not take place at all during some shifts or during busy periods. In some cases where a Statistics Canada representative obtained permission to be present during collection (and this was difficult), procedures were adhered to more closely, clearly indicating the desirability of ongoing monitoring by the actual survey taking organization. It is vital that such collection activities have the support of the collection staff and their managers.

Data collection for the large majority of Statistics Canada's surveys is done through its network of regional offices spread out across the country. Results are usually excellent. These are organizations with long histories in survey taking with very experienced staff and an outstanding culture in which doing a top notch job, including minimal nonresponse, is an integral part.

Over the last two or three years, data collection for a number of special surveys has been conducted through a new centralized telephone interviewing facility located at head office in Ottawa. These have all been conducted via CATI and some have been by RDD. This facility has been successful in getting these surveys up and running quickly and at reasonable cost, in fact more quickly and more cheaply than would have been the case through the regional offices. There are a number of reasons for this and a number of consequences that are illuminating.

As a centralized and smaller organizational unit, the centralized telephone interviewing facility is more flexible in its ability to respond rapidly to client needs. Many management activities are simplified and logistics of project implementation are easier. Questionnaire instruments can be reliably developed more quickly perhaps because the unit has fewer surveys active at any given time and hence its programming staff have fewer competing priorities. Development of training materials is done more quickly. Implementation of interviewer training is certainly easier and cheaper as fewer trainers are needed; in fact, the trainer is typically the same person who develops the training materials.

Staffing of interviewer positions, or "CATI clerks" as they are called, has been awkward. The volume of interviewing business for this group has not yet been large enough to make hiring a permanent interviewing staff viable. Consequently, they have been hired on a temporary or term basis.

The strong culture that exists in the regional offices does not yet exist in this relatively new group. This is made difficult also because turnover of CATI clerks has been comparatively quick. Thus, they are also a less experienced staff, not as adept at the various means used to persuade respondents. For example, two recent RDD surveys, one on electronic media usage and another a pilot test of the Sun Exposure Survey, had response rates of just under 50%. By contrast, Canada's Alcohol and Other Drug Survey - a RDD survey conducted through the regional offices in 1994 - achieved a response rate of 76%.

Further illustration of the point is provided via the subsequent Sun Exposure Survey, also conducted by the centralized facility, which got a response rate of 70% after making a number of changes including the advance letter test described earlier, a simple change to the call scheduler described above and some changes to the training for the CATI clerk staff.

Most of the staff had previously worked on both the Electronic Media Usage Survey and the Sun Exposure pilot test as well as on another survey in the meantime, so they had become somewhat more experienced. Their training was upgraded from that used in the pilot test. Additional emphasis was placed on basic interviewer skills. A new module on refusal avoidance and conversion was developed. Based in part on the interviewers' own comments, the background information on the survey, including questions and answers useful to reply to respondents' questions and to help persuade them to comply was lengthened and made easier to use. These efforts facilitated a decline in the refusal rate from 25% in the pilot to 19% (about 13% for those sent an advance letter and about 21% for the remainder). This remains high, but further reductions can likely be achieved with a still more experienced and better trained staff.

There are some indications that some nonresponse to interviews done from the central location may result from the call coming from Ottawa and not from more local origins. It is suspected that public mood does affect respondent relations, especially when the relationship is with the federal government.

2.10 Collection

In order to maximize response to the longitudinal surveys on health (NPHS) and on children and youth (NLSCY), a supplementary collection period was added in June of 1995 in which unresolved cases were addressed. In fact, the NPHS has made this additional collection period a permanent feature of the survey and has dubbed it "quarter 5". Response rates were increased by two percentage points through this follow-up for NPHS in each of the 1994 and 1996 surveys with over one-third of all unresolved cases being converted to full or partial response.

One of the reasons for adding the supplementary period resulted from restrictions associated with the introduction of CAI. For example, in the first wave of the NPHS core sample, collection for each quarter was restricted to a two-week period in one month outside the Labour Force Survey collection period. This occurred because unresolved cases could not be stored on the laptop computers for later action. Storage has since been made possible and collection has been extended within the quarter and into subsequent quarters. The NLSCY carries over unresolved cases into its second period of collection.

In the vacation months of July and August, the Labour Force Survey extends its collection period with enumeration on Monday and Tuesday of the second week of collection rather than on the Sunday.

The supplementary buy-in RDD samples for the NPHS were done by independent sets of interviewers from the NPHS core sample. Part way through the collection period for these buy-ins, revised sample assignments were introduced to help improve response rates for harder to reach households. In earlier collection periods, the month's sample was assigned to each interviewer at the beginning of the month. At the end of the month, all unresolved records were assigned a final status code and returned to Head Office. This resulted in interviews for easy to contact households being done in the first two-thirds of the month with the hardest to reach ones being left to the final third. The revised procedure provided for a collection period of two months with the sample being assigned at the start of the first month and around the 23rd of the first month. Efforts were made for the sample sent out in the first wave to have a greater likelihood of containing telephone numbers for residential addresses. An advantage of this method is that more time is provided to contact the harder to reach households. In addition, at the end of the first month, instead of working only on the difficult households, a combination of these along with the easy ones from the second wave can be worked on at the same time.

In the second wave of the NLSCY, parental consent forms to administer tests in the child's school and to send questionnaires to the child's teacher and principal were incorporated into CAI as well as maintained on paper. This gave a better control over the consent process and thereby improved response through reduction of lost paper forms or late arrivals.

In general, the use of CAI permits a more rigorous and more frequent monitoring of response rates while field work is underway. Actions can be taken sooner to understand and alleviate problems. In fact, target response rates are set for many surveys and knowledge of shortfalls is more immediate with the use of CAI in both the regional offices and head office.

For SLID, to aid in response, there are two collection periods: one in January for the labour portion of the survey as this month immediately follows the reference year for questions; the other in May for the income part as this is immediately after the deadline for completion of tax forms. In fact, for the income questions, respondents are asked for permission to have tax records accessed rather than have them complete the May income questionnaire.

2.11 Incentives

Although Statistics Canada does not generally offer incentives for survey response, some limited research has been carried out in a test in the 1990 Family Expenditure Survey to evaluate the impact of incentives. The Statistics Canada publication "A Portrait of Canada" and a clipboard bearing the Statistics Canada logo were used as incentives to part of the sample. The incentive was given at the time of introduction of the interviewer and was not conditional on the household's later decision to respond or not. No significant effects of the incentives were found nationally although in some cities response rates were significantly improved. Of course, these results apply specifically to the two incentives used and do not necessarily generalize.

More recently, part of a test of redesigned collection procedures for the International Travel Survey (ITS) included the use of incentives (poster, travel planning guides, interactive CD-ROM). Results are currently being analysed.

Non-monetary incentives are being considered for the upcoming Assets and Debt Survey.

2.12 Other related initiatives

The International Travel Survey (ITS) questionnaires are distributed by Canadian Customs officials to returning Canadians and to international visitors for mail-back to Statistics Canada. Although the ITS is a survey of individuals but not a household survey, it is worth noting some of the efforts underway since the major goal of its current testing is to improve response rates. In addition to the testing of Japanese questionnaires, Japanese speaking interviewers and incentives already noted, the following are being tested or are under consideration for future testing: distribution of questionnaires by Statistics Canada, short entry interviews to obtain key variables to examine nonresponse bias adjustment methods, questionnaires available in other languages, exit surveys, use of administrative data for content or as a sampling frame, and an improved harmonization of data with the domestic travel survey.

2.13 Remarks

Survey managers are finding it harder and harder to maintain acceptably high response rates due to a variety of societal pressures, including a growing collective burden of telephone enquiries from telemarketers and survey taking organizations. They are dealing with this by putting a premium on varied and effective communications strategies with respondents. Communications experts are more often included on project teams. Interviewers are being better trained and equipped to be politely and effectively

persuasive. Advance letters are being used more frequently than in the past, including in new situations, viz. RDD surveys.

All of the above discussion leads to an important issue. What level of nonresponse, or more generally of data quality, is required given the survey objectives? While it is reasonable for an ongoing survey of national importance such as the LFS to spend the budget required to achieve nonresponse rates of five percent - with substantial quality and nonresponse dividends for its supplements - this is often not affordable or perhaps even appropriate for many clients. Reducing or minimizing nonresponse must be taken in the context of some kind of Total Quality Management or fitness for use point of view in which there are many elements of quality to be managed so that the best overall quality is achieved given the resources available. Some examples of other aspects of quality among which there is some element of tradeoff are: making theoretical concepts operational, developing and testing questionnaires that will collect the necessary information in a problem free manner, developing a suitable sampling plan, minimizing assorted sources of non-sampling error other than nonresponse, etc. Survey managers need to take the actions necessary to minimize nonresponse within the constraints of their own data quality standards (Statistics Canada's, in our case) as well as the budget and data quality requirements of clients.

3 Adjusting for nonresponse

Ignorable nonresponse can be compensated through increases in sample size in anticipation of an estimated nonresponse rate. Subsequently, adjustments are made in the weights applied to survey data. However, if nonrespondents differ from respondents in the characteristics measured, then nonresponse will contribute to the bias of estimates. As limited budgets generally preclude extensive research into the measurement of biases, the weight adjustments are generally done in ways that attempt to account for possible biases.

For the Canadian Labour Force Survey (LFS), if a nonrespondent was a respondent in the previous month of the survey, the previous month's data are used for the current month. This accounting for nonresponse applies only if there was a response in the previous month, i.e., data are carried forward only once. All remaining nonresponse is treated through weight adjustments within weighting classes. Weighting classes are defined in such a way that nonrespondents in a weighting class are more likely to have similar characteristics as respondents in the same class. For the LFS, weighting classes contain all households that belong to the same Employment Insurance Region (a subprovincial administrative unit used also for analytical purposes), belong to the same type of area (based on the type of frame from which the sample is drawn, i.e., a combination of

geographic, sample design and urban/rural/remote factors), and have been in the sample the same number of months (i.e., in the same rotation group). The last adjustment using rotation group is new to the LFS and was implemented as part of the most recent sample redesign.

Nonresponse adjustment for supplements to the LFS is usually a fairly straightforward matter. Multiplicative adjustments are calculated using the same nonresponse adjustment groups - sets of strata - as the LFS. Sometimes, due to subsampling for the supplement, it is necessary to collapse strata further for nonresponse adjustment. When the ultimate sampling unit is the person, poststratification to province, age group and sex demographic control totals is also included. Adjustment for nonresponse for other special surveys is usually similarly straightforward. For the General Social Survey (GSS), nonresponse adjustment is done within each geographic stratum and month group.

Canada has a mobile population with about 12% of the population changing address each year. For movers, addresses and telephone numbers on frames can get out of date rapidly, typically resulting in higher nonresponse rates than for non-movers unless considerable time and energy are devoted to tracing. Since these two groups differ in substantive ways, careful nonresponse adjustment is needed to reduce bias.

Exploratory research was carried out on the Survey of Consumer Finances (SCF) in which nonresponse was modelled and a profile of nonrespondents determined. Logistic regression techniques were used to model the probability of nonresponse in the SCF given a response in the LFS (since the SCF is a supplement to the LFS). Nonresponse adjustment factors were calculated on the basis of observed proportions and predicted proportions, along with a poststratification adjustment by age-sex groups. Future work may include examination of the impact on the standard error of survey estimates, the inclusion of earnings data instead of income data in the model, extension from the examination of the conditional SCF nonresponse categories to the unconditional probabilities, and the use of CHAID (Chi-Square Automatic Interaction Detection) techniques to generate nonresponse adjustment classes.

The School Leavers Survey was first conducted in 1991 to collect data on young adults aged 18 to 20 years. Four years later, respondents were contacted again in a follow-up survey to collect more data on their education and work experiences. The response rate in the School Leavers Follow-up Survey was about 66%. An exploratory study was carried out to compare several weight adjustment methods and investigate the impact of revised weights on survey estimates. Logistic regression techniques were used to create response propensity classes with weight adjustments based on response probabilities rather than proportions of respondents.

It is generally known that nonresponse to longitudinal surveys does not occur completely at random and there tends to be differential nonresponse among different subpopulations. As a result, the method of adjustment for nonresponse requires special consideration. Logistic regression techniques may be used to compare estimates using model-based weights with those using standard weighting. Modelling could reduce the nonresponse bias that results from attrition in a longitudinal sample. Its success, however, depends on the availability of good predictors of the nonresponse behaviour.

For SLID, logistic regression was used to identify predictors of response propensity from which weighting classes were created for longitudinal weighting. With the preliminary SLID survey and two subsequent waves, questions naturally arose on which prior wave's data should be used in the model for nonresponse adjustment. For SLID, the decision was made that data for modelling nonresponse would come from the preliminary survey for relatively stable variables (which represent most of the variables) and the most recent wave for the more dynamic variables.

In the NPHS, weighting classes for the 1994-1996 longitudinal file of individuals were determined at the province level using the CHAID algorithm within the software Knowledge Seeker. Geographic and design variables, household level variables and person level variables were considered. For cross-sectional household nonresponse from the 1996 NPHS, the same software was used but resulted in different weighting classes. For cross-sectional individual nonresponse for the 1996 NPHS, weight adjustments were at the provincial level. In the first wave in 1994, adjustment for household nonresponse was by geographic stratum/season combinations while individual nonresponse adjustment was based on province-age-sex classes. In the third wave of the NPHS, imputation may replace nonresponse adjustments in order to create a "square" or complete data set to facilitate analysis using standard packages. For the first wave of the NLSCY, nonresponse adjustment was done by geographic region. For the second wave, research is being carried out to use Knowledge Seeker or logistic regression techniques to determine weighting classes for nonresponse.

More generally, research has shown that commonly used methods of reweighting for nonresponse do not ensure internal consistency of estimates of gross flows. In addressing this issue, a method of Generalized Weight Calibration in Sampling has been considered for nonresponse adjustment for longitudinal surveys.

4 Conclusion

Although the overview provided above contains a wide variety of information and detail, a few main themes arise. Longitudinal surveys bring forth a more complex set of considerations in the adjustment, determination and reporting of nonresponse for both longitudinal and cross-sectional purposes and therefore standards need to be developed. As well, with the longitudinal surveys, information from previous waves or related questionnaires in the household have permitted the use of modelling in the adjustment for nonresponse. The introduction of new technologies has provided new information on nonresponse, more immediate access to data about nonresponse and improved call scheduling to increase the chance of getting response. Societal changes seem to require increased efforts on the part of field staff to maintain the levels of response to which we have become accustomed. Lastly, although more information is now available on nonresponse and nonrespondents, a sufficient understanding of response mechanisms and measurement of the bias component of nonresponse are still lacking.

References

An extensive reference list is included in:

Swain, L. and Dolson, D. (1997). Current Issues in Household Survey Nonresponse at Statistics Canada, Methodology Branch Working Paper HSMD-97-005E, Statistics Canada

Assessment of Efforts to Reduce Nonresponse Bias: 1996 Survey of Income and Program Participation (SIPP)[1]

PRESTON JAY WAITE, VICKI J. HUGGINS AND STEPHEN P. MACK

Abstract: Concern over increasing levels of nonresponse in the 1991-1993 SIPP Panels and new information about the existence of bias in time series estimates of poverty from the SIPP surfaced prior to fielding the newly redesigned SIPP 1996 sample. A tremendous amount of effort and expense has been dedicated by the U.S. Census Bureau to reducing nonresponse and adjusting for its bias. This paper will summarize these efforts and provide a preliminary assessment of the success of the efforts for the first year of the 1996 panel. We will discuss the use of monetary incentives, fielding nonresponse surveys, and changes in field staffing and procedures.

1 Introduction

The SIPP is a complex panel survey conducted by the U.S. Census Bureau to provide information for federal policy makers and academia on topics such as part-year poverty, government program participation and eligibility, health insurance coverage, and income distributions. The SIPP has been used as a multi-purpose survey providing *cross-sectional, longitudinal* and *current event information.* The primary goal of the survey though is a longitudinal one - select a nationally representative sample of households and follow the people in those households to assess changes in their characteristics over time. Quite often, the multi-purpose uses of the data have compromised the longitudinal uses in terms of sample size, data product availability, and important longitudinal analyses.

The 1996 SIPP Panel is the first sample from the new abutting panel design of the survey. The 1984-1993 panels were longitudinal and overlapping - up to 3 panels were in the field simultaneously. Approximately 37,000 sample households will be interviewed every 4 months for about 4 years which will provide analysts with more longitudinal

1 The views expressed are attributed to the authors and do not necessarily reflect the views of the U.S. Bureau of the Census.

observations than the old longitudinal design of 2 2/3 years. A new panel will be introduced every 4 years, e.g., 2000 and 2004. The 1996 panel also includes an oversample of the low income population to enhance poverty analyses.

The change from an overlapping panel to an abutting panel design beginning with the 1996 sample supports the primary goals of the SIPP: producing longitudinal estimates of income and program participation, paying most attention to improving the information for people who are economically at risk, and improving the capability to respond to current policy needs in topical areas. However, the change also inherently exacerbates panel nonresponse issues in the SIPP because overlapping panels will not be available for combined panel analyses which help reduce the level of nonresponse.

The Bureau has conducted a great deal of research on nonresponse issues in the SIPP attempting to 1) assess the differences in the responding and nonresponding universes, 2) estimate the effect of attrition on specific estimates such as monthly mean income amounts, poverty and program participation estimates and 3) investigate alternative imputation and weighting procedures to reduce nonresponse bias.

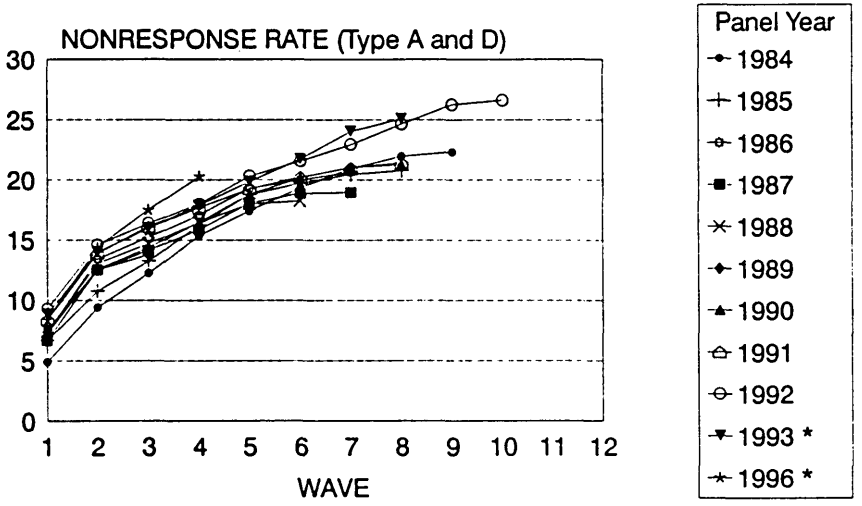
Until recently, there was little evidence that nonresponse bias posed major problems for many important SIPP estimates (Lepkowski et al. 1992, McCormick 1994). Over the past few years though, the Bureau observed some important phenomena in current poverty and low income data series estimates from the SIPP that are cause for concern. Specifically, there is a consistent drop in poverty and some income estimates across panels from the first to second interview that is larger than expected (Huggins and Winters 1995). Also, there is a consistent pattern of decrease in poverty over the life of a panel. These phenomena in and of themselves are troubling, but it becomes even more troubling when the time series carries over from the end of one 4-year panel sample to the beginning of another 4-year panel sample. With the observed decline in poverty estimates over the life of a panel and the higher level reporting at the first interview of a new panel, the jump in the time series resulting from switching to a new panel could be substantial.

Nonresponse rates for the 1984-1990 panels were 5-8% for the first interview and about 21% after eight interviews. Average levels of nonresponse increased in the 1991-1993 panels as compared to earlier panels and the 1995 dress rehearsal first interview response rates were discouragingly low - 88%. The observed bias in poverty statistics described above and the increasing levels of nonresponse since 1990 prompted the Census Bureau to focus even more effort and resources on procedures to reduce nonresponse and improve adjustment methodologies. In the 1996 panel specifically, we

1. *Researched the use of monetary incentives at reducing levels of nonresponse.*
2. *Conducted a nonrespondent study to collect information for improvements in the Wave 1 weighting and to assess whether interview observations may act as good proxy information for weighting. The study was comprised of 2 surveys - one of nonrespondents, one of interviewers.*
3. *Enhanced field procedures for tracking people who move, updated field evaluation procedures, and stepped up the feedback to Field Division on the importance of high response for longitudinal surveys. We also added clerical staff and improved training of interviewers for conversion of refusals.*

Below, we present the picture of nonresponse in SIPP, then we discuss in detail the concerted efforts towards improvement and current results to date as they relate to the success of reducing/adjusting for nonresponse bias in the 1996 SIPP.

Figure 1: Nonresponse rates for the 1984-1996 Sipp Panels



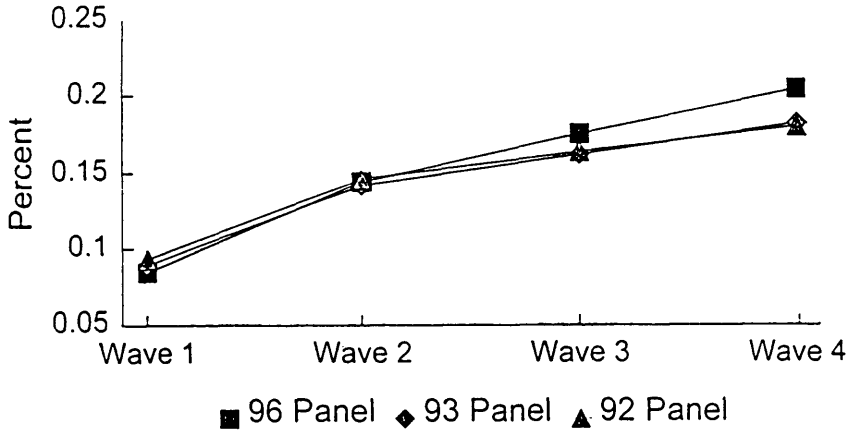
*Preliminary Rates;
Type A: refused/not at home/temporarily absent; Type D: unlocatable mover

2 Patterns of nonresponse and existence of nonresponse bias

As seen in Figure 1, there was a significant increase in nonresponse rates between the 1984 and the 1992 and 1993 panels. When the decision was made to implement a 4 year non-overlapping panel design, we assumed we would reach a nonresponse rate of about 25% after 12 waves of interviews. This in fact became our goal for the 1996 panel. Unfortunately, we now project nonresponse of about 35% after 12 waves. (Wave refers to the 4 month time period it takes to interview the entire sample.)

Figure 2 graphically presents nonresponse rates from the 1992, 1993 and 1996 panels through wave 4. It shows that even with elevated efforts by the Bureau, response rates for the 1996 panel have *not* improved overall compared to the 1992 and 1993 panels.

Figure 2: Household sample loss rates, Wave 1, 2, 3 and 4



Figures 3 and 4 present attrition for the 1991 and 1996 panels respectively by poverty/non-poverty status at the time of survey drop out. Comparing the two graphs, it appears that the unlocatable mover nonresponse rates (type D nonresponse) are down in the 1996 panel as compared to the 1991 panel. (Rates for the nonpoverty group at Wave 2 are not statistically different from each other). However, the refusal rates (type A nonresponse) have increased in the 1996 panel. (Rates for the poverty group at Waves 2 and 3 were not statistically different). These are the two major components of household level nonresponse that we attempted to reduce in the 1996 panel.

Figure 3: SIPP 1991 panel wave 2 + nonresponse rates by poverty status

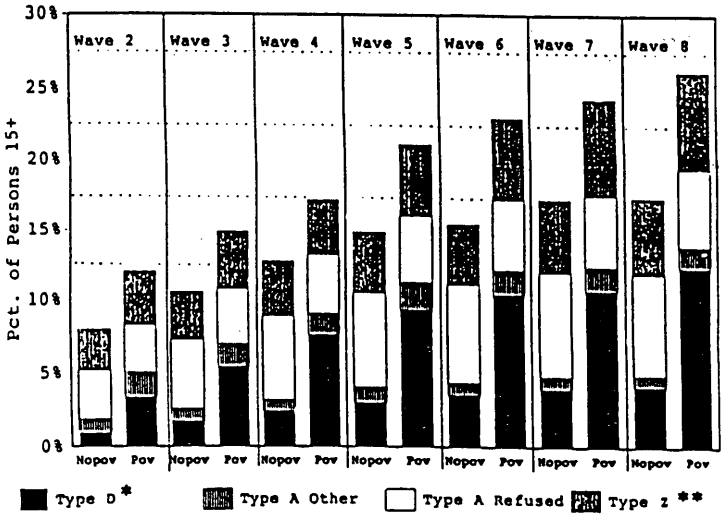
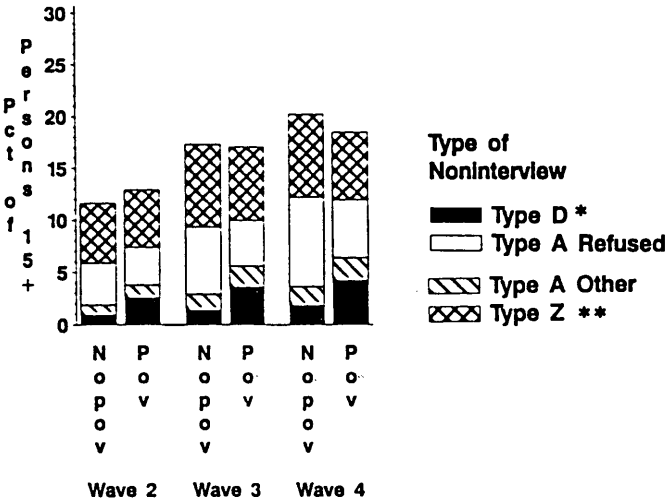


Figure 4: SIPP 1996 panel wave 2 + nonresponse rates by poverty status



* Unlocatable mover ** Within household person nonresponse

3 Efforts to reduce nonresponse in the 1996 SIPP Panel

3.1 Incentives

Without some changes in procedure, the Bureau recognized in 1995 that nonresponse could rise to an unacceptable level by the end of a 4-year panel. A plausible means of maintaining higher response rates is to offer incentives to SIPP sample households. Research has shown that incentives are effective at reducing nonresponse in mail surveys, but little has been done in personal visit interviews. One of the few intensive studies on a personal visit survey offered a nice ball point pen, which increased response rates from 76% to 81% (Willimack et al. 1995). SIPP also reported limited success in the 1987 panel with a one-time non-cash incentive (Butler 1991). More specifically, incentives have been shown to decrease refusal rates (Willimack et al. 1995) and are most effective with minorities and undereducated persons (Berlin et al. 1992, Ferber and Sudman 1974). Since these groups of persons are more likely to have low incomes, the incentive may have higher value to them. A major objective of the SIPP is to provide measures of economic well-being among the low income population, so it becomes important to keep this population well represented in the SIPP sample.

We designed the SIPP experiment to answer the following questions:

1. Do incentives reduce nonresponse at the first (Wave 1) interview?
2. Do incentives reduce nonresponse among low income households at the first interview?
3. Do incentives at Wave 1 reduce nonresponse at subsequent interviews?
4. Do incentives at Wave 1 reduce nonresponse at subsequent interviews for the low income population?

At the initial Wave 1 interview of the 1996 SIPP Panel, the incentives were given as early as possible in the personal visit to the addresses of test cases. The incentives were introduced to the respondent as „a token of appreciation". The respondents were given a paper voucher that resembled cash with the denomination of the incentive printed in the corners. The respondents were instructed to fill in their name and check their address and return it to the Census Bureau in a preaddressed stamped envelope. They were told that they would receive a check for the amount of the incentive in 2 to 3 weeks. Interviewed and noninterviewed households received the incentive; i.e., incentives were given out regardless of the household interview status. Approximately one-fourth of the sample received vouchers for $10, one fourth received vouchers for $20, and one-half did not receive vouchers. This corresponded to a sample size of about 10,000 households each per voucher treatment groups and 20,000 for the control group.

Figure 5: Rot 2-4 sample loss

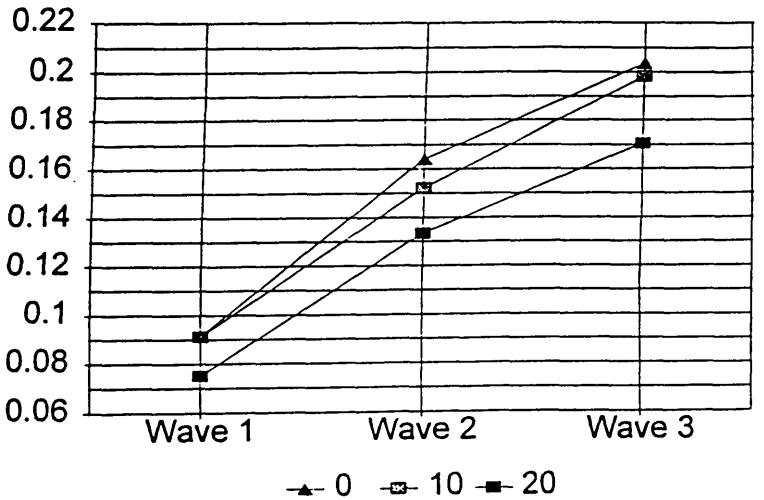
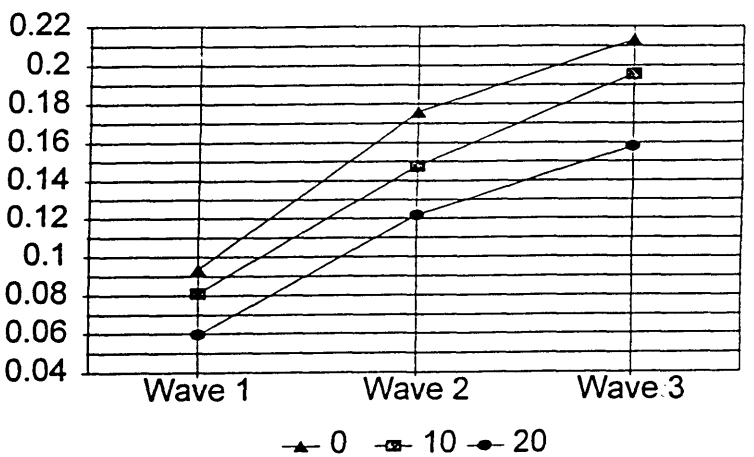


Figure 6: Rot 2-4 sample loss: Poverty stratum



Treatment groups were assigned at the stratification Primary Sampling Unit (PSU) level. Typically, a stratification PSU is made up of one or more counties in the U.S. The PSUs were sorted into 11 blocks based on their 1990 Decennial Census number of households. Each block was composed of 23 to 39 PSUs. The PSUs were ordered by size within each block and then the sample was randomly assigned to the $0, $10, and $20 groups.

Generally, interviewers were assigned to only one treatment group. The exception came when cases had to be reassigned due to reluctant respondents or interviewers leaving. Interviewers were aware of the experiment and the treatment groups, which probably affected their motivation for getting completed interviews.

As shown in Figure 5 and Figure 6, the $20 incentive significantly reduced overall nonresponse rates and nonresponse rates in the high poverty stratum, i.e., the SIPP design stratum with a high proportion of poverty units. There is also some evidence that the $10 and the $20 incentives are effective at Wave 2 at reducing nonresponse rates both overall and in the high poverty stratum. By Wave 3, the evidence is quite substantial that the $20 incentive is effective both overall and in the high poverty stratum. For more detailed information on the results, see James (1997). There is also evidence that incentives reduce the number of callbacks needed to obtain a complete interview. This is important in balancing the cost of the incentive with the cost of repeated visits. This result is consistent with findings in incentive literature cited earlier.

Evident from the decrease in nonresponse rates in the high poverty stratum, the $20 incentive is very likely to reduce the nonresponse bias associated with low income households in the SIPP.

This effort in the 1996 panel was a success. We will continue to analyze the data for all waves and look at results for other subgroups such as participants in government programs, Black, Non-Black and Hispanic subgroups.

3.2 Nonresponse study

SIPP Wave 1 nonrespondents are not contacted in subsequent waves, and the Wave 1 nonresponse adjustment is an integral part of each future wave's weighting adjustment. Also, very little is known about nonresponse bias in Wave 1, other than what is available on the sampling frame. Because of this, the quality of Wave 1 nonresponse adjustment is a high priority, especially in light of a four-year panel.

To assess the feasibility of improving adjustments for Wave 1 unit nonresponse in the 1996 Panel, we conducted two surveys:

- *One survey was a mail-out/mail-back questionnaire to gather limited information from nonrespondents after close-out of SIPP data collection for Wave 1.* ***Nonrespondent Questionnaire*** *(see appendix, Figure 7)*
- *The other was filled in by our interviewing staff called field representatives (FRs) after each noninterview to collect observational information;* ***Field Representative Questionnaire*** *(see appendix, Figure 8)*

Together, the surveys provide insight into the quality of the SIPP Wave 1 noninterview adjustment, as well as providing information to assess Wave 1 nonresponse bias.

Methodology

To evaluate whether the study results can be used to reduce the nonresponse bias associated with important subject matter estimates such as poverty, program participation, and income distributions, two types of analyses were performed:

1. We compared responses across the two surveys to determine how well FRs served as a resource in imputing for nonresponse. High correlations indicated that field representatives serve well as such a resource. We calculate three measures of association: a nonparametric percent concordance, a continuous simple correlation, and a categorical Cramer V association measure. The Cramer V association measure is described in Kendall and Stuart (1979) and the percent concordance measure is one minus the gross difference rate, commonly used in reinterview analysis.
   We then consolidated the two surveys, taking answers from the nonrespondent when we had them, taking answers from the FR when we did not. We compared the aggregate to the production database to determine how well responses were in concordance for those households that were eventually converted to completed interviews in production and for those households that remained type A noninterviews. High correlations indicate that responses from FRs and nonrespondents are in agreement with the production database. We calculated the same association measures as discussed above.
2. Distributional properties of respondents and nonrespondents were also analyzed. This was performed at the aggregate level. We produced crosstabulations of key characteristics by their nonrespondent status. FRs filled out questionnaires for type A nonrespondents who, after further follow-up, were finally converted to a completed response. Because of this we were able to partition respondents and nonrespondents into three categories:

- "Early Respondents" are those households that responded in the production database and did not have FR questionnaires filled out.
- "Late Respondents" are those households that responded in the production database and had FR questionnaires filled out because they were originally type A noninterviews.
- "Type A Noninterviews" are those households that have completed FR or nonrespondent questionnaires and are in the production database as type A noninterviews.

Distribution of demographic and housing factors such as tenure, race, and income were compared for the three types of respondents using a polytomous logistic regression. The higher the log odds ratio in absolute value, the stronger the relationship between the demographic or housing factor and whether the household was an early respondent, a late respondent, or a type A noninterview.

The objective here was to determine whether we could correct or weight for nonresponse bias by examining the characteristics of nonrespondents who are reluctant to participate in the initial phases of SIPP but later consent and characteristics of individuals who remain nonrespondents. This analysis should help to identify other variables to be used in developing a new nonresponse adjustment procedure, where original SIPP sample and respondents to study differ.

To obtain as high a response rate as possible from the nonrespondents and to keep the incremental workload for field representatives as low as possible, we kept the nonrespondent questionnaires short (one page front and back for both field representatives and nonrespondents) and changed the mode of the study instrument of nonrespondents from personal interview to mail-out/mail-back. Specifically, we limited the questions to those items used for Wave 2+ nonresponse adjustment and specific measures of interest concerning income, poverty, and program participation. We also asked questions of the nonrespondents that we believed the field representatives could also answer.

For monthly household income, three categories (<1200, 1200-3999, and 4000+) were provided to field representatives while six categories (<500, 500-1199, 1200-2999, 3000-3999, 4000-8999, and 9000+) were provided to nonrespondents because we believed that nonrespondents could provide a more precise estimate of their monthly household income than field representatives. We wanted as many income categories as possible to compare nonrespondents to the respondents in the production database. The six categories are collapsed into the three field representative categories when comparing answers between nonrespondents and field representatives.

Shortening the length and changing the mode of the questionnaire was highly successful in the nonrespondent survey. Of 3,194 type A noninterviews in Wave 1 of the 1996 SIPP Panel, all were sent questionnaires. Counting only those forms that were completely filled in or partially filled by the respondent, and excluding ineligibles we obtained a response of 716 questionnaires or a rate of 22%. Of the remaining nonrespondent questionnaires returned to the Bureau, approximately half were undeliverable as addressed or out of scope. We found limited success with the mail-out/mail-back short questionnaire for those from whom we were unable to obtain complete SIPP interviews in a personal interview. We need to determine the proportion of them that were refusals versus not at home, etc. to access what portion of the nonrespondent population was willing to cooperate under this scenario versus the full SIPP experience.

After matching the nonrespondents survey data with the interview observational data, we conclude that *field representatives do as well as nonrespondents* in providing some information. These are the variables with relatively high correlation or Cramer V statistics.

- *race of reference person*
- *number of residents in household*
- *number of children in household*

There are three variables that lead us to conclude that *field representatives do not perform as well as nonrespondents* in providing information. They are:

- *rented in public housing project*
- *received rent subsidy*
- *household monthly income*

For the other variables, we believe that further research is prudent.

This research has led to interesting observations:

1. We did not expect to see the results concerning income when comparing field representatives and nonrespondents. When asked to obtain proxy information or estimate household income themselves, FRs tend to understate household income.
2. We did not expect to see the disparity concerning public housing. That leads us to believe that nonrespondents and field representatives may have differing definitions of public housing projects. In fact, the FR may even be more correct in their definition if they talked to a knowledgeable respondent such as a superintendent. We wish to look into a greater understanding of this issue in the future as it potentially indicates a self-identification problem on the part of respondents and an identification problem

on the part of field representatives, both of which can be problematic to researchers when conducting any poverty survey.

The next step is to develop a new nonresponse adjustment procedure using those variables that we determined were most associated with response. We will construct a test database of all respondents and nonrespondents to Wave 1 and determine which adjustment cells nonrespondent households will reside.

1. *We will reweight nonrespondent households in those adjustment cells and compare those weights to their original nonresponse adjustment.*
2. *We will test whether the values of income, poverty, and program participation estimates show a statistically significant change as a result of the reweighting.*

It is possible to extend this study in the future to incorporate administrative records for nonresponse adjustment. These efforts may be undertaken to validate reporting error. For interviewed cases, we could compare values reported in the survey to values derived from administrative records. We could use auxiliary information where possible (e.g., reinterviews) to determine which measures are biased if they disagree. We could then consider whether the differences are systematic; e.g., due to conceptual or time period differences, and whether such differences could be modeled and then form the basis for adjusted values.

Whatever the case, having two questionnaires supplement for nonresponse brings us a long way in understanding who nonrespondents are and how to adjust for them as necessary. Final results could lead to changes in SIPP Wave 1 weighting and routine collection of nonrespondent and/or field representative data.

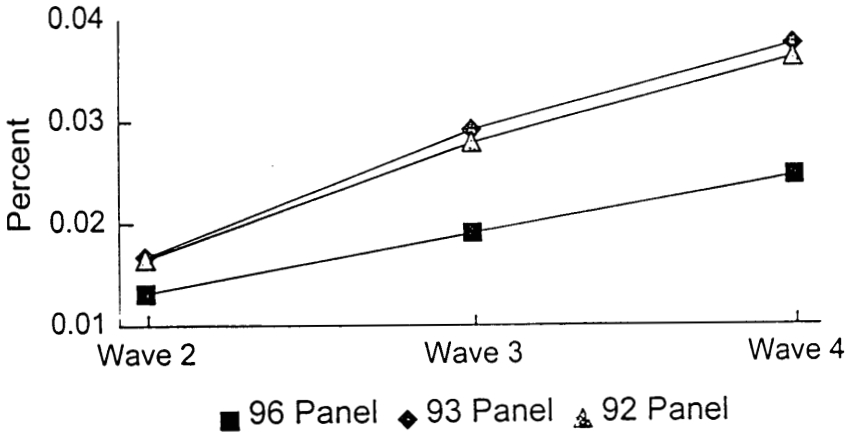
3.3 Field improvements

We implemented several changes in the field work to try to improve response rates. In general, we disseminated more information at all levels in our Field Division to educate and enlighten staff about the existence and harmful effects of nonresponse on longitudinal survey estimates. We focused on following mover activities since Field had not traditionally recognized mover noninterviews as causing serious bias as compared to refusals, not-at-home, and temporarily absent cases. Below are some specific efforts.

3.3.1 Centralizing locating activities

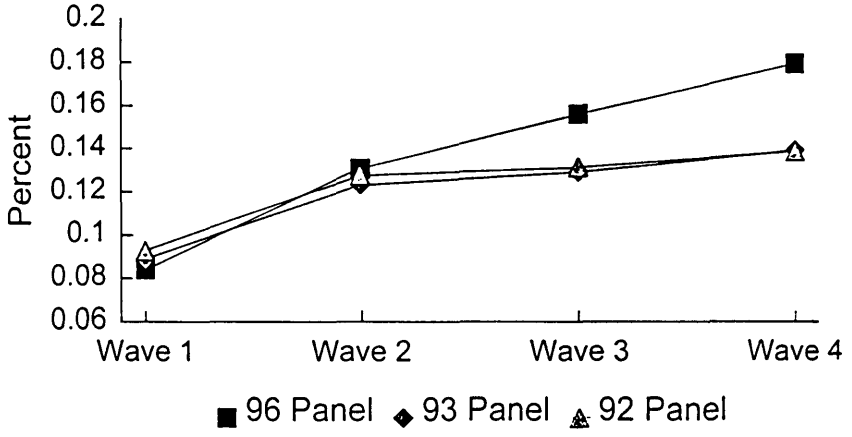
Each Regional Office (RO) hired a clerk to assist the local FRs in tracking SIPP movers. This tracker is not just for SIPP cases exclusively, but for all surveys. Once a case is identified as a type D (unlocated mover) by the FR, the tracker is assigned the case. The locator begins to try to find the mover during what previously had been the resting months, i.e., the months between interviews. FRs and Regional locators would work together, communicating by telephone, for a total of eight months on each case designated as a type D noninterview.

Figure 9: Type D (unlocated mover) rates: Waves 1, 2, 3 and 4



As seen in Figure 9, type D rates in the 1996 panel were practically cut in half as compared to the 1992 and 1993 panel rates. Figure 10 indicates however, that type A refusal rates increased - almost by 4 percentage points. The type D improvement is overshadowed by the increase in refusals.

Figure 10: Type A (refused) rates: Waves 1, 2, 3 and 4



3.3.2 Extending the length of time to track movers

The goal is to increase the rate of successful locations by extending the time to track from 5 months to 9 months. This is accomplished by increasing the number of waves a type D noninterview is reassigned to the field. Previously, it was thought that missing two interviews in a row would make the case longitudinally worthless. However, we plan to implement a new imputation procedure to impute for one or two consecutive missing waves. Therefore conversion after two waves can still benefit longitudinal analysis.

We will begin to tabulate the number of cases that will be improved for longitudinal analyses through this effort after Wave 6 becomes available - this will give two years worth of longitudinal data to evaluate the increase in useable sample.

3.3.3 Feedback of total sample loss rates to ROs monthly

The idea behind this initiative was to feedback not only the type A rate but also the type D rate to the ROs. This had not been done systematically in the past. The quality of the SIPP survey is judged by total sample loss, thus the ROs should be judged by the same standards. Field Division is feeding back the *interview* rate to ROs which is defined as the number of interviews divided by assigned workload. This rate is affected by type A, B, C,

and D nonresponse (nonresponse of eligible (A & D) and ineligible (B & C) units). So far, reaction has been negative because this rate is adversely affected by cases that FRs never had a chance to interview, namely type Bs (vacants, converted to business) and type Cs (condemned, demolished).

3.3.4 Including type D rates in field representative rating standards

Up until recently, FRs were evaluated on their type A rate and their production time. Thus, locating a mover was a lower priority than keeping refusal rates and production times low. We believe that including type D rates in FRs' ratings reinforces the fact that locating a mover is as important as keeping a reluctant sample household participating. Field Division initially felt that the ROs would disagree with the initiative and simply ignore the standard which is within their rights. The principal reason for disagreement is the difficulty in implementing the standard fairly. There are some areas in the U.S. that have little migration and/or where tracking is easy, but there are also areas that have a high mobility rate and offer few leads for movers. This initiative was implemented with some flexibility accounting for regional differences in migration patterns. This change may contribute to the different pattern of type A and D cases we see in the 1996 panel. Perhaps FRs coded refusals as unlocatable movers in the past. However, changes in rating standards would no longer make that advantageous to them.

3.3.5 Automatic and consistent transference of type D noninterviews across regional offices

If a type D is known to move to another RO's area it should *automatically* be transferred, not *optionally* transferred based on past positive or negative experiences with movers or whether it is an interview or not. In the past, we relied on ROs to work this out on an ad-hoc basis but it often caused delays and hard feelings. Computer interviewing makes transference/control of mover cases easier. Field division currently has written guidelines and computer programs that speed up the process and reduce problems.

4 Conclusion

Unfortunately, our 75% response rate goal for the 1996 SIPP 12-Wave panel is eluding us. However our efforts did make a positive difference in the following ways:

- *Wave 1 response rates improved with the use of incentives; particularly in the low income population. In addition, attrition rates have dropped significantly for this incentive group over subsequent waves of interviewing.*

- *The collection of nonrespondent data still has potential for improving Wave 1 weighting. We need to complete the analyses and reweighting pieces to determine the best way to use the nonrespondent information.*
- *Recapture of noninterview cases at later interviews will improve the quality of the data for longitudinal analyses, i.e., we should have more useable cases longitudinally.*

We are concerned to think how high nonresponse *may have been* in the 1996 had we not committed substantial resources to minimize its level and reduce its biasing effects.

A question we have raised repeatedly in light of increasing nonresponse is "when is the level of nonresponse unacceptable?" We were used to observing 25% nonresponse in the 1984-1990 panels, so an increase to 28% in the 1991-1993 panels was not alarming. However, 35% from 25% is quite a jump, especially since we know poverty estimates are adversely affected by attrition. We even have to question whether other estimates we evaluated in the past that seemingly were not adversely effected by attrition before now have crossed over and are seriously biased with the new patterns and levels of attrition.

We will continue to investigate the experimental results and apply positive measures to the 1996 and 2000 panels. We are hopeful that we will once again be able to use incentives even if it is a test in 2000 to determine the number of times incentives are needed to minimize nonresponse. We will also continue to document known effects of nonresponse bias and re-visit some of the earlier findings to determine if they are still valid.

References

Berlin, M., L. Mohadjer and J. Waksberg (1992). An Experiment in Monetary Incentives. Proceedings of the Survey Research Section of the American Statistical Association, pp. 393-398

Butler, D. (1991). SIPP 87: Gift Experiment Results. Internal Census Bureau Memorandum dated April 2 to R. Singh

Ferber, R. and S. Sudman (1974). Effects of Compensation in Consumer Expenditure Studies. Annals of Economic and Social Measurement, pp. 319-331

Huggins, V. J. and F. Winters (1995). Income and Poverty Time Series Data from the Survey of Income and Program Participation. Presented at the International Conference on Survey Measurement and Process Quality, Bristol, England

James, T. M. (1997). Results of the Wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation. Proceedings of the Survey Research Methods Section of the American Statistical Association. Forthcoming

Kendall, M. and A. Stuart (1979). The Advanced Theory of Statistics, Volume 2. New York, McMillan Publishing Company, Inc.

Lepkowski, J., S. Pennel, D. Miller and G. Kalton (1992). Time in Panel Effects in the SIPP. Report for Census Bureau JSA Contract 90-28. Ann Arbor, University of Michigan

McCormick, M. K. (1994). SIPP 1984-1988 Panel: Time-in Sample Effect Study Results. Internal Census Bureau memorandum for the Record

Willimack, D., H. Schuman, B. Pennel and J. Lepkowski (1995). Nonmonetary Incentives in Face-to-Face Surveys. Public Opinion Quarterly, 59, pp. 78-92

APPENDIX

Figure 7: Nonrespondent questionnaire

OMB No. 0607-0813: Approval Expires 09/30/9

FORM SIPP/NFR-16001
(3-12-88)

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

SURVEY OF INCOME AND PROGRAM PARTICIPATION

RETURN TO

BUREAU OF THE CENSUS
1201 East 10th Street
Jeffersonville, IN 47132-0001

NOTICE – The information you provide is **confidential** by law (title 13, U.S. Code) and may be seen only by sworn Census employees and may be used only for statistical purposes.

(Please correct any error in name, address, and ZIP Code)

INSTRUCTION – Please answer these questions as accurately as you can. If you do not know an answer, write "DK."

PGM 2

1. Were you or any other member of your household living at this address last month?
Mark (X) one box.

1 ☐ Yes – *Please continue below with question 1.*
2 ☐ No – ***STOP**, please return this form in the envelope enclosed or mail to the address above.*
THANK YOU FOR YOUR HELP.

PGM 3

1. LIST THE PERSONS WHO LIVED HERE LAST MONTH *(First name, middle initial, last name)*	2. SEX *Use the one-letter codes below to indicate the sex of each household member* **F** – Female **M** – Male	3. AGE *(Years)*	4. HEAD OF HOUSEHOLD *Place an "X" below by one person who owns or rents this residence.*	5. RELATIONSHIP TO OWNER/RENTER *Use the one-letter codes below to indicate how each household member is related to the owner or renter of this residence.* **S** – Spouse (Husband/Wife) **C** – Child (Natural or adopted) **R** – Other relation (Parent, uncle, cousin) **N** – Not related

PLEASE CONTINUE ON THE BACK

▶ *Mark (X) the appropriate box below. If "Yes", provide your best estimate of the amount received for each income type received last month.*

6. Last month, did anyone in this household work for pay –

a. on a job?

1☐ Yes – **Total earned by all who worked** $ _____ . 00

2☐ No

b. as self-employed?

1☐ Yes – **Total earned by all who worked** $ _____ . 00

2☐ No

7. Last month, did anyone in this household receive income from the following?

a. General Assistance (GA)

1☐ Yes – **How much?** $ _____ . 00

2☐ No

b. Aid to Families with Dependent Children (AFDC)

1☐ Yes – **How much?** $ _____ . 00

2☐ No

c. Supplemental Security Income (SSI)

1☐ Yes – **How much?** $ _____ . 00

2☐ No

d. Foster child payments

1☐ Yes – **How much?** $ _____ . 00

2☐ No

e. Women, Infants, and Children (WIC)

1☐ Yes – **How much?** $ _____ . 00

2☐ No

8. Last month, did anyone receive pension income from the following?

a. Social Security

1☐ Yes – **How much?** $ _____ . 00

2☐ No

b. Other pensions

1☐ Yes – **How much?** $ _____ . 00

2☐ No

9. Last month, did anyone get income from the following?

a. Rent from real estate

1☐ Yes – **How much?** $ _____ . 00

2☐ No

b. Dividends from stocks or mutual funds

1☐ Yes – **How much?** $ _____ . 00

2☐ No

c. Interest from bonds, bank accounts *(Do not count interest from a checking account.)*

1☐ Yes – **How much?** $ _____ . 00

2☐ No

d. Any other income

1☐ Yes – **How much?** $ _____ . 00

2☐ No

10. Last month, what was the TOTAL HOUSEHOLD INCOME before deductions? Total household income equals the combined income of all household members from all sources from which money was earned last month. *Mark (X) one box.*

1☐ $ 0 – $ 499

2☐ $ 500 – $ 1,199

3☐ $ 1,200 – $ 2,999

4☐ $ 3,000 – $ 3,999

5☐ $ 4,000 – $ 8,999

6☐ $ 9,000 or more

11. Does anyone in your household have health insurance with Medicaid?

1☐ Yes

2☐ No

PGM 4

12. Are your living quarters – *Mark (X) one box.*

1☐ **Owned or being bought by you or someone in your household?**

2☐ **Rented for cash?**

3☐ **Occupied without payment of cash rent?**

13. Is this residence in a public housing project? *Mark (X) one box.*

1☐ Yes

2☐ No

14. Is this residence owned by a local housing authority? *Mark (X) one box.*

1☐ Yes

2☐ No

15. If rented, is the government paying part or all of the rent for this property? *Mark (X) one box.*

1☐ Yes

2☐ No

16. What race is the head of this household? The head of the household is the person who owns or rents this residence. *Mark (X) one box.*

1☐ White

2☐ Black

3☐ American Indian, Eskimo, or Aleut

4☐ Asian or Pacific Islander

17. Is the head of the household Hispanic? (Mexican, Mexican American, Chicano, Puerto Rican, Cuban, Spanish, Latino, or other Spanish/Hispanic/Latino.) *Mark (X) one box.*

1☐ Yes

2☐ No

18. What is the highest grade of school that the head of the household completed? *Mark (X) one box.*

1☐ 0 – 8th

2☐ 9th – 12th, no diploma

3☐ High school graduate, some college (no degree), vocational/technical school, Associate's degree

4☐ College graduate – any post graduate degree

Please return in the envelope enclosed or mail to the address on the front page.
THANK YOU FOR YOUR HELP.

ORM 64PP/NFR-18001 (3-12-86)

Figure 8: Field representative questionnaire

FORM SIPP/NFR-16003
U 20-96U

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

SIPP TYPE A RECORD FOR WAVE 1 OF THE 1996 PANEL

Section I – TYPE A NONINTERVIEWS – *To be completed by Field Representative for every Type A transmitted to RO. Answer all Items by interview or observation.*

1. FR Code ☐☐☐

2. Date prepared
Month | Day | Year

3. Control number
PSU | Segment No. | Serial No.

4. Name of reference person (owner/renter of house)
Last | First | Initial

5. Household type *(Mark (X) one)*
1 ☐ Female reference person, no husband present, with own children under age 15
2 ☐ Reference person is 65 or over
3 ☐ Other

6. Household size
a. Number of adults age 15 and older ..... ☐☐
b. Number of children age 14 and younger .. ☐☐
c. **Total** (Sum of lines a and b) ..... ☐☐

7. Race of reference person
1 ☐ White
2 ☐ Black
3 ☐ American Indian, Eskimo, or Aleut
4 ☐ Asian or Pacific Islander

8. Tenure
1 ☐ Own – *SKIP to 11*
2 ☐ Rent
3 ☐ Occupied without payment of cash rent – *SKIP to 11*

9. Rented in public housing project
1 ☐ Yes
2 ☐ No

10. Received rent subsidy
1 ☐ Yes
2 ☐ No

11. Monthly household income (Combined income of ALL household members.) *(Mark (X) one)*
1 ☐ Low (less than $1,200)
2 ☐ Medium (between $1,201 and $3,999)
3 ☐ High ($4,000 or more)

12. Market/Rental value of housing unit
a. **If owned, what would you estimate to be the market value of the house or condominium? Market value equals the price the unit would sell for if sold today.**
$ _____ . 00 per month
b. **If rented, what would you estimate to be the monthly rent of the house or apartment?**
$ _____ . 00

13. Number of visits and telephone calls to the household before transmitting to RO
Number of visits to the household ..... ☐☐
Number of telephone calls to the household ☐☐

14. Type A non-interview reason at time transmitted to RO *(Mark (X) one)*
1 ☐ No one at home
2 ☐ Temporarily absent
3 ☐ Refused
4 ☐ Unable to locate
5 ☐ Language problem
6 ☐ Other Type A – *Explain* _____

ENTER COMMENTS ON THE BACK!

Section II – FOLLOWUP – *To be completed by Regional Office Staff.*

1. Was a Type A letter sent to the household?

1 ☐ Yes
2 ☐ No

2. Did an (S)FR visit or contact the household after the Type A was received?

1 ☐ Yes → Number of visits made to the household ..... ☐☐

Number of telephone/other contacts made with the household ..... ☐☐

(S)FR who performed visits/contacts – *(Enter up to 3 (S)FR codes)*

☐☐☐ ☐☐☐ ☐☐☐

2 ☐ No

3. Was the type A converted?

1 ☐ Yes – *SKIP to 5*
2 ☐ No

4. Final Type A reason *(Number from Section I, item 14 (e.g., 3 for refused))* ☐

5. Explain what happened.

Comments

PLEASE RETURN THIS FORM TO YOUR REGIONAL OFFICE WHEN FINISHED WITH MONTHLY ASSIGNMENT. THANKS FOR THE HELP!

FORM SIPP-NFR 16003 (7 70 98)

The Impact of Nonresponse on the Unemployment Rate in the Current Population Survey(CPS)[1]

CLYDE TUCKER AND BRIAN A. HARRIS-KOJETIN

Abstract: *In the present research we matched CPS data from all consecutive months from January, 1994 to June, 1997 and conducted an analysis similar to a gross flows analysis that included nonrespondents to examine the "flow" of persons from respondent to nonrespondent status in the CPS and the resulting effect on labor force estimates. Persons who were nonrespondents to the CPS one month had higher rates of unemployment, labor force participation, and employment than those who were respondents both months. There were also moderate, but significant positive correlations between the differences on unemployment rates and the overall level of nonresponse in the CPS. There were also some differences in labor force characteristics between refusals and noncontacts.*

Keywords: *gross flows analysis; panel nonresponse.*

1 Introduction

The presence of nonresponse can pose problems for drawing inferences from survey data. Thus, it is not surprising that response rates are often used as an indicator of the quality of survey data. However, nonresponse rates only provide an indication of the potential for bias entering into survey estimates. What is critical is the degree to which respondents and nonrespondents differ on the variables of interest. Ideally, one would hope that nonrespondents are a random cross-section of the sample, reflecting the same demographic, geographic, and economic groups. However, it is typically the case that nonrespondents differ from respondents on these characteristics (for reviews see Goyder 1987, Groves 1989). Therefore, even surveys with high response rates may have some degree of bias in their results if the nonrespondents are strikingly different from the respondents.

---

[1] The views expressed in the paper are those of the authors and do not necessarily represent those of the U.S. Bureau of Labor Statistics.

The purpose of the present study was to examine the effect of nonresponse in the Current Population Survey (CPS) on the monthly unemployment rate, the most widely reported result from the CPS, as well as its effects on the labor force participation rate, and percent of the population that is employed. Of course, it is the nature of nonresponse that we don't know the labor force status of those individuals who did not respond to the CPS in a particular month. However, because the CPS is a panel survey in which households are in sample for a total of eight months, we often do eventually obtain labor force information some months even from households who don't respond every month. In this paper, we utilize the analytic technique of gross flows analysis, which is used by economists to examine changes in the labor force status from one month to another, to examine the flow of persons from respondent to nonrespondent status (and vice versa) in the CPS.

1.1 Economic analysis of gross flows

Economists have used gross flow data from panel labor force surveys to examine underlying changes in the labor force classification from one month to the next (e.g., see Barkhume and Horvath 1995, Williams 1995). The monthly labor force "stock" counts give only the total number of people employed or unemployed each month but do not give any indication on what is happening to individuals and how long they may remain unemployed. Indeed, it is possible for there to be little or no month to month change in the overall unemployment rate, but large numbers of people may actually move in or out of unemployment from one month to the next. Specifically, people may move from unemployment to employment (U - E) or vice versa (E - U), or people may search for a job and become unemployed from outside the labor force (N - U), or decide not to keep looking for a job and leave the labor force (U - N).

Data on gross flows is often not published like the monthly totals, and economists have tended to neglect these data for research, perhaps largely due to several methodological problems (see Barkhume and Horvath 1995, Flaim and Hogue 1985, U.S. Dept. of Commerce and U.S. Dept. of Labor 1985). One obvious problem is that the gross flows include only those cases in the survey two consecutive months, which reduces the sample by ¼ due to rotation pattern for the CPS, while movers and nonrespondents effectively reduce the sample further. This results in a discrepancy between the gross flows and the overall labor force counts that can be difficult to estimate. Prior to the 1994 CPS conversion to computerized data collection, matching persons across months was also prone to errors.

1.2 Using gross flows for the study of nonresponse

Although the presence of nonresponse for a particular month leads to these households being excluded for economic analysis that month, we sought to utilize the available information from an adjacent month that the household responded to understand better the characteristics of nonrespondents and the consequences of nonresponse on labor force statistics in the CPS (see also Flaim and Hogue 1985, Stasny and Fienberg 1985, for treatments of non-matches and nonrespondents in analysis of gross flows). Although the vast majority of households that are in the sample in two consecutive months are respondents both of those months (R - R), some households are respondents the first month but do not respond the second month (R - NR), and some are nonrespondents the first month and then become respondents the second month (NR - R). Comparisons can be made between households who responded both months to those that responded only one of the two consecutive months on the labor force characteristics of each of these groups during the month both responded. In addition, we also examined the reason for nonresponse (refusal or noncontact) to see if persons who were interviewed one month, but who refused to participate another month have different labor force characteristics than those who were interviewed one month but were not contacted another month.

2 Design

The CPS is the monthly household labor force survey for the United States conducted by the U.S. Census Bureau for the U.S. Bureau of Labor Statistics. The data collected beginning in January, 1994, are from a redesigned CPS, which incorporates computer-assisted interviewing and a new questionnaire and several improvements in data quality, including a longitudinal identification number that would allow better matching of CPS data from month-to-month. Approximately 50,000 eligible households are sampled each month in a two-stage clustered design.[2] Households selected for the sample are interviewed for 4 consecutive months, are not interviewed 8 months, and then are interviewed again for 4 consecutive months. Furthermore, in any given month, one eighth of the sample is composed of households participating for the first time (month-in-sample 1 (MIS 1)), one eighth the second time (MIS 2), etc. All households except those in for the first time and the fifth time were in sample the previous month; and, therefore, ¾ of the households are the same from month-to-month, and ½ of the households are in the sample the same month from one year to the next.

[2] Approximately 60,000 households were sampled each month during the period January 1994 - December 1995. The sample was cut to 50,000 households due to budgetary constraints beginning in January, 1996.

2.1 Analysis

Data for the present investigation were drawn from monthly CPS files from January, 1994 to June, 1997. Person and household level records from each consecutive pair of months were matched and the major labor force characteristics (unemployment rate, civilian labor force participation rate, and percentage of the population that was employed) from each month was examined by interview status for the other month. Thus, for a given pair of months, four unemployment rates were calculated. The unemployment rate for the first month was calculated separately for persons who responded both the first and second months and for persons who were nonrespondents the second month. The unemployment rate for the second month was also calculated separately for persons who responded both months and for persons who did not respond the first month. The same procedure was followed for calculating the civilian labor force participation rate and the percentage of the population that was employed.

These labor force characteristics were then compared by examining each series over time (from January 1994 to June, 1997) to determine if there were consistent differences in labor force characteristics for persons who responded both months compared to persons who responded the first month but not the second and persons who did not respond the first month but responded the second month. T-tests were used to compare average levels of each group's labor force estimates over the entire time period studied. When significant differences were demonstrated between respondent and nonrespondent labor force estimates, we further examined the correlations of those differences over time with the levels of overall nonresponse to the CPS to see if increasing levels of nonresponse were associated with greater or lesser differences between respondents and nonrespondents. All of the labor force estimates were calculated using base weights, which reflect only the probability of selection, from the month that the labor force data was obtained.[3]

3 Results

3.1 Overview of nonresponse rates

The overall type A nonresponse rate (refusals, noncontacts, and other noninterviews) for January, 1994 to June 1997 averaged 6.6% and ranged from 5.7% to 9.2.%. There was an increase in nonresponse in January, 1994 with the conversion to a redesigned questionnaire with computerized data collection and some other procedural changes.

[3] There were no further adjustments made for nonresponse or for population controls. Also, no composite estimation was done, nor seasonal adjustment. Therefore, these labor force estimates are NOT comparable to any published figures and are presented here for illustrative research purposes only.

There are often peaks of nonresponse for the annual March Income supplement to the CPS, which is typically the highest nonresponse rate achieved during the year. The highest level of nonresponse occurred in December, 1995 at the time of the U.S. Government Shutdown due to budgetary battles between the White House and Congress. Data collection on the CPS was curtailed with approximately 3,000 cases left in the field.

An examination of the month-to-month match of respondent and nonrespondent households across this 3½ year time period revealed that for any given pair of consecutive months, less than 2% of the households that matched across those months were nonrespondents in either the first or the second month and respondents in the other month. Thus, even when the nonrespondents differed from the respondents in their labor force status (in a month they responded), they had quite a small impact on the overall estimate. Approximately 2 percent each month were nonrespondents both months and, therefore, were not included in these analyses.

3.2 First month unemployment rate by second month interview status

Table 1 shows the average unemployment rate for the first month for persons who were interviewed both months and for those who were interviewed the first month, but became nonrespondents in the second month. Persons in households that did not respond in the second month had consistently higher unemployment rates than those persons who were interviewed both months. As can be seen in Table 1, across this entire time series the average unemployment rate for persons interviewed both months was 5.4%, while it was 6.9% for those who were nonrespondents in the second month, and this difference is significant.

A closer examination of the differences in unemployment rates between persons who responded both months and those who responded the first month but did not respond the second month revealed that the differences were higher in the first six months of the new survey (average difference 2.4%), and also higher during the March Supplements (average difference 2.8%). Since these months had relatively high nonresponse rates, we examined the extent to which the nonresponse rate was related to these differences in unemployment rates between these two groups. The correlation between these two series was significant, but only moderate, $r = .30$, $p < .05$. A further examination of refusals and noncontacts separately showed no consistent pattern of differences between these two subgroups of nonrespondents for unemployment rates (see Table 2).

Table 1: Average levels for labor force status from the CPS by interview status

1st month Labor Force Status	Interview in 2nd month	Nonresponse in 2nd month	Difference
Civilian Labor Force	65.98%	68.51%	2.53%**
Employed	62.45%	63.80%	1.35%**
Unemployment Rate	5.35%	6.87%	1.52%**
2nd month Labor Force Status	**Interview in 1st month**	**Nonresponse in 1st month**	**Difference**
Civilian Labor Force	65.79%	67.41%	1.62%**
Employed	62.39%	63.74%	1.35%**
Unemployment Rate	5.18%	5.48%	.30%*

* $p < .05$, ** $p < .01$

3.3 First month labor force participation rate by second month interview status

The average civilian labor force (CLF) participation rate for the first month for persons who were interviewed both months and for those who were interviewed the first month but became nonrespondents in the second month can be seen in Table 1. Persons who became nonrespondents to the CPS had higher labor force participation rates than those interviewed both months, and the overall average difference in CLF between these two groups was 2.5%.

The differences between respondents both months and nonrespondents the second month in labor force participation rates were not related to the overall nonresponse rate in the CPS. However, the reason for nonresponse did affect labor force participation rates. As can be seen in Table 2, refusals the second month had significantly higher labor force participation rates in the first month than persons who were noncontacts the second month.

Table 2: Differences in labor force status from the CPS by type of nonresponse

1st month Labor Force Status	Refusal in 2nd month	Noncontact in 2nd month	Difference
Civilian Labor Force	69.56%	67.59%	1.97%**
Employed	65.00%	62.80%	2.20%**
Unemployment Rate	6.53%	7.10%	-.57%
2nd month Labor Force Status	**Refusal in 1st month**	**Noncontact in 1st month**	**Difference**
Civilian Labor Force	67.16%	67.34%	-.19%
Employed	63.33%	63.78%	-.45%
Unemployment Rate	5.70%	5.31%	.38%

* $p < .05$, ** $p < .01$

3.4 First month percent employed by second month interview status

The percentage of persons employed is shown separately in Table 1 for those interviewed both months and those interviewed the first month who were nonrespondents the second month. The overall average difference between these two groups was significant over the entire 3 ½ years. The differences between respondents both months and nonrespondents the second month in percent employed were not related to the overall nonresponse rate in the CPS. However, the reason for nonresponse did affect the percent employed. As can be seen in Table 2, refusals the second month had significantly higher percentage employed the first month than persons who were noncontacts the second month.

3.5 Second month unemployment rate by first month interview status

There were small, but significant differences between the unemployment rate for persons who were interviewed both months and for those who were nonrespondents the first month but became respondents in the second month (see Table 1). We also examined the extent to which the nonresponse rate was related to these differences in unemployment rates between these two groups and found that the correlation between these two series was significant, but only moderate, $r = .32$, $p < .05$. There were no significant differences between the first month refusals and noncontacts (see Table 2).

3.6 Second month labor force participation rate by first month interview status

As noted in Table 1, the civilian labor force participation rate for persons who were interviewed both months and for those who were nonrespondents the first month but became respondents in the second month was significantly different. There was no significant relation between the differences and the overall CPS nonresponse rate, and there were no differences between refusals and noncontacts (see Table 2).

3.7 Second month percent employed by first month interview status

The pattern of results for percent employed follows that of labor force participation. The percent employed for persons who were interviewed both months and for those who were nonrespondents the first month but became respondents in the second month were significantly different (see Table 1). However, there was no significant relation between the differences and the overall CPS nonresponse rate, and there were no differences between refusals and noncontacts (see Table 2).

4 Discussion

In the present research we matched CPS data from all consecutive months from January, 1994 to June, 1997 and conducted an analysis similar to a gross flows analysis that included nonrespondents to examine the "flow" of persons from respondent to nonrespondent status in the CPS and the resulting effect on labor force estimates. There appeared to be consistent differences between persons who were respondents to the CPS both months and those who were nonrespondents one month for unemployment rates, labor force participation rates, and percent employed. Persons who were nonrespondents to the CPS one month had higher rates of unemployment, labor force participation, and employment than those who were respondents both months. Although persons who respond one month and are nonrespondents another month represent a very small portion of the total sample any given month, they have consistently different labor force statistics than persons who are respondents both months. There is thus potential for nonresponse bias in the CPS labor force estimates.

We also examined the degree to which the differences between respondents and nonrespondents were related to the magnitude of nonresponse. There were moderate significant positive correlations between the differences on unemployment rates and the level of nonresponse. Thus, there is the potential for increasing nonresponse bias in CPS estimates of unemployment with increasing levels of nonresponse.

Finally, we examined the degree to which refusals and noncontacts have different labor force characteristics. There were fewer differences observed, but second month refusals demonstrated a significantly higher labor force participation rate and percent employed in the first month than second month noncontacts.

4.1 Limitations

Although we took advantage of all available data for consecutive months to examine labor force characteristics of persons who were respondents one month and nonrespondents one month, we obviously still do not know the labor force characteristics of persons in households that were nonrespondents both months, or whether labor force status was different for the nonrespondents during the month they failed to respond than the month they responded. In addition, some people who were consistent respondents in one pair of consecutive months were actually the partial respondents in another pair of months. A complete longitudinal data file for all eight months the household is in the sample would allow a more consistent comparison between partial and complete respondents (see Harris-Kojetin and Tucker, pp. 263-272, this volume).

The labor force estimates compared also utilized only base weights, which reflect only the probability of selection. Further evaluation should be made of nonresponse adjustments and adjustments to population controls to see whether these weighting adjustments decrease the observed differences.

In conclusion, while the presence of nonrespondents may be a methodological problem for normal economic analysis of labor force gross flows, nonrespondents from one month who are interviewed in the other month can be included in the analysis to obtain one measure of the effect of survey nonresponse on labor force estimates. In the CPS, it appears that there is the potential for some small nonresponse bias in the unemployment rate, the labor force participation rate and percent employed due to persons who are nonrespondents one month. However, because of the small number of these people in any given month (usually less than 2% of the ¾ of the overlapping sample), the effects on the overall estimates are quite small. Nonetheless, the effects for the unemployment rate appear to increase with increases in nonresponse rates which reinforces the close monitoring of survey response rates.

References

Barkhume, A. J. and Horvath, F. W. (1995). Using gross flows to explore movements in the labor force. Monthly Labor Review, 118(4), pp. 28-35

Flaim, P. O. and Hogue, C. R. (1985). Measuring labor force flows: A special conference examines the problems. Monthly Labor Review, 108(7), pp. 7-17

Goyder (1987). The Silent Minority: Nonrespondents in Sample Surveys. Boulder, CO: Westview Press

Groves, R. M. (1989). Survey Errors and Survey Costs. New York: Wiley

Harris-Kojetin, B. and Tucker, C. (1998). Longitudinal Nonresponse in the Current Population Survey. In: Koch, A. and Porst, R. (eds.). Nonresponse in Survey Research. Proceedings of the Eighth International Workshop on Household Survey Nonresponse 24-26 September 1997. Mannheim: ZUMA, pp. 263-272

Stasny, E. A. and Fienberg, S. (1985). Some stochastic models for estimating gross flows in the presence of nonrandom nonresponse. Proceedings of the conference on gross flows in labor force statistics. Washington, DC

Williams, D. R. (1995). Women's part-time employment: A gross flows analysis. Monthly Labor Review, 118(4), pp. 36-44

U. S. Department of Commerce and U. S. Department of Labor (1985). Proceedings of the conference on gross flows in labor force statistics. Washington, DC

An Evaluation of Unit Nonresponse Bias in the Italian Households Budget Survey

CLAUDIO CECCARELLI, GIULIANA COCCIA AND FABIO CRESCENZI

Abstract: *The effect of nonresponse is a crucial aspect which has received considerable attention in literature. Nonresponse causes an increase in sampling variance and the estimates of units nonresponse bias are useful to give a concrete measure of this increase. Scarce co-operation of respondents gives inaccurate or incomplete information, and this part of the answers can't be used in the estimation. Since 1991, Istat erases from its sample file all households with incompatible data on expenditure and income. We call this the "UR (Unreliable Respondents) sub-population". UR data can't be used to build estimates, but can be used to give approximate information on NR. The idea is that, concerning behaviour, the NR population is more similar to UR than to R.*
The results of an attempt to evaluate unit nonresponse bias in the 1995 Italian Household Budget Survey (HBS) are shown in this paper. These results were obtained without performing a specific survey on nonrespondents.

Keywords: *unreliable respondents, unit nonresponse bias*

1 Introduction

The results of an attempt to evaluate the unit nonresponse bias in the 1995 Italian Household Budget Survey (HBS) are shown in this paper. These results were obtained without performing a specific survey on nonrespondents.

Quality of estimates depend on sampling and non sampling errors and one of the major components affecting non sampling errors is the rate of non response (Platek and Gray 1986). A population can be divided in respondents (R) and nonrespondents (NR), but there are several difficulties in the correct evaluation of non response bias because of the lack of information on NR. The high cost of additional surveys force statisticians to look for new methods to build approximate estimates of bias (Lessler and Kalsbeek 1992).

Scarce co-operation of respondents gives inaccurate or incomplete information, and this part of the answers can't be used in the estimation. Since 1991, Istat erases from its sample file all households with incompatible data on expenditure and income. We call this the "UR (Unreliable Respondents) sub-population". UR data can't be used to build estimates, but can be used to give approximate information on NR. The idea is that,

concerning behaviour, the NR population is more similar to UR than to R.

Non respondent bias is affected by two components. We call V_I or Variable Independent, the first of these two components, which is independent from the value of the considered variable in the NR population. The second one, which we call V_D or Variable Dependent, depends on the differences between values assumed by the considered variable in the R and in the NR populations.

Concerning the first component, NR data are usually available, concerning the second one, it must be estimated and our approach is to use UR data to approximate NR data.

2 HBS survey design

HBS gives a continuous flow of data on Italian households' expenditures, incomes and other characteristics (Istat 1995). There are two main survey techniques on which data collection is based:

- a self-compiled diary. For ten consecutive days consumers record detailed information on expenditures, aiming especially at small purchases, which are often difficult to recall;
- a face to face interview that, at the end of the month, allows information on large expenditures to be collected. It also provides data on family size, dwelling characteristics, durable consumer goods, etc.

The sample scheme is a two stages stratified sample. Communes represent the Primary Sample Units (PSUs). Every quarter are 550 PSUs (out of 8103 Italian Communes) surveyed, stratified by demographic size, altitude zone and prevalent economic activity.

The PSUs of each geographical region are divided in two groups:

1. Communes with more than 50,000 inhabitants, permanently included in the sample;
2. other Communes, included in the sample on a rotating panel basis (one month quarterly).

PSUs are selected from each stratum without replacement and probabilities of inclusion proportional to size. Every quarter 148 PSUs come from the first group and 402 PSUs from the second one.

Secondary Sample Units (SSUs) are households listed in municipal registers. SSUs are sampled using a systematic scheme, without replacement and with equal probabilities of inclusion. All the components of selected households are included in the sample. Every

quarter about 9,000 households are interviewed (yearly about 36,000 households).

To ensure obtaining the fixed sample size, nonrespondent households are replaced by other households of the same PSU.

Quarterly expenditure estimates are available at the national level, while annual estimates are available at the regional level.

The estimator used to produce each of the estimates is post-stratified by household size.

3 HBS response rates and unreliable respondents

Table 1 shows the percentage of households, response rates and mean expenditures by household size.

Table 1: 1995 HBS-percentage of households, response rates and mean expenditure, by household size

	Household Size				
	One component	Two components	Three components	Four or more components	Overall
Households (%)	25.6	29.2	22.3	22.9	100.0
Response rates (%)	86.8	91.8	93.6	94.6	91.3
Mean expenditure (thousands of liras)	1,886	2,763	3,687	4,078	3,2218

It can be highlighted that, not surprisingly, the response rate is an increasing function of the number of components: one component households reach the lowest rate, four or more component households attain the highest.

Nonrespondents can be divided by nonresponse categories. Table 2 shows the percentage of households by those categories.

Table 2: 1995 HBS-percentage of NR households by nonresponse categories

	Percentage
Cannot be found	26.3
No one at home	23.3
Refusals	32.6
Others	17.8
Total nonrespondent	100.0

The "Cannot be found" category includes those households which couldn't be contacted for the entire survey period (i.e. wrong address). The "No one at home" category includes households which couldn't be found at home after repeated attempts at various times. The "Refusals" category includes households refusing to participate in the survey.

Table 3 gives the distribution of nonrespondents by nonresponse categories and household size.

Table 3: 1995 HBS-percentage of NR households by nonresponse categories and household size

Household Size	Nonresponse categories				
	Cannot be found	No one at home	Refusals	Others	Total Nonrespondents
One person	40.0	35.3	23.4	46.1	33.8
Two persons	22.6	26.7	28.5	23.1	25.7
Three persons	18.2	16.2	22.9	13.4	18.7
Four or more persons	18.6	21.5	24.8	15.1	21.0
Not indicated	0.6	0.3	0.4	2.3	0.8
Total	100.0	100.0	100.0	100.0	100.0

Identification of unreliable households (UR) is based on the comparison of data on food and non-food expenditures, income and family size.

Households showing very low food and non-food expenditures, and a high percentage of the non response item is labelled as unreliable.

Unreliable data are deleted from the original file and are not used to estimate expenditures.

Dividing population in two sub-sets: r = RR and $\bar{r} = (NR + UR)$, it is possible to define the nonresponse rate as:

$$T = \frac{\bar{r}}{r + \bar{r}} \times 100 = \frac{NR + UR}{RR + NR + UR} \times 100 \qquad (1)$$

Table 4 shows the number of UR, NR, R and T by region. T assumes the highest values in Friuli V.G. (17.7%) and Piemonte (13.9%), the lowest values in Marche (4.8%) and Puglia (6.8%).

Table 4: 1995 HBS-nonresponse rates by region

Region	Respondents RR	Nonrespondents NR	Unreliable UR	$T = \frac{NR + UR}{NR + UR + RR} \times 100$
Piemonte	2,631	382	44	13.9
Valle d'Aosta	936	41	35	7.5
Lombardia	3,211	468	40	13.7
Trentino Alto Adige	1,940	208	27	10.8
Veneto	1,760	117	26	7.5
Friuli Venezia Giulia	998	199	16	17.7
Liguria	1,713	168	12	9.5
Emilia Romagna	2,029	258	43	12.9
Toscana	2,437	232	35	9.9
Umbria	1,035	58	32	8.0
Marche	1,416	41	30	4.8
Lazio	2,418	175	69	9.2
Abruzzo	960	38	57	9.0
Molise	1,028	77	32	9.6
Campania	2,564	116	98	7.7
Puglia	1,924	74	67	6.8
Basilicata	923	15	73	8.7
Calabria	1,251	118	41	11.3
Sicilia	2,334	146	110	9.9
Sardegna	895	71	41	11.1
ITALY	34,403	3,002	928	10.3

4 Evaluation of unit nonresponse bias

1995 HBS expenditure estimates are based on a family size post-stratified estimator and a two stage stratified sample. Post stratification reduces the effects of nonresponse on estimates (Cochran 1977; Holt and Smith 1979).

In this paper, we show the results of an evaluation of bias obtained by using a simplified sample scheme, which does not consider the first stage PSUs, but only a stratification by regions of sampled households and a post stratification by household size. This is to reduce the formula's complexity, with slight impact upon results. However, it is easy to generalise formulas to the more complex case.

Let us call:

h	stratum index (regions) (h=1,...,H);
j	SSUs (households) by stratum index (j=1,..., M_h);
l	size of households class (l=1,...,L);
M_{hl}	number of SSUs of size l in the stratum h;
m_{hl}	number of sampled SSUs of size l in the stratum h;
M_h	number of SSUs in the stratum h;
m_h	number of sampled SSUs in the stratum h;
M	number of SSUs;
m	number of sampled SSUs;
y_{hlj}	expenditure y of the j SSU of size l in the stratum h.

The mean population value of y is then:

$$\bar{y} = \frac{1}{M} \sum_{h=1}^{H} \sum_{l=1}^{L} \sum_{j=1}^{M_{hl}} y_{hlj} \tag{2}$$

As in par. 3 we divide population in two sub-sets r=RR and $\bar{r} = (NR + UR)$. So it is possible to write:

$$\bar{y} = \frac{1}{M} \sum_{h=1}^{H} \sum_{l=1}^{L} \left(\sum_{j=1}^{{}_rM_{hl}} y_{hlj} + \sum_{j=1}^{{}_{\bar{r}}M_{hl}} y_{hlj} \right) = \frac{1}{M} \sum_{h=1}^{H} W_h \sum_{l=1}^{L} W_{hl} \left[t_{hl}\, {}_r\bar{y}_{hl} + \left(1 - t_{hl}\right) {}_{\bar{r}}\bar{y}_{hl} \right] \tag{3}$$

where ${}_rM_{hl}$ is the total number of respondent households of size l in the stratum h,

${}_{\bar{r}}M_{hl}$ is the number of non respondents and unreliable households of size l in the stratum h, and

$$W_h = \frac{M_h}{M}, \quad W_{hl} = \frac{M_{hl}}{M_h}, \quad t_{hl} = \frac{{}_r M_{hl}}{M_{hl}}.$$

Let us consider $\hat{\bar{y}}$, a direct estimator based only on respondent households:

$$\hat{\bar{y}} = \frac{1}{M} \sum_{h=1}^{H} \sum_{l=1}^{L} \frac{M_{hl}}{{}_r m_{hl}} \sum_{j=1}^{{}_r m_{hl}} y_{hlj} \tag{4}$$

Using $\hat{\bar{y}}$ as estimator of $\bar{y}$: it is possible to show that

$$E(\hat{\bar{y}}) = \sum_{h=1}^{H} \frac{W_h}{t_h} \sum_{l=1}^{L} t_{hl} W_{hl} \, {}_r\bar{y}_{hl} \tag{5}$$

where:

$$t_h = \frac{{}_r M_h}{M_h}.$$

The bias of $\hat{\bar{y}}$, $B(\hat{\bar{y}})$, is given by:

$$B(\hat{\bar{y}}) = E(\hat{\bar{y}}) - \bar{y} \tag{6}$$

$$B(\hat{\bar{y}}) = \frac{1}{t} \sum_{h=1}^{H} \frac{W_h}{t_h} \sum_{l=1}^{L} W_{hl} (t_{hl} - t_h) {}_r\bar{y}_{hl} + \sum_{h=1}^{H} W_h \sum_{l=1}^{L} W_{hl} (1 - t_{hl}) ({}_r\bar{y}_{hl} - {}_{\bar{r}}\bar{y}_{hl}) = V_I + V_D \tag{7}$$

where:

$${}_r\bar{y}_{hl} = \frac{1}{{}_r M_{hl}} \sum_{j=1}^{{}_r M_{hl}} y_{hlj} \quad \text{and} \quad {}_{\bar{r}}\bar{y}_{hl} = \frac{1}{{}_{\bar{r}} M_{hl}} \sum_{j=1}^{{}_{\bar{r}} M_{hl}} y_{hlj} \tag{8}$$

This result confirms the above mentioned results: two components affect nonrespondent bias. The first of these is independent from the values of the considered variable observed on the nonrespondents. The second one is a function of the differences between values assumed by the considered variable in the respondent and in the nonrespondent populations.

We estimate $B(\bar{y})$ substituting in (6) population values with sample estimates as follows:

$$\hat{t}_{hl} = \frac{{}_r m_{hl}}{m_{hl}} \tag{9}$$

$$\hat{t}_h = \sum_{l=1}^{L} \hat{t}_{hl} W_{hl} = \sum_{l=1}^{L} \frac{{}_r m_{hl}}{m_{hl}} \times \frac{M_{hl}}{M_h} \tag{10}$$

$$\hat{t}_h = \sum_{h=1}^{H} \hat{t}_h W_h \tag{11}$$

$${}_r\hat{\bar{y}}_{hl} = \frac{1}{{}_r m_{hl}} \sum_{j=1}^{{}_r m_{hl}} y_{hlj} \tag{12}$$

Our sample doesn't give us the possibility to estimate ${}_{\bar{r}}\bar{y}_{hl}$, so we use UR data to approximate $\bar{r}$ in the following way:

$${}_{\bar{r}}\hat{\bar{y}}_{hl} = \frac{1}{{}_{\bar{r}} m_{hl}} \sum_{j=1}^{{}_{UR} m_{hl}} {}_{UR} y_{hlj} k_{hl} \tag{13}$$

where:

$$k_{hl} = \frac{{}_{\bar{r}} m_{hl}}{{}_{UR} m_{hl}} \tag{14}$$

5 Results

Table 5 shows the results of this study. At this initial stage only some of the most important food and non-food items have been considered.

The underlying assumptions are restrictive, but a lack of other information sources and the quality of results encourages to go ahead in this direction.

Table 5: 1995 HBS-estimates of bias for some items

Items	Monthly Mean Expenditures (MME)	BIAS (B)	BIAS in percentages (B / MME)×100
Bread	43,746	260.75	0.60
Meat	42,017	95.22	0.23
Fruit and vegetables	33,439	359.85	1.08
Electricity	46,640	-149.83	-0.32
Gas	56,552	-402.65	-0.71
Telephone	55,543	139.07	0.25

For all items it is possible to observe small absolute bias values which surpass 1% of monthly mean expenditures only in the fruit and vegetables case.

Concerning food items, the method highlights an overestimation of expenditures evaluated only on the basis of respondents. Expressed as a percentage, this overestimate ranges from 0.23 for meat, to 1.08 for fruit and vegetables.

For non-food items the results reveal an underestimation for gas and electricity (respectively -0.71 and -0.32), and an overestimation for telephone (0.25).

It is our intention to repeat this exercise for 1996 HBS data increasing the number of items considered.

Further studies will include tests of robustness of methods and the possibility of specifying a probabilistic model that, starting from the results of a given number of items, allows us to generalise results so as to estimate bias for the complete set of expenditures.

References

Bethlem, J.G. and Kersten, H.M.P. (1981). The nonresponse problem in the Netherlands. Paper presented at the U.N. Conference of European Statisticians, Meeting relating to problems in household Survey, Geneve

Cochran, W.G. (1977). Sample Techniques. J.Wiley: New York

Hansen, M.H. and Hurwitz, W.N. (1946). The Problem of Nonresponse in Sample Survey. Journal of The American Statistical Association, 41, pp. 517-529

Holt, D. and Smith, T.M.F. (1979). Post-stratification. J.R. Statist. Soc. A., 142, pp. 33-46

ISTAT (1994). Indagine sui consumi delle famiglie - Anno 1993. Collana di informazione, 22, Roma

Kalton, G. and Kasprzyk, D. (1986). The Treatment of Missing Survey Data. Survey Methodology, 12, 1, June 1986, Statistics Canada

Lessler, J.T. and Kalsbeek, W.D. (1992) Non Sampling Error in Surveys. Wiley Series in probability and mathematical statistics, chapter 8

Madow, W.G., Nisselson, H. and Olkin, I. (eds.) (1983). Incomplete Data in Sample Surveys, Vol. 1. Academic Press: New York

Madow, W.G., Olkin, I. and Rubin, D.B. (eds.) (1983). Incomplete Data in Sample Survey, Vol. 2. Academic Press: New York

Mopsik, J.A. and Dippo, C.S. (1984). The adjustment for the consumer Expenditure Interviewed Survey. Proc. Comp. Sect., A.S.A

Platek, R. and Gray, G.B. (1986). On the Definition of Response Rate. Survey Methodology, 12, 1. Statistics Canada

Rubin, D.B. (1983). Conceptual Issues in the Presence of Nonresponse. Incomplete Data in Sample Survey, Vol. 2. Academic Press: New York

Thomsen, I. (1973). A note on the efficiency of weighting subclass means to reduce the effects of non response when analyzing survey data. Statisstisk Tidskrift, 4

Nonresponse in the 1996 Income Survey (Supplement to the Microcensus)

ÉVA HAVASI AND ÁDÁM MARTON

Abstract: *Income suvrvey: supplementary voluntary questionnaire to the randomly selected one quarter of the April 1996 microcensus, 18,117 households, 16% (2,988) refusals. The characteristics of the households which refused to answer the income survey (they are not randomly distributed) were carefully studied. The most important results and measures which were done to reduce the bias are presented. A substantial number of census variables for households were found to be associated with nonresponse. The characteristics most strongly associated with Income Survey response rate were the qualification level of head of household and the type of region. Higher response rate was found in the countryside among the older households with low income and low qualification and high refusal rate in Budapest, mainly in high income groups with high qualification level, as well.*

Keywords: *nonresponse rate, nonrespondents, income survey*

The Unified System of Household Surveys (USHS) has been operating since 1976. It is part of the Hungarian Central Statistical Office (CSO), and covers the noninstitutional population.

There are two large ongoing rotating household sample surveys:

- The Family Expenditure Survey (FES)
- The Labor Force Survey (LFS).

Up to 1991 a separate income survey was also carried out in every four or five years.

Supplementary, short questionnaires on different topics are sometimes added to the FES and/or LFS.

In most household surveys, which are face to face interviews, the response rate is still fairly high. This is due partly to the good sampling frame and field work and also to the long expertise, and to a certain extent, the good respondent-interviewer relations. But the response rate differs depending on the topic of the survey and the strata of the population. In recent years a decline was observed.

The interviewers' activities are organized and supervised by the County Offices of the CSO. The interviewers in the smaller towns and villages establish, in many cases, some kind of personal contact with the members of the households in the sample. This helps to reduce refusals in the consecutive waves but as far as response error is concerned, it has some disadvantages too. The given arrangement of field-work is crucial: it determines the way how to handle nonresponse, nonsampling error, etc.

Careful (follow-up) studies, call-backs help to reduce the number of nonresponse and also try to get some basic information on nonresponding households. Bias can be also reduced through poststratification, reweighting.

It is obvious that the estimates for the "tails" of the target population (very poor, very rich) are unreliable due to high nonresponse rate and also certain nonsampling errors[1].

It appears now that life is getting tougher for survey methodologists and the populations we need to survey. On the one hand, people are more vehemently cautioned about sharing "private" information with anyone (government officials included) because of misuse of information and fraud. The population has many more family situations with both spouses working, complicating their lives, and limiting their available time to surveys. And there is a concern that information will somehow be used with negative reprecussions on them personally.

On the other hand, those of us administering government sponsored surveys rely on the existence of civic responsibility in the population, well-trained and talented field staff emphasising the importance of the survey data to maximize response rates[2].

There are population censuses every ten years in Hungary and a microcensus in the middle of these decades. To answer the census questionnaire is mandatory, while in the case of household surveys it is voluntary.

The last microcensus was carried out in the spring of 1996.

Stratified, multilevel sample was designed for the microcensus. The primary sampling units were the settlements. All cities with more than 15000 inhabitants were included, the other settlements were selected proportionate to size. The secondary sampling units were the enumeration districts while the final sampling unit was the address (the household), selected from the updated list of addresses of the enumeration districts. The USHS has a sample of fairly similar character to the microcensus. The experienced field staff helped to conduct the microcensus.

---

1 Marton 1995

2 P.J. Wait and V.J. Huggins 1996

The "new" personal income and comsumption tax system in Hungary was introduced in 1988. Then the private sector was very small. The privatization procedure of the former stately owned (socialist) sector will soon be finished, but the more important point for us is that there are more than a million small private entrepreneurs (fewer than 20 employees).

The willingness of tax evasion is very strong. While employees (wage earners) cannot avoid paying taxes, the small (private) businesses are very successful in that activity. This is a complicated issue. According to certain estimates, the volume of the "black" economy is about 30 pc. of the GDP. In such circumstances it is understandable that the results of an income survey will be certainly biased. Still we have to try to collect data which are fairly reliable at least in certain strata of the population.

Due to the changing survey climate in the country since 1991 there have been no more separate income surveys. Within the new market oriented society data on the widened income distribution are very much needed (10 pc. unemployment and some with high incomes).

To a randomly selected one quarter of the microcensus (0.5 pc. of the total population) a non-mandatory supplementary questionnaire was attached on the amount and sources of the incomes. Another supplementary questionnaire with the same sample size (another quarter of the microcensus) dealt with travelling habits.

One quarter of the microcensus contained 20,087 households for the income survey. The microcensus was successful in 18,117 households (90 pc.). The reason of non response was vacant flat, no contact, not at home, etc. also refusal[3]. The interview started with the microcensus questionnaire (mandatory) and it was not mentioned in advance that some further questions would follow. When it was completed, the interview with the supplementary (non-mandatory) income questionaire followed. The result: 15,126 completed income questionnaire (Table 1).

Having matched the approx. 2,900 refusing households, in the case of the income survey with their census data it was possible to analyse the characteristics of the families whose basic characteristics were available from the census.

---

[3] Refusals 4.6% and non-contact from other reasons 3%. No measures were taken although to answer was mandatory.

Table 1: Share of households by type of nonresponse in the Hungarian Income Survey, 1996

Response type	Share of households %		Number of households	
Respondents	83.5		15126	
Nonrespondents	16.5		2989	
Refusals		12.7		2301
Partly refusals		3.3		599
Non-contact		0.5		90
Total	100.0		18117	

As it was already mentioned, the very poor, homeless people are missing from the survey as well as those who are very rich or with very high social status, because they generally refuse to answer the questions. In the case of censuses the coverage was more complete. The voluntary supplements are confined to the "average" population about 80-90 pc. of the total. *Who are those who refuse to answer? Where do they live, which strata of the population do they belong to, etc.?*

The different strata of the population behave differently. It is a general observation that the young one person households – practically regardless of the topic of the survey – are not willing to answer or it is difficult to contact them. At the same time pensioners are sometimes happy to answer the interviewer. Some people refuse to discuss topics like income, wealth, religion etc. But they are happy to answer the other kinds of questions: health care, vacation, travelling habits, etc. (The refusal in the latter case was only 6.8 pc., while that of the income supplement was 16.5 pc.) In Hungary income is a real sensitive topic not only because many people do not like to talk about it, but – as it was already mentioned – the so-called black or grey economy is very strong. Those who answer the questions usually say what they declared in their tax returns. Employees generally are not interested in underreporting their incomes, while the family income in the private sector is very underestimated.

The observation unit is the household. It is also a general phenomenon, that if one member of the household refuses to answer the others follow his or her example. When in the following we talk about refusal, it will always refer to a household. It is true, that if the interviewer can build up some contact with the family ("foot in the door") it is highly probable that the interview will be succesful. But the opposite of the above argument is also true: if one person is ready to answer, so are the others.

The characteristics of the households where the rate of refusal was higher than the average:

- households in Budapest and in the surrounding settlements
- economically active households
- private (small) entrepreneurs
- highly educated people (relatively high social position)
- households with good housing
- large households

Table 2: Main household types with nonresponse rates significantly above and below average

Groups with significantly low nonresponse rate	
Demographical composition	Households with old persons only (4.5%)
Educational level of head of hh.	Less than elementary school (5.1%)
Economic type of hh.	Inactive hh. (7.0%)
Quality of dwelling	Very bad quality of dwelling (9.8%)
Area type	Small villages (11.2%)
Groups with significantly high nonresponse rate	
Economic status of head of hh.	Self-employed, with more than 20 employees (45.2%)
Quality of dwelling	Very good quality of dwelling (33.9%)
Educational level of head of hh.	University degree (30.1%)
Area type	Budapest (28.5%)
Number of active earners	3 or more adults in employment (24.5%)

In the case of refusals the following characteristics can be observed:

- the higher the educational attainment (social status) the higher the nonresponse rate,
- the nonresponse rate is higher in Budapest, lower in towns and the smallest in the villages,
- the older generation (pensioners) is generally ready to answer
- the response of the one member households is generally high but it depends on their age.

Figure 1: Refusal rates of HIS by age composition of the households,%

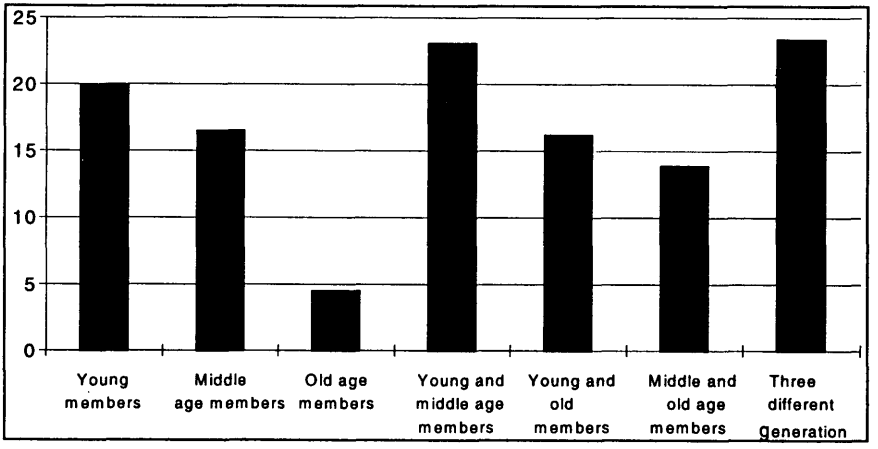


Figure 2: Refusal rates of HIS by educational attainment of the head of the households, %

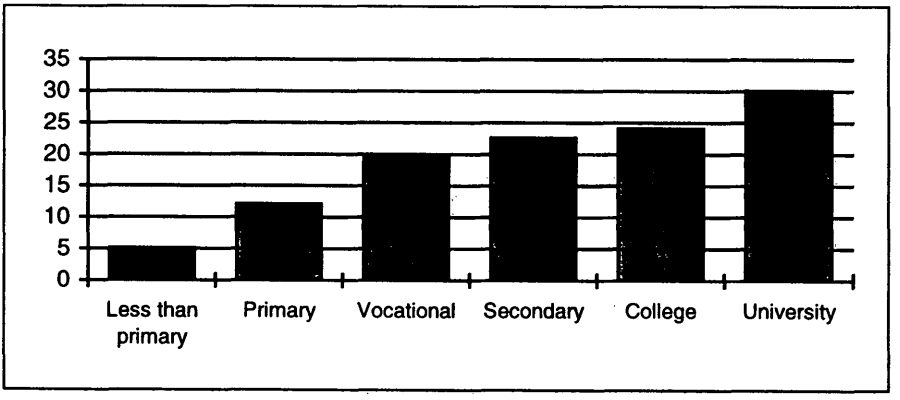


Figure 3: Refusal rates of HIS by type of occupation of active earners

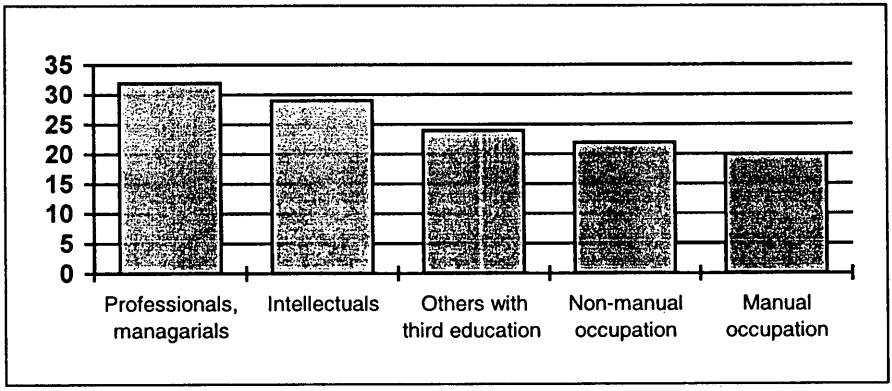
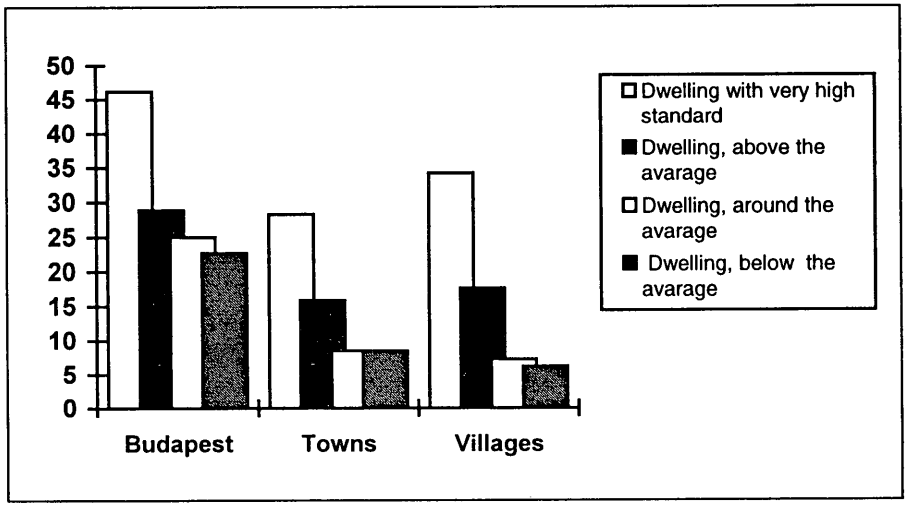


Figure 4: Refusal rates of HIS by type of settlements and characteristics of dwellings, %



Based on this information

- further studies should be carried out to try to establish better information on the behaviour of the nonrespondents with different characteristics. (The same type of study is going on in the case of the survey of travelling habits. That will add more information on how refusals depend on the sensitiveness of the topic.)
- the imputation techniques can be improved by the use of the nearest neighbour methods. This seems to be more effective than simple reweighting on the basis of a few demographic characteristics. (The results, the evaluation of the effectiveness of the imputation will be available soon.)

Some remarks on the comparison of the 1995 English FES and the 1996 Hungarian income supplement. The most important differences (Table 3).

Table 3: Main household types with high nonresponse rate in the UK and Hungary ***(the average nonresponse rate of the country is 1[4])***

Household characteristics	UK	Hungary
3 or more wage earner adults	1.52	1.48
Couples with non-dependent child(ren) only	1.49	1.55
Head belonging to ethnic minority	1.45	-
3 or more adult members	1.33	1.13
Head self-employed	1.18	2.09
Head without post-school qualification	1.05	0.70

Conclusions

Up to the early 1990's the rate of nonresponse in household surveys was not a great concern within the CSO. We enjoyed good and stable response rates in practically all surveys except the FES as a result of its heavy burden on respondents.

Recently the problems of coverage (not at home, wrong address, empty dwellings etc.) and refusal are getting more serious.

In Hungary, like in the other East-Central European (former socialist) countries the social

[4] K. Foster 1995

climate has radically changed, which had a negative impact on the quality of survey data: higher nonresponse, bias, etc. That is reflected in the fairly high refusal rate (4.2%) in the microcensus, which is mandatory, and contains no sensitive data.

The response rate strongly depends on the topic (sensitiveness) of the survey and also on the changing social status: the quickly growing private sector, especially the numerous small private businesses. The black (grey) economy is very strong.

Sometimes the CSO is criticized due to the unfavourable economic developments (high inflation rate, unemployment, not reliable register of small businesses), although the CSO strictly follows the very rigorous laws on privacy, the use of individual data, and does everything possible to provide timely and reliable data.

Finally it has to be mentioned that among the new circumstances more and more financial and human resources are needed for call-backs, follow-ups, greater samples etc., which are in many cases not available.

References

Foster, K. (1995). Weighting the Family Expenditure Survey in Great Britain to compensate for nonresponse: an investigation using census-linked data. Paper presented on the Sixth International Workshop on Household Survey Nonresponse, Helsinki, October 1995

Lillard, L., Smith, J.P. and Welch, F. (1986). What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation. Journal of Political Economy, vol. 94, no. 31, pp. 489-503

Lopdell, J. and Ramsay, D. (1996). Emerging issues in household survey non-response in statistics New Zealand. Paper presented on the Seventh International Workshop on Household Survey Nonresponse, Rome, October 1996

Marton, A. (1995). Nonresponse in the Hungarian household surveys. Paper presented on the Sixth International Workshop on Household Survey Nonresponse, Helsinki, October 1995

Waite, P.J. and Huggins, V.J. (1996). Nonresponse research in the 1996 survey of income and program participation (SIPP): Monetary incentives and a nonrespondent study. Paper presented on the Seventh International Workshop on Household Survey Nonresponse, Rome, October 1996

The Stability of Nonresponse Rates According to Socio-Demographic Categories

METKA ZALETEL AND VASJA VEHOVAR

Abstract: *The data from various sources (e.g. Central Population Register, Census data, Tax Register, Register of Territorial Units, etc.) were merged at the level of enumeration areas for the whole territory of Slovenia. The nonresponse and refusal rates of face-to-face surveys were connected to this model. The paper shows the stability of responses according to the given predictors.*

1 Introduction

During the last years, a lot of research has been conducted about reasons for the nonresponse in household surveys. Some of the results can be generalised to different countries, but some of them are really country - or region - specific. A typical example is the case of families with young children, which are generally more likely to respond; that was not proven to be true in Slovenia. On the other hand, there is a result which holds more generally: people living in the same neighbourhood or village do behave similarly when they get a request for the interview. Usually there is only one interviewer involved in a survey in a certain neighbourhood, so the interviewer's influence cannot be separated from the responding behaviour within a neighbourhood.

The paper analyses response rates at the level of primary sampling units and then constructs a model with explanatory variables at the same level, which enables us to predict response rates at the level of primary sampling units in the future. We first summarise the past research of nonresponse in Slovenia and its results in Section 2, then we explain the background and motivation of the present analysis in Section 3. In Sections 4 and 5 we first introduce all available data and then explain the methods used for the analysis; after this, the results are presented.

2 Past research and results

We have been searching for the reasons for nonresponse in Slovenia almost since the establishing of the country as an independent state in 1991. The nonresponse and refusal rates started to grow from the very beginning and, unfortunately, the growth has not stopped yet. In the case of the Labour Force Survey, we notice approximately 2% growth of nonresponse rate per year. With re-designing of the survey in 1997 and with employment of a small number of experienced interviewers we tried to stop the growth, but at the same time we introduced CAPI which - only in some regions - pushed it up again. Similar development is noticed in other surveys independent of the organisation - Statistical Office, academic organisations or marketing companies.

First, it was proven that substitutions cause biased estimations (Vehovar, 1993). At that time, substitutions were widely used in all kinds of surveys.

In 1995, the matching of persons selected in different samples with all available data sources was performed (Vehovar and Zaletel, 1995). The matching was possible because of the personal identification number that every person in Slovenia has. We matched the results of two surveys (Labour Force Survey and Household Budget Survey) with Census'91, Central Population Register, Database on Employed Persons, Taxation Register and Register of Unemployed Persons. We found that the results cannot be generalised across surveys since the Labour Force Survey gave us different predictors of response behaviour then the Household Budget Survey. Some interactions of two variables also appeared as predictors but they also cannot be generalised. The main predictors were education, income and type of dwelling. At the same time we estimated bias of some of the main estimates from both surveys. In some cases, the relative bias was up to 5% of the estimate.

In 1996, research of impact of confidentiality concern was made (Vehovar and Zaletel, 1996). The results are rather surprising: in general, people are not aware of confidentiality issues and possible abuse of survey data. The impact of confidentiality concern on nonresponse rates was thus not proven. There are certain topics which make some groups of people worried but we could not confirm that they are less likely to respond.

3 Motivation and background

Almost all samples of official surveys in Slovenia have the same sampling design - they are two stage stratified samples and primary sampling units are usually enumeration areas. The post-survey adjustment for nonresponse is also quite similar for most of the surveys: weights are calculated at the level of primary sampling units. If adjustment is done at the level of enumeration areas, perhaps we can also predict the nonresponse at the level of

enumeration areas. Another motivation for this idea are certainly the results from some of the countries (e.g. King, 1996) where the division of the country into small areas according to the socio-economic variables was made in advance. Then it was proven that the nonresponse rates vary across socio-demographic types of areas. We decided to generalise the idea: to build socio-demographic types of enumeration areas according to nonresponse rates achieved in some of the official surveys. This model would enable us to predict nonresponse rates for similar surveys in the future.

There are about 14,000 enumeration areas (EA) in Slovenia with 45 households each on average. Unfortunately, some of the EAs are very small or even empty, especially in remote areas. This fact caused a lot of problems in the process of sample designing and selection. In 1996, we merged all small EAs with their larger neighbours. We ended up with 9,872 clusters of enumeration areas (CEA) with an average of 65 households. The problem of small EAs vanished almost completely. Since 1996, primary sampling units in the majority of official surveys are CEAs.

4 Surveys and administrative data sources

To build a model for prediction of response rates at the level of enumeration areas, we selected some of the major (and largest) official surveys in Slovenia on one side to get data on achieved response rates at the level of enumeration areas and the register and census data on the other side to get socio-demographic data. In this section, we will first describe the surveys, then the administrative data and finally the variables selected from these rich sources of data.

4.1 Surveys

We included the following surveys:

- Labour Force Survey (LFS) 1994, 1995, 1996: this survey was conducted annually in May every year. Sample sizes were approximately 8,000 households per year. The whole fieldwork organisation was very similar from year to year: five follow-ups, advance letters, about 140 free-lance interviewers, face-to-face surveys in PAPI mode. Average length of interview was 18 minutes. The nonresponse rates were as follows: 8.9% in 1994, 9.0% in 1995 and 10.1% in 1996.
- Household Budget Survey (HBS) 1993, 1994, 1995, 1996: this survey was also conducted annually in December every year. The sample size in 1993 was 4,500 households, sample sizes in years from 1994 to 1996 were about 1,400 households. The fieldwork organisation was similar to that of the LFS with the exception of the

number of interviewers. In 1993, there were 109 free-lance interviewers. In later surveys, about 30 interviewers were involved. Average length of interview was about 90 minutes. The nonresponse rates were as follows: 19.7% in 1993, 17.8% in 1994, 18.0% in 1995 and 34.6% in 1996.

- Household Survey on Energy and Fuel Consumption (HSEFC): the survey was conducted for the first time in Slovenia in May 1997. The sample size was 5,000 households. The fieldwork was not organised by the Statistical Office of the Republic of Slovenia as for other surveys, but the organisation of the fieldwork was very similar. The number of interviewers was about 100. Average length of interview was 23 minutes. The nonresponse rate was 17.9%.

4.2 Administrative data sources

All major administrative data sources available at the Statistical Office of the Republic of Slovenia and some other organisations were used:

- Central Population Register (CRP)
- Census '91 database
- Database on Employed Persons in the Republic of Slovenia (DEP)
- Register of Territorial Units (RTU)
- Telephone Database (TD)

At this stage of research, the Taxation Register (maintained by the Ministry of Finance) has not been included in the estimations, but when the TR is available, the model will be re-estimated.

There are two important points which need to be stressed here. During the Census in 1991, there was a centroid determined for every building in Slovenia. Later, also the height above sea level of every building was estimated. According to these data, we defined the centroid for each CEA and for each settlement.

The second point concerns the time distance from the Census '91. All the data from the Census are obviously now 6 years old, but we took from the Census mostly data on dwellings and migrations. The Slovenian population is very stable and only about 2% of population is moving per year. In fact, most of those 2% are migrations within the same towns or villages. The situation concerning dwellings has not changed much in Slovenia since 1991 because not a lot of new dwellings have been build in-between. We can assume that Census data are good enough for our purposes.

4.3 Independent variables

First of all, we defined five sets of variables, concerning (1) persons, (2) dwellings, (3) households, (4) settlements and (5) clusters of enumeration areas. Then we re-calculated all these variables at the level of clusters of enumeration areas. We started the estimation of the model with the following variables:

Table 1: Available variables and their sources

		data source				
set	variable	CRP	Census	DEP	RTU	TD
1	proportion of children under 15 years					
1	proportion of persons over 65 years					
1	proportion of employed persons					
1	proportion of persons with higher education					
2	proportion of privately owned dwellings					
2	proportion of dwellings in apartment buildings					
2	proportion of weekend or summer houses					
3	proportion of farming households					
4	proportion of migration for school or work out of the settlement					
4	if the settlement is a centre of municipality or not					
4	type of settlement					
5	density of population					
5	air distance from centre of municipality					
5	telephone coverage					

4.4 Dependent variable

The first and natural idea for the selection of the dependent variable was the response rate at the level of enumeration areas. After merging all data we realised that there are some problems with data from Household Budget Survey 1993. We were able to define the initial sample size and responses for each CEA, but that was not the case for ineligible persons. In every survey we usually experience about 5% of ineligible households because of some differences between de-iure and de-facto addresses of those persons. After some investigation of the problem we concluded that this problem is equally spread all over the country and that results for the response rate and the completion rate (i.e.

number of responses divided by number of initial sample) are the same. So we simplified the problem and took the completion rate at the level of CEA as a dependent variable.

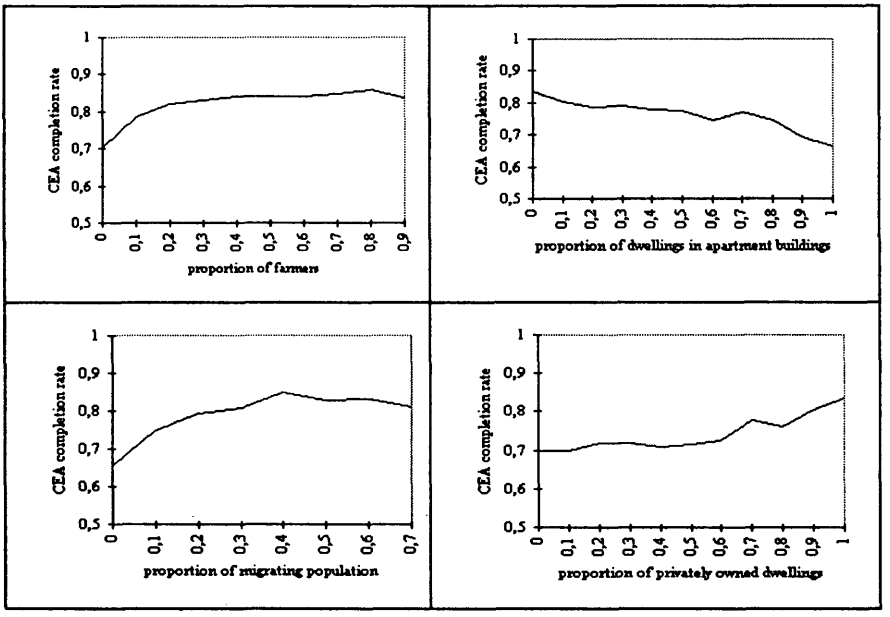
5 Analysis and results

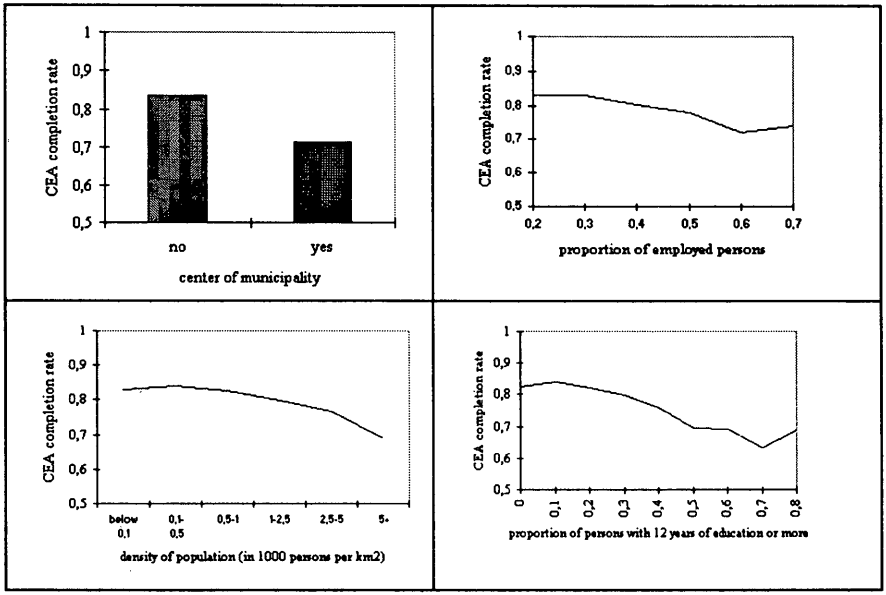
Firstly, we observed the dependence of completion rates on separate variables. Secondly, we took into account also the topic of the survey. Finally, the linear regression model was estimated.

5.1 Completion rates

Let us first observe a few figures presenting the dependence of CEA completion rates on selected variables. We calculated general completion rates irrespective of the survey.

Figure 1: CEA completion rates according to selected variables





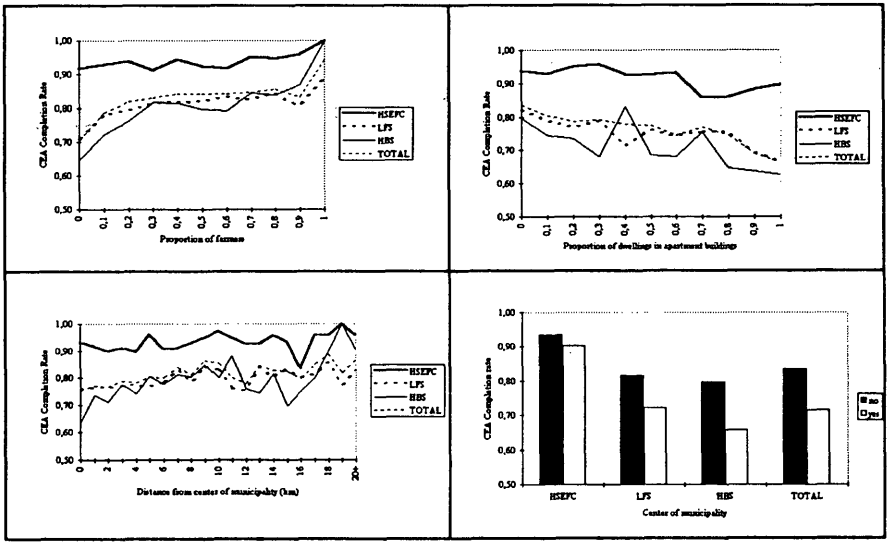
We can see that there exists a dependence of completion rates on most of the displayed variables. Here we do not present any numerical tables showing the dependence. There are two questions appearing right away:

1. Are the presented results survey dependent?
2. Is there any interaction between two variables?

5.2 Completion rates across different surveys

We notice that the HSEFC is behaving very differently in comparison with the other two surveys which are very similar. The same picture would be given with other variables which are not shown here. Even before the estimation of the model we can expect that we have to estimate separate models for each of the surveys included. At the same time we can say that the model for the HSEFC will not explain a lot of variability in completion rates. But let us first have a look at the estimation of the models.

Figure 2: CEA completion rates across different surveys



5.3 Regression model

The estimation of the regression model has shown what we expected and predicted according to the results of the previous section: the results cannot be generalised independently of the survey topic. Another result seen on the figures above was proved: available variables do not explain the variability in completion rates for the HSEFC at all.

In the table below we labelled the variables which were significant in the regression model. The level of significance is 0.05.

We can see that more or less the same variables are significant in the models for the LFS and the HBS. Only one variable is significant for the HSEFC, but even this one does not explain any variability of completion rates.

Table 2: Regression coefficients across different surveys

set	variable	HBS	LFS	HSEFC
1	proportion of children under 15 years			
1	proportion of persons over 65 years			
1	proportion of employed persons	-0.19		
1	proportion of persons with higher education	-0.20	-0.04	
2	proportion of privately owned dwellings			
2	proportion of dwellings in apartment buildings	-0.07	-0.11	
2	proportion of weekend or summer houses			
3	proportion of farming households		0.02	
4	proportion of migration for school or work out of the settlement	0.03		0.12
4	if the settlement is a centre of municipality or not	0.06	0.03	
4	type of settlement	-0.02	-0.02	
5	density of population			
5	air distance from centre of municipality			
5	telephone coverage			
	intercept	1.02	0.86	0.87

6. Conclusions and future work

The paper shows the modelling of completion rates at the level of clusters of enumeration areas according to given socio-demographic variables. We used data from three different official surveys and five different administrative sources. The achieved results somehow confirmed expectations we had according to similar research in the past:

- the results depend strongly on the survey topic and cannot be generalised;
- the key variables are education, type of dwelling and type of settlement;
- data on distances and population density are not important at this stage of research.

The research on enumeration areas will proceed with the taxation data added to the model. Income proved to be a very important explanatory variable in the past research. Data on the average number of contacts per household will also be added which will enable the estimation of costs of the survey in certain enumeration areas in advance.

References

Groves, R.M. (1989). Survey Errors and Survey Costs. New York. Wiley

Groves, R.M. and Couper, M.P. (1993). Correlates of Nonresponse in Personal Visit Surveys

King, J. (1996). The use of geo-demographic coding schemes for understanding household nonresponse. 7th International Workshop on Household Surveys Nonresponse. Rome, October 1996

Vehovar, V. (1993). Field Substitutions in Slovene Public Opinion Survey. Contributions to Methodology and Statistics. Ljubljana, FDV, pp. 39-66

Vehovar, V. and Zaletel, M. (1995). The Matching Project in Slovenia - Who are the Nonrespondents? Sixth International Workshop on Household Surveys Nonresponse. Helsinki, October 1995

Vehovar, V. and Zaletel, M. (1996). Does the confidentiality concern increase the nonresponse rate? 7th International Workshop on Household Surveys Nonresponse. Rome, October 1996

Understanding Household Survey Nonresponse through Geo-demographic Coding Schemes[1]

JOHN KING

Abstract: *A geo-demographic coding system was used to analyse response to a large household survey. This provided a contemporaneous indicator of response and of nonresponse to the survey using terms external to the survey. Geo-demographic coding information is used to describe the non-respondents to the survey. Use of this information for reducing some fieldwork problems is also indicated.*

Keywords: *household sample surveys; survey nonresponse; geo-demographic coding systems; non-respondent characteristics; Mosaic.*

1 Background

Geo-demographic coding systems have been used extensively for marketing and planning. In partnership with Experian, owners of the Mosaic geo-demographic coding system, Mosaic codes were attached to the eligible sample drawn for the British Family Expenditure Survey (FES) for 1995-96. Sample response and nonresponse was examined using these codes. Usually response to surveys is analysed either by sample design or fieldwork characteristics; or by socio-economic descriptors. The geo-demographic codes have given a description of survey nonresponse using an alternative set of descriptors. A simple, but basically multi-factor, picture of respondents and non-respondents can be obtained using geo-demographic coding schemes.

A previous paper (King 1996) indicated ways in which these schemes could be used to describe survey respondents. This paper extends those analyses to the characteristics of the eligible drawn sample. This leads to a description, in geo-demographic terms, of the non-respondents as well as of the respondents.

[1] Acknowledgements: FES data are Crown copyright and are used with permission of the Office for National Statistics, United Kingdom. Mosaic is a product of Experian and is used here with permission. The views expressed are those of the author and do not necessarily represent those of the Office for National Statistics.

2 Geo-demographic coding schemes

Geo-demographic coding schemes have been developed over the last couple of decades to help with analyses in several subject areas; particularly in market research, and especially targeted marketing, and in urban social geography and planning. There are several major schemes available in Great Britain based on the 1991 Census of Population and Housing. A brief description of these, and the principles upon which they are built, is given by Openshaw (1993).

The coding of households by geo-demographic coding schemes is not exact, in the sense that the characteristics of the individual household are not used. First, the coding scheme uses information from different geographic levels. Some information may be specific to the unit postcode, some information may be averaged over several adjacent unit postcodes and other information may refer to a wider geographic area. Second, all households with a particular postcode will be given the same code; no allowance is made for the characteristics of the individual household. By using the geo-demographic codes in this study, it is assumed that the codes do adequately describe the responding households.

The system used for this study, and the previous one, is the Mosaic system: a product of Experian which is used here with their permission. The Mosaic system is based on 86 variables and is described in more detail in the earlier paper and in The Multimedia Guide to Mosaic (Experian 1997). Essentially, about half of the input variables are Census variables (or census-based) which are available at the level of the census Enumeration District; other variables, including some financial information and data on accessibility to shopping centres, are available at the level of the postcode.

The Mosaic classification system has 52 household types. These are grouped into 12 lifestyle groupings as shown in Annex A. Households are allocated a Mosaic type using only the postcode of the household.

3 The survey and attaching the codes

The Family Expenditure Survey (FES) is a continuous survey of private households conducted in the United Kingdom by the Office for National Statistics (ONS). Detailed information about the expenditure of all those aged 16 and over is obtained through interviews and a diary kept for a period of 2 weeks. Information about income of individuals in the household is also obtained through interviews, along with information about employment, receipt of state security benefits and ownership of consumer durable goods. Only the sample for Great Britain has been used for this study.

In Great Britain the sampling frame is the small-user Postcode address file (PAF) maintained by the Post Office. The design is a multi-stage stratified systematic sample with clusters. The sample is stratified by Standard Statistical Region (SSR) and by three further variables: metropolitan/non-metropolitan areas (the latter are further split into 3 strata of low, medium and high population density); the proportion of owner-occupiers and the proportion of privately renting households.

The Social Survey Division (SSD) of ONS draws the FES sample and undertakes the fieldwork in Great Britain. Information is collected under a promise of confidentiality. To ensure this, any names and addresses are removed before processing, and records are further anonymised before being released as micro-data. Attaching a geo-demographic code is not thought to compromise this undertaking.

The Mosaic code for a household is determined solely by its postcode. SSD used a directory of Mosaic codes provided by Experian to Mosaic-code the drawn FES sample. This was subsequently modified to a list of the eligible sample by excluding ineligible addresses and adding those additional households sampled at addresses where multiple households were found. This list also showed the outcome of the fieldwork: response, refusal or non-contact.

4 The FES sample by geo-demographic codes

Table 1 shows the eligible sample by Mosaic main lifestyle groupings for 1995-96 and also the composition of the household population in Mosaic terms. The distribution of FES households across these groups is compared with the distribution of all households. The Index in the final column of Table 1 is the ratio of FES households to all households: an Index value of 100 means that there was the same proportional representation in the FES as in the population.

Differences between the two distributions are attributable to several factors. One factor is the slightly varying definition of a household: the FES uses the pre-1981 census household definition, based on sharing common housekeeping and meals; while the Mosaic system was built using the 1981, and subsequent, household definition based only on the criterion of shared accommodation. Estimates of the distribution of households by Mosaic type are updated using Electoral Roll and PAF data. The differences created by the change in Census definitions are thought to be small (Todd and Griffiths 1986). Other differences are due to sampling variability and to changes over time from the 1991 Census basis of some of the variables underlying the Mosaic classification system.

Table 1: Family expenditure survey 1995-96 - Great Britain The eligible sample by Mosaic codes

Lifestyle grouping		Number of eligible FES households	Percentage of households in FES	Mosaic	Index
L1	High Income Families	1171	11.7	9.9	118
L2	Suburban Semis	1162	11.6	11.0	105
L3	Blue Collar Owners	1297	12.9	13.0	99
L4	Low Rise Council	1482	14.7	14.4	102
L5	Council Flats	695	6.9	6.8	102
L6	Victorian Low Status	822	8.2	9.4	87
L7	Town Houses & Flats	974	9.7	9.4	103
L8	Stylish Singles	489	4.9	5.2	94
L9	Independent Elders	720	7.2	7.4	97
L10	Mortgaged Families	550	5.5	6.2	88
L11	Country Dwellers	654	6.5	7.0	93
L12	Institutional Areas	29	0.3	0.3	96
	not Mosaic coded	109			
	Total	10154			

5 Response by geo-demographic codes - the results

Table 2 shows the number of households in the FES in 1995-96 in Great Britain in each of the Mosaic main lifestyle groupings by the outcome of the fieldwork: full response, refusal and non-contact. Annex B shows similar information for each of the 52 Mosaic types.

Table 2 shows that the response rate for the FES in GB in 1995-96 was 65 per cent. Nearly all the nonresponse was due to refusals rather than non-contact. Refusals were 32 per cent and non-contacts were 3 per cent.

Table 2: Family expenditure survey 1995-96 - Great Britain Analysis of the results of the fieldwork by Mosaic codes

Lifestyle grouping		Number of eligible FES households			
		responding	refusing	non-contacted	total
L1	High Income Families	776	376	19	1171
L2	Suburban Semis	781	364	17	1162
L3	Blue Collar Owners	843	422	32	1297
L4	Low Rise Council	991	457	34	1482
L5	Council Flats	411	245	39	695
L6	Victorian Low Status	494	280	48	822
L7	Town Houses & Flats	630	322	22	974
L8	Stylish Singles	282	164	43	489
L9	Independent Elders	465	237	18	720
L10	Mortgaged Families	386	146	18	550
L11	Country Dwellers	458	190	6	654
L12	Institutional Areas	24	4	1	29
not Mosaic coded		70	38	1	109
	Total	6611	3245	298	10154

Table 3 shows the distributions across the Mosaic groupings for the respondents and non-respondents. The distribution of responding households differs from the eligible sample shown in Table 1 mainly in that there were significantly fewer responding households in groups L5 Council Flats, L6 Victorian Low Status and L8 Stylish Singles. This is a similar pattern to that found in the analysis of respondents to the 1994-95 FES (King 1996).

Table 3 also shows the characteristics of the non-responding households. This is direct description of these households. There are proportionately more non-responding households in groups L5, L6 and L8, corresponding to the under-representation of these groups in the responding households. There are also proportionately fewer households in the groups L2 Suburban Semis, L4 Low Rise Council, L10 Mortgaged Families and L11 Country Dwellers. Only four groups have a representation similar to that in the eligible sample.

The type of nonresponse - either refusal or non-contact - is known for each household. Table 3 also shows the characteristics of household by reason for nonresponse. Non-contact is low on the FES and the number of households not contacted is small. This is because of the efforts of the interviewers and also because the design of the sample and the field work permits several calls. But although the number of these households is relatively low, the characteristics of these are very different from those of the refusing households and of the eligible sample. In particular, again, households in groups L5, L6 and L8 are proportionately over-represented. These categories are known to be problematic: the access to many council flats is difficult because of the security arrangements; and stylish singles are often difficult to find at home.

Table 3: Family expenditure survey 1995-96 - Great Britain Response and nonresponse by Mosaic codes

		Percentage of eligible FES households			
Lifestyle grouping		**responding**	**nonresponse (total)**	**refused**	**no contact**
L1	High Income Families	11.9	11.3	11.7	6.4
L2	Suburban Semis	11.9	10.9	11.4	5.7
L3	Blue collar Owners	12.9	13.0	13.2	10.8
L4	Low Rise Council	15.2	14.0	14.3	11.4
L5	Council Flats	6.3	8.1	7.6	13.1
L6	Victorian Low Status	7.6	9.4	8.7	16.2
L7	Town Houses & Flats	9.6	9.8	10.0	7.4
L8	Stylish Singles	4.3	5.9	5.1	14.5
L9	Independent Elders	7.1	7.3	7.4	6.1
L10	Mortgaged Families	5.9	4.7	4.6	6.1
L11	Country Dwellers	7.0	5.6	5.9	2.0
L12	Institutional Areas	0.4	0.1	0.1	0.3
	Total Mosaic coded	6541	3504	3207	297

Table 4 shows another way of looking at the data. For each Lifestyle grouping the row shows the proportions of households responding, refusing and not being contacted. Thus, L10 and L11 are groups with higher response. The groups most difficult to contact are L5, L6 and L8. These have only a slightly higher refusal rate than other household types: the main difficulty is in contacting them. Contrary to general beliefs, L1 High Income Families are not poor participants in the FES. Their response rate is about average, and higher than that of several other groups.

Table 4: Family expenditure survey 1995-96 - Great Britain Response and nonresponse patterns by Mosaic codes

Lifestyle grouping		Percentage of eligible FES households responding	refusing	no contact	Total (number)
L1	High Income Families	66	32	2	1171
L2	Suburban Semis	67	31	1	1162
L3	Blue collar Owners	65	33	2	1297
L4	Low Rise Council	67	31	2	1482
L5	Council Flats	59	35	6	695
L6	Victorian Low Status	60	34	6	822
L7	Town Houses & Flats	65	33	2	974
L8	Stylish Singles	58	34	9	489
L9	Independent Elders	65	33	3	720
L10	Mortgaged Families	70	27	3	550
L11	Country Dwellers	70	29	1	654
L12	Institutional Areas	83	14	3	29
	Total Mosaic coded	65	32	3	10045

6 Comparisons with other analyses

Foster (1996) analysed FES respondents and non-respondents using certain census variables, and described the factors most closely affecting nonresponse in terms of these variables. This was based on an exercise matching the FES sample with individual household records from the 1991 Census of Population and Housing. The number of adults in the household, especially where there were no children under the age of 16, was the variable most closely associated with response, probably because of the strict response rules for the FES. Other characteristics linked with higher nonresponse were: if the household was in London; if the household had a head with no post-school qualifications, was born outside the UK or was self-employed. Also, nonresponse was associated with households with no dependent children or with a head of household aged 35 or more.

Because the categories used for Foster's analysis were different from those used in this study it is not possible to make direct comparisons between the conclusions of the studies. However, there are some indications of these results. The high representation of L5 and

L6 groups in the non-responding households is reflected in the finding about London and other metropolitan areas. Similarly, the higher representation of L8 is part of the description of households with no dependent children.

Analysis by geo-demographic coding schemes will not replace the need for Census comparisons to understand the true picture of response and nonresponse. But such studies are infrequent and costly. By comparison, geo-demographic coding analyses are cheap and very fast. They could be used to give a cheap means of monitoring trends in nonresponse. Monitoring trends would provide checks on the continued appropriateness of the assumptions made when using Census-based studies several years after the Census date.

A more recent study of nonresponse to the FES has been undertaken by Hansbro and Foster (1997). This study reports on an analysis of a nonresponse questionnaire (NRQ). The NRQ is used by SSD to obtain as much information as possible about non-responding households on certain key descriptive variables. Again, the analysis is in terms of standard, Census-like, socio-economic descriptors. Much of the picture emerging from this study is similar to the Census comparison and to the implicit results of the geo-demographic code analysis.

7 Value and uses of geo-demographic analyses

A particular advantage of geo-demographic coding is that information is available quickly and relatively cheaply. There is a built-in picture of nonresponse as the data flows in from the field. It is thus possible to monitor levels and trends in response immediately. Month by month tracking of trends would allow problems to be seen and even anticipated.

Geo-demographic coding information for each sampled household is, of course, available as soon as the sample is drawn. This means it can be used by fieldwork organisers to anticipate the likely difficulty of a workload. Allocation of interviewers can be targeted; matching interviewers with particular skills to particular workloads - the nature of which will be better known.

Interviewer performance and variability can also be understood in the context of the actual workload and the difficulty of it. Geo-demographic codes provide a more objective measure of the difficulties and problems that the interviewers may have faced.

8 The future

We hope to look further at the stability of these analyses over time. We hope to look at the bias resulting from nonresponse in both expenditure and income patterns. We also hope to compare in more detail these results with the traditional studies mentioned above.

References

Experian (1997). The Multimedia Guide to MOSAIC. Nottingham: Experian

Foster, K. (1996). A comparison of the census characteristics of respondents and non-respondents to the 1991 Family Expenditure Survey. London: Office for National Statistics; Survey Methodology Bulletin, no 38, January 1996

Hansbro, J. and Foster, K. (1997). Characteristics of nonresponse on the 1995-96 Family Expenditure Survey. London: Office for National Statistics; Social Survey Division

King, J. (1996). The use of geo-demographic coding schemes for understanding household survey nonresponse. Seventh International Workshop on Household Survey Nonresponse. Rome, October 1996

Openshaw, S. (1993). Special Classifications. Chapter 7 in: B. Leventhal, C. Moy and J. Griffin (eds.). An Introductory Guide to the 1991 Census. Henley-on-Thames: NTC Publication Ltd in association with the Market Research Society

Todd, J. and Griffiths, D. (1986). Changing the definition of a household. London: Office for National Statistics; Social Survey Division

Annex A
Structure of the Mosaic coding system

- L1 High Income Families
  - M1 Clever Capitalists
  - M2 Rising Materialists
  - M3 Corporate Careerists
  - M4 Ageing Professionals
  - M5 Small Time Business
- L2 Suburban Semis
  - M6 Green Belt Expansion
  - M7 Suburban Mock Tudor
  - M8 Pebble Dash Subtopia
- L3 Blue Collar Owners
  - M9 Affluent Blue Collar
  - M10 30s Industrial Spec
  - M11 Lo-Rise Right to Buy
  - M12 Smokestack shiftwork
- L4 Low Rise Council
  - M13 Coop club and Colliery
  - M14 Better off Council
  - M15 Low Rise Pensioners
  - M16 Low Rise Subsistence
  - M17 Problem Families
- L5 Council Flats
  - M18 Families in the Sky
  - M19 Graffitied Ghettos
  - M20 Small Town Industry
  - M21 Mid Rise Overspill
  - M22 Flats For The Aged
  - M23 Inner City Towers
- L6 Victorian Low Status
  - M24 Bohemiam Melting Pot
  - M25 Victorian Tenements
  - M26 Rootless Renters
  - M27 Sweatshop Sharers
  - M28 Depopulated Terraces
  - M29 Rejuvenated Terraces
- L7 Town Houses & Flats
  - M30 Bijou Homemakers
  - M31 Market Town Mixture
  - M32 Town Centre Singles
- L8 Stylish Singles
  - M33 Bedsits and Shop Flats
  - M34 Studio Singles
  - M35 College and Communal
  - M36 Chattering Classes
- L9 Independent Elders
  - M37 Solo Pensioners
  - M38 High Spending Greys
  - M39 Aged Owner Occupiers
  - M40 Elderly in Own Flats
- L10 Mortgaged Families
  - M41 Brand New Areas
  - M42 Pre Nuptial Owners
  - M43 Nestmaking Families
  - M44 Maturing Mortgagees
- L11 Country Dwellers
  - M45 Gentrified Villages
  - M46 Rural Retirement Mix
  - M47 Lowlands Agribusiness
  - M48 Rural Disadvantage
  - M49 Tied / Tenant Farmers
  - M50 Upland and Small Farm
- L12 Institutional Areas
  - M51 Military Bases
  - M52 Non Private Housing

Mosaic codes are used by permission of Experian.

Annex B
Family expenditure survey 1995-96 - Great Britain
Response type by Mosaic codes

	Description	Percentage of eligible FES households				
		respondents	non-respondents	refusals	no contact	Total (number)
M1	Clever Capitalists	1.4	1.8	1.7	2.0	153
M2	Rising Materialists	2.8	1.9	1.9	2.0	247
M3	Corporate Careerists	2.9	2.9	3.1	0.7	291
M4	Ageing Professionals	1.6	1.7	1.8	0.7	165
M5	Small Time Business	3.2	3.0	3.2	1.0	315
M6	Green Belt Expansion	3.9	2.6	2.7	1.7	347
M7	Suburban Mock Tudor	3.1	3.0	3.1	2.4	310
M8	Pebble Dash Subtopia	4.9	5.2	5.6	1.7	505
M9	Affluent Blue Collar	2.9	3.5	3.6	2.4	312
M10	30s Industrial Spec	4.0	4.0	4.1	2.4	398
M11	Lo-Rise Right to Buy	3.0	3.1	3.0	3.7	304
M12	Smokestack shiftwork	3.0	2.4	2.4	2.4	283
M13	Coop club and Colliery	3.6	3.1	3.3	1.0	341
M14	Better off Council	2.0	2.0	2.0	2.4	204
M15	Low Rise Pensioners	4.0	3.9	4.1	1.7	399
M16	Low Rise Subsistence	3.7	3.3	3.1	4.7	357
M17	Problem Families	1.8	1.7	1.7	1.7	181
M18	Families in the Sky	1.2	1.8	1.6	4.7	141
M19	Graffitied Ghettos	0.3	0.3	0.2	0.3	27
M20	Small Town Industry	1.1	1.2	1.2	1.0	115
M21	Mid Rise Overspill	0.6	0.6	0.6	0.3	58
M22	Flats for the Aged	1.5	1.6	1.7	1.0	152
M23	Inner City Towers	1.7	2.6	2.3	5.7	202
M24	Bohemiam Melting Pot	1.8	3.0	2.6	7.1	221
M25	Victorian Tenements	0.0	0.2	0.1	1.0	7
M26	Rootless Renters	0.8	0.9	0.9	1.3	87
M27	Sweatshop Sharers	0.9	0.9	0.9	1.7	92
M28	Depopulated Terraces	0.6	0.8	0.9	0.3	70
M29	Rejuvenated Terraces	3.4	3.5	3.3	4.7	345
M30	Bijou Homemakers	3.4	3.9	4.0	2.4	358
M31	Market Town Mixture	4.2	3.5	3.7	1.7	397

	Description	Percentage of eligible FES households				
		respondents	non-respondents	refusals	no contact	Total (number)
M32	Town Centre Singles	2.0	2.4	2.3	3.4	219
M33	Bedsits and Shop Flats	0.7	1.1	0.9	2.7	82
M34	Studio Singles	1.3	2.4	2.0	7.4	172
M35	College and Communal	0.5	0.4	0.3	1.3	45
M36	Chattering Classes	1.8	2.0	1.9	3.0	190
M37	Solo Pensioners	2.0	2.1	2.1	2.4	206
M38	High Spending Greys	0.9	0.9	1.0	0.3	93
M39	Aged Owner Occupiers	2.9	2.5	2.5	1.7	278
M40	Elderly in Own Flats	1.3	1.7	1.7	1.7	143
M41	Brand New Areas	0.2	0.2	0.2	0.3	20
M42	Pre Nuptial Owners	1.4	0.6	0.5	1.3	113
M43	Nestmaking Families	1.8	1.6	1.5	2.0	174
M44	Maturing Mortgagees	2.5	2.3	2.3	2.4	243
M45	Gentrified Villages	1.5	1.5	1.6	0.0	150
M46	Rural Retirement Mix	0.5	0.3	0.2	0.3	43
M47	Lowlands Agribusiness	1.6	1.7	1.7	1.3	164
M48	Rural Disadvantage	1.2	0.9	0.9	0.3	110
M49	Tied / Tenant Farmers	0.8	0.7	0.7	0.0	73
M50	Upland and Small Farm	1.4	0.7	0.7	0.0	114
M51	Military Bases	0.4	0.1	0.1	0.3	29
M52	Non Private Housing	0.0	0.0	0.0	0.0	0
Total Mosaic coded		6541	3504	3207	297	10045
Households not coded		70	39	38	1	109
Total all Households in FES		6611	3543	3245	298	10154

Response Distributions when TDE is Introduced[1]

HÅKAN L. LINDSTRÖM

Abstract: *Three surveys at Statistics Sweden use Touch-tone Data Entry (TDE) for their data collection: Producers Price Index (PPI), Short-term Turnover Statistics for Domestic Trade and Services (STS), and Statistics on Municipalities' Costs for Social Assistance (MCSA). PPI is a monthly survey, MCSA is a quarterly and STS has both monthly and quarterly parts. MCSA is a census and the two other surveys use the same sample each month during the year. All three are mandatory surveys. The TDE implementation in PPI was preceded by a large experiment. Methodological studies were also made in STS and MCSA. Details are given to show the development of TDE and overall response rates over time and to what extent the surveys have turned into multi-mode surveys. The TDE response rates differ in these surveys. This report presents the differences and offers some explanations to them. It is a condensed version of a report written for Statistics Sweden. The complete report will include the details of MCSA and response distributions for all months of 1997.*

Keywords: *touch-tone data entry; nonresponse; multi-mode survey.*

1 The conclusive TDE experiment in the 1993 PPI

Since we knew that TDE already was in use at other statistical agencies, it was natural to investigate if it could be successfully adapted in some Statistics Sweden's surveys. Research reports by Clayton, R. (1989), Phipps, P.A. and Tupek, A. R. (1990) and Weeks, M. F. (1992) convinced us that the prospect of successful implementation was good.

About half a dozen surveys have the restricted number of numerical variables which is a necessary prerequisite for the use of TDE. PPI was the first choice for test and implementation as its sample size was large enough to make cost reductions possible. We did an experiment in that part of the sample for which TDE responding is most

[1] Acknowledgment: The author would like to thank Siv Grimsvik-Laskaridis, Gunilla Hedengren, Erik Henriksson and Karin Larsson (all at Statistics Sweden), who all have provided information and helpful comments on the report.

complicated, which is for imported goods. There were two complications which were not tested in any other application we knew of. Prices of imported goods might be given in different currencies during the year. For low priced goods it was necessary to register the price with a decimal point. If the respondent could manage these complications it would be sensible to recommend TDE also in less demanding applications.

The functioning of TDE was tested in an embedded experiment. 200 companies were sampled randomly from the "import prizes part" of the regular survey in October 1993. The experiment was extended into November and December to show if there was a persistent effect or a trend. The size of the net sample was 181, 180 and 179 respectively. The standard group was not offered to use TDE and continued to used the standard data collection method (shuttle paper form). It consisted of 474 companies. The standard and experiment groups are compared regarding response rates in Table 1.

Table 1: Response rates of net sample in experiment and standard groups

Sample	October	November	December
Treatment group	0.86	0.94	0.94
of which by TDE	0.70	0.74	0.74
Standard group	0.93	0.93	0.92

The total response rate in the experiment group is seven percentage points lower in October but one respectively two percentage points higher during the last two months. Too detailed information given to the respondents may explain the low response rate of the experiment group in October. It gave the impression of putting a heavier burden on the respondent than before. The increase in both general and TDE response rates of November and December may depend partly on the learning process, partly on more concise and efficient information about TDE in the information letter for these months. It was obviously important that the experiment continued for three months. The seven percentage points decrease of the overall response rate in September might otherwise have been a conclusive argument against implementation of TDE. The final TDE-use in the experiment is shown in Table 2.

Table 2: TDE users in December 1993

Sample group	Number	Per cent of :		
		gross sample	net sample	offered TDE
Gross sample	200	100	--	--
Net sample	179	90	100	--
Offered TDE	155	78	87	100
Used TDE	132	**66**	**74**	**85**

Each TDE respondent has to register the good's identity and price for at most ten types of goods or services together with company identity. When there is no new price he/she can use simple codes to register either if the price is the same as in the previous month or if they did not sell that kind of goods.

Table 3: Response time per item and type of goods - in seconds

Type of price information	October average	median	November average	median	December average	median
Same price	35.5	32.0	36.0	31.0	32.1	30.0
Did not sell	35.5	32.5	42.3	35.0	35.2	33.0
New prices	59.1	57.0	58.4	54.0	53.4	55.0

To register "new price" takes almost twice as long as the two other alternatives. The high average value for "did not sell" in November found no explanation. There is some tendency towards faster registration from September to December. The average time per response occasion was reduced from 122 to 102 seconds and the median time from 87 to 82. We guess that respondents eventually learn to spend less time to listen to instructions. Less than ten per cent of the respondents registered more than four prices at one occasion. During the three months only about a dozen contact persons used more than four minutes to register. The maximum time used was eight minutes and ten seconds. The total time to collect the information and prepare it for TDE registration is not known.

107 companies used TDE at all three occasions. 150 (84 %) of 178 companies, initially eligible for TDE, used TDE at least once. Only half a dozen respondents stopped using TDE after having tried it and became nonrespondents. We also made a debriefing

interview about the attitude to TDE after the September survey with one of ten of the experiment group. Few expressed a negative opinion to TDE.

We regarded the final TDE rate of the net sample to be very good. About 90 per cent of all companies had the technical facilities to use TDE at that time. The outcome of the test was judged as successful considering both cost and data quality aspects. Since the technical solutions worked as planned and the respondents used TDE although the demands on them were comparatively high, we could also recommend TDE to be used not only by PPI but also by other surveys, with similar types of respondents. The experiment is described in Lindström (1995).

2 Response rates in current PPI during 1997

Due to changes in organisation and responsibilities for the survey the implementation of TDE in PPI did not take place immediately after the experiment. To be able to encounter disturbances, TDE was introduced stepwise to parts of the sample. This introduction started only in February 1996.

Not only import prices as in the experiment but also companies reporting export prices and prices on the home market were included. Inclusion rules for companies were more restrictive. Companies which report goods with infrequent or very small price variations during the year are allowed to report their prices for two or more months at one occasion and they do it on paper forms.

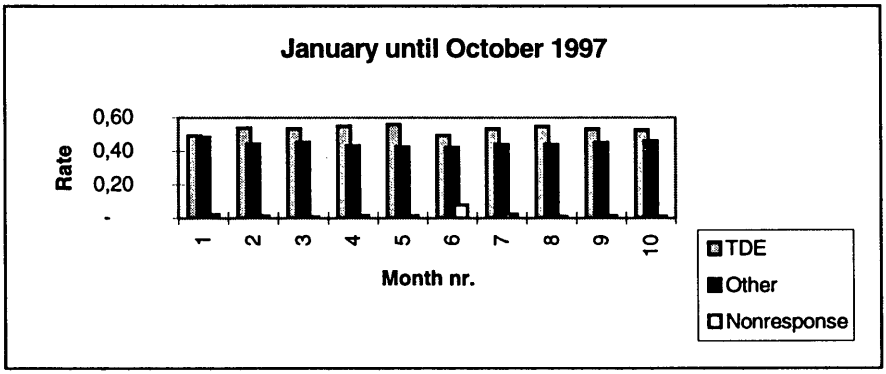
All sampled units of the selected types were offered to register (in February) the prices of January 1997 by TDE. In the beginning of 1997 the gross sample included 1 480 contact persons. 1 280 (86 %) were originally offered to use TDE and finally 1 060 of these were identified as positive to and equipped for registration by TDE. On an average 910 (61 % of the gross sample and 86 % of the TDE-group) use TDE each month. The rate within the TDE-group is almost the same as in the experiment in December 1993. The experiment's rate of TDE users (66 %) within the gross sample is an uncertain predictor as inclusion rules, coverage and nonresponse are different in 1997. However a s.r.s.-approximation of the confidence interval for the outcome of the experiment is 66 ± 7 % and includes the value of 1997 when selection effects are not taken in regard.

The lower overall TDE response rate during 1997 may depend on several factors. The experiment included import prices only, but the survey covers all prices. The TDE inclusion rules were more restrictive. In the regular surveys respondents reporting goods with rare and small price variations are not offered TDE. Nonresponse and overcoverage in the experiment was close to five and ten per cent respectively and much larger than in production. More detailed studies are needed to tell if technical factors are the explanation

for the difference or if willingness among respondents to use TDE has changed. The relevant observation for PPI is that the TDE response level turned out to be high enough to ensure savings.

It is the same sample during all months and the completed table for all the year that will tell how fast TDE is accepted by the respondents. Each new year a small jump in the time series may appear as there is a rotation of the sample of small companies (the big companies are permanently included) at the turn of the year. About five per cent of the respondents will be offered to use TDE for the first time when they report their January prices. Contact persons are only identified monthly among those who are selected to respond by TDE. An exact calculation of company response rates is laborious to do as there may be more than one contact person at a company and one contact person may split his/her reporting on two or more occasions. The monthly response distributions for all items of goods since January 1997 are shown in Figure 1 and Table 4.

Figure 1: PPI response distribution for items of goods



"Other" includes mail, fax and telephone replies, companies responding for several consecutive months at one occasion and those who have many items to report.

TDE is now the main data collection method, but there is no obvious trend yet. Up to October the monthly TDE rates for goods vary between 49 and 56 per cent and are mostly close to 55 per cent. As companies with many goods use paper forms the response distribution of goods is rather different from the distribution of contact persons. The nonresponse rate is very low except for in June. Table 4 shows the details better.

Table 4: Monthly response rates for goods by mode in PPI during 1997

Month	TDE	Other	Non-response	Net sample
January	0,49	0,48	0,02	4229
February	0,54	0,44	0,02	4226
March	0,54	0,45	0,01	4226
April	0,55	0,43	0,02	4229
May	0,56	0,42	0,02	4226
June	0,50	0,42	0,08	4223
July	0,53	0,44	0,03	4217
August	0,55	0,44	0,01	4204
September	0,53	0,45	0,01	4204
October	0,53	0,46	0,01	4220

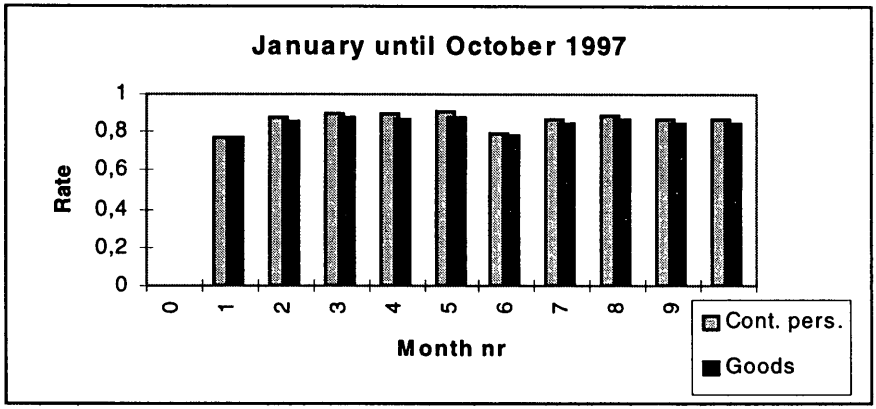
The five percentage points increase in TDE-rate from January to February may depend on the fact that some contact persons in the new panel were unfamiliar with TDE at the start, but learnt the new method quickly. The comparatively high nonresponse rate of June is likely to depend on that vacations in Sweden are most frequent in July.

Figures 2 and 3 show the response distributions in the group of around 1 060 companies that were offered to use TDE. The total number of contact persons who respond by TDE is on average 910 and the number of goods items they report is on average slightly below 2 250. Figure 2 shows that TDE rates are higher for contact persons than for goods every month.

Except for the low values of January and June the TDE response rates vary between 85 and just above 90 per cent for contact persons but one to three percentage points lower for goods.

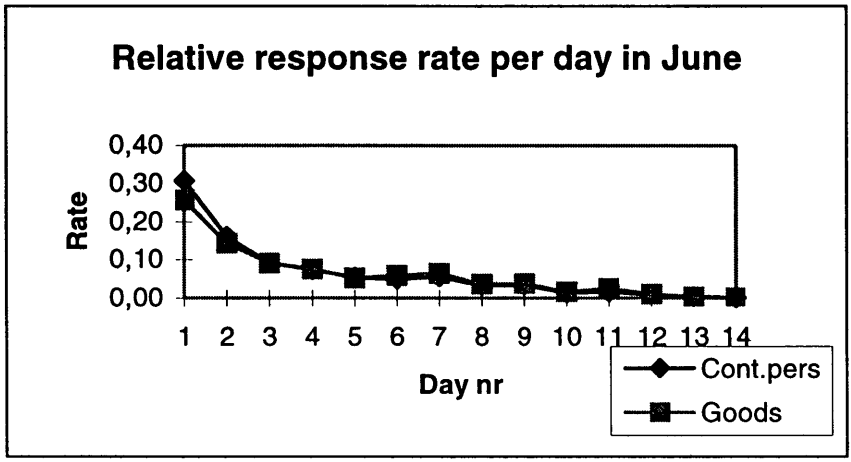
PPI has a rather short data collection period often ranging from 13 to 15 days. The arrival rates per day of the responses varies among the months. The presence of holidays during the data collection period is one important explanation for this.

Figure 2: PPI response rates among contact persons offered TDE



In Figure 3, June 1997 is chosen to represent those months where no holidays interfere. June 2 (day 1) is a Monday and the deadline for data collection is before the beginning of the Midsummer holidays.

Figure 3: Relative arrival rate by day in June 1997. Contact persons and goods items



The arrival rate of goods by day in the TDE group is rather close to the rate of TDE respondents. The response rates during the initial days are higher per contact person than per goods item as Figure 3 shows. There is a slight tendency for contact persons with few items to use TDE more frequently than contact persons in the TDE group with many items, especially obvious during the first few days. Around fifty per cent responded during the first three days. The data inflow follows a simple pattern and has the highest response rates during the first few days of the data collection period. After that the rates, on the whole, decrease day by day.

Some measurements of indicators on the respondent burden for registration and editing are computerised like the time used to register, the number of corrections, the number of spoken comments. The indicators are there to draw attention to such major variations between the months, which may be due to unexpected disturbances. Some variations will be a consequence of variations on the rate of price changes from one month to another. These routines are still under development. Other useful indicators would be the response distribution on "New price", "Same price", "No item sold this month" and the number of corrections distributed on price, goods identity and company identity.

3 Response rates in current STS

In the Short-term Turnover Statistics for Domestic Trade and Services (STS) a company only reports one figure each month/quarter - its turnover during the period of reference. Companies in retail trade report after each month and other companies after each quarter. As the response burden is the same at each occasion for all companies, there was no reason to except some companies from TDE as PPI did. All companies were offered to use TDE, both in the experiment and later when TDE was introduced in production.

The companies in the STS are on average smaller than the companies responding to the PPI. It was feared that the respondents of this survey were less technique-minded and less prone to accept to use TDE so they were approached very softly, when they for the first time were offered to register that way.

During 1996 the TDE response rate of the gross sample was very close to ten per cent in both monthly and quarterly surveys. Even if there is a small cost reduction already at this response level a substantial further reduction was desirable and believed to be within reach as the TDE response rate was substantially below those of MCSA and PPI. Since the respondent burden was lower than in PPI (only one answer, all prizes are given in Swedish currency and no decimal point), there were two main hypothesis to explain for the differences between STS and PPI.

- The populations are different. STS has a larger proportion of small firms in its' sample and fewer of them have the technique/ capacity/ experience/ organisation to respond by TDE.
- The introduction of TDE was too cautious. The respondents did not perceive that it was important for Statistics Sweden that they turned to TDE.

To find answers to the two assumptions an experiment was planned. A sample of 300 enterprises was drawn among the regular sample and equally distributed on three strata by company size. A new introductory letter was written for this experiment group - and these companies no longer received a response envelope. Both the layout and the phrasing of the "new" message indicated very clearly that Statistics Sweden's demand for answers by TDE was strong.

Table 5 presents the outcome of the experiment distributed on strata and the data collection mode chosen by each company.

Table 5: TDE - sample by response mode in three strata by number of employed in the third quarter of 1996

Response mode Number and per cent in an embedded study						Non-response	Net sample size
Number of employed	**TDE**	**Mail**	**Fax**	**Phone**	**All**		
Stratum I 0-1	33	33	6	4	76	21	97
per cent	*34*	*34*	*6*	*4*	*78*	*22*	*100*
Stratum II 2-10	38	43	5	5	91	11	102
per cent	*37*	*42*	*5*	*5*	*89*	*11*	*100*
Stratum III > 10	41	45	5	6	97	4	101
per cent	*41*	*45*	*5*	*6*	*96*	*4*	*100*
All strata	112	121	16	15	264	36	300
per cent	*37*	*40*	*5*	*5*	*88*	*12*	*100*

The effect of the information being more demanding and the absence of response envelopes was strong. The TDE response rate increased from about ten per cent to 37 per cent (unweighted average) of the net sample and 42 per cent among the respondents. The

increase was large in all strata. There are both size and nonresponse effects. A high nonresponse rate in the strata of small companies contributes to a low TDE rate in these. The larger the companies, the larger is also the TDE rate of the net sample and the lower the nonresponse rate. But even in stratum III the level is about 15 percentage points below the TDE response rate of PPI. Probably there is a higher proportion of small companies in STS even in this stratum. Mail was still the dominant data collection mode but TDE came close.

Due to the positive outcome of the experiment it was decided to offer TDE to all companies in the survey. The same information and encouragement as in the experiment was used to prompt them. This was done for the first time when the turnover for January 1997 was to be collected. Table 6 shows the monthly response distributions during 1997. The information to the firms has been uniform through the period as no additional steps have been taken to increase the TDE response rate further. Accordingly the time series show the spontaneous development of TDE use among the contact persons.

Table 6: Response distribution. Contact persons in per cent during 1997. Complete monthly and quarterly surveys

Month	TDE	Mail	Fax	Phone	All response	Nonresponse	Sample size
January	48,5	12,9	10,4	1,2	73	27	2764
February	45,0	10,6	8,4	15,0	79	21	2738
March	43,2	17,7	6,7	10,4	78	22	7723
April	46,5	10,2	9,0	13,3	79	21	2684
May	46,7	10,3	9,9	12,2	79	21	2677
June	40,9	15,9	7,7	11,5	76	24	7372
July	43,8	9,8	12,9	12,5	79	21	2664
August	44,7	8,8	15,4	12,1	81	19	2664
September	43,1	13,7	9,4	15,8	82	18	7202

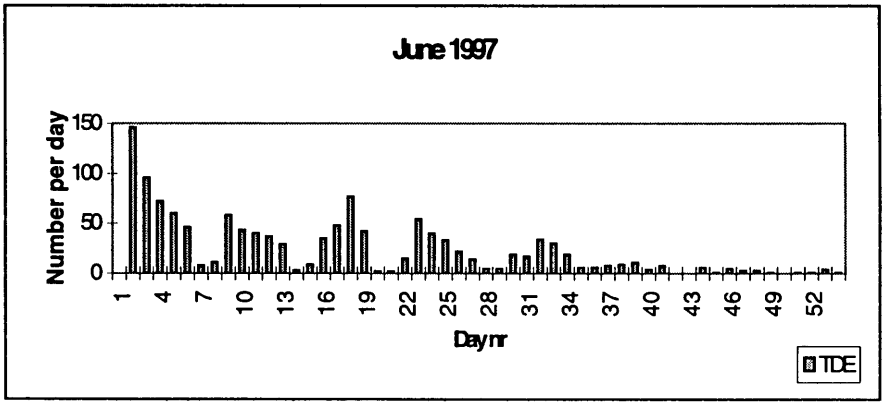
TDE turned out to be by far the dominant data collection mode during 1997. This is in contrast to the experiment where mail was slightly more frequent. In the net sample of 1997 the TDE response rate is higher than in the experiment during all months. No trend is noticed. In average about ten per cent of the respondents chose each of the other modes. The mail response rate is lower but the fax and phone rates higher than in the experiment. A possible explanation for this difference is that in the experiment in 1996 the sample was "old" and had responded by mail since January and was used to put their filled-in forms

in an envelope and mail it. The new sample in 1997 was never given any response envelopes. As a consequence it was more prone to chose TDE. On average the TDE rate of the net sample is about 15 percentage points lower than in the PPI.

The high mail response rates in March, June and September may reflect that the quarterly surveys cover a broader population which is surveyed with larger time intervals. The unweighted nonresponse average in the survey has varied a lot over the years. In 1995 it was 23 per cent and in 1996 only 18 per cent. This material is not analysed enough to verify whether the increase from 1996 to 1997 depends on the introduction of TDE or has other explanations.

The answers are registered by the day they arrived except for answers by phone. Most of them arrive late in the data collection period which is more extended than in PPI. The pattern of inflow is different. June is chosen as an example. Figure 4 shows the arrival TDE rate by day and Figure 5 the cumulative rates for all modes (except telephone) in June 1997. Answers by phone are often a result of a contact made by the STS staff to remind a late respondent.

Figure 4: Number of TDE responses by day during June



The pattern of the number of respondents per day has a number of local peaks and is different from that of PPI.

Figure 5: Cumulative response rate by mode in June 1997

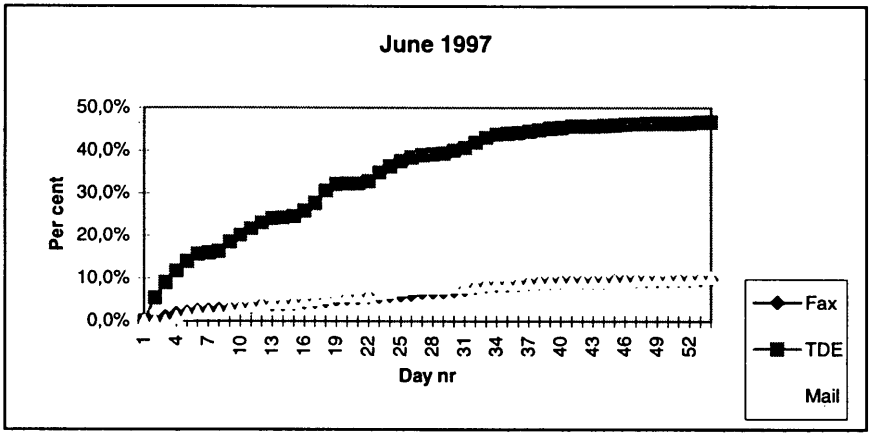


Figure 5 shows that those who used TDE tended to respond earlier than respondents by other modes. For example during the first few days of June 1997 the proportion of TDE respondents among all respondents (except telephone) was almost 80 per cent. Thereafter the rate fell slowly and stopped at 70 per cent.

4 Concluding remarks

4.1 TDE in production - a summary

The three examined surveys are all periodical and mandatory. The respondents are companies and municipalities. The recent introduction of TDE has turned the surveys into multi-mode (more than before) surveys with TDE as the main data collection mode. TDE response rates have varied between 40 and 60 per cent of the net sample of contact persons. Differences between surveys depend both on the sampled population and the way to argue for TDE. When the three surveys are compared, PPI has the highest TDE response rate. PPI told the respondents most explicitly that it demanded TDE answers. MCSA started in 1994. Its TDE rate is increasing but it is still 10 percentage points below PPI. STS increased its TDE response rate substantially when it used similar arguments as PPI. Still it is about 15 percentage points below. The lower TDE rates in MCSA may also be due to the importance for respondents to check that sums are correctly recorded or otherwise edit them. So the response burden measured in time at the phone is high even if there only are three sums to register. The time series are still too short to allow any major

conclusions about the trend of the rates, except in one case. In the MCSA there is an increase by two percentage points per year. TDE together with other data collections modes has kept the overall response rates at least at the same level as before in PPI and MCSA. In the STS the case is more dubious as the response rate was higher the year before the introduction but lower than during the year before that.

4.2 Experiment and production

There were reasons to expect good predictive abilities of the experiments, which this report refers to. The experimental conditions were under good control and no problems disregarded. The experiments were embedded and had the best prerequisites to give correct results. Still there turned out to be differences. When it came to production there were sensible reasons not to stick to all the experimental conditions any more. To make inferences from an experiment to the outcome of a continuing survey is obviously a problem that is easily underrated. Production defines its own goals. When these are attained there is seldom time and resources left to verify in retrospect to what degree the results of the experiment were valid.

In the 1993 PPI experiment it was vital that the experiment continued for three months. The conclusion and the decision had been different if we had not adapted the experiment to the experiences of the first month and let it continue two additional months. The TDE rate of the gross sample turned out to be about 5 percentage points lower than in the experimental group. This may be explained by random variation, but different inclusion rules, overcoverage and nonresponse may have caused systematic differences that will not be fully understood without further analysis.

In the STS on the other side the TDE rates turned out to be 5-10 percentage points higher than in the experiment. In this case we think that we know the explanation. The experiment was performed in the second half of the year and the respondents were used to respond by mail. The following year the new panel of the sample did not have this experience and was more prone to use TDE.

As expected, it was not possible to make precise predictions of TDE rates in one survey from experiments or experiences from another. The populations are different, the approach to the respondents is different and so is the response burden and the variation of response burden among respondents.

The observed differences underline the importance of a close connection between experiment and implementation in production together with a subsequent follow-up. This is not without problems. The production staff in each survey is pressed for time and resources and has seldom the time needed to find out in detail how different factors

operate to achieve improvement/changes in their survey as long as the outcome is good enough. They have still less resources for cross-survey comparisons. To make comparisons across surveys - necessary to understand how a method works under different conditions, additional resources must be available for independent evaluators.

4.3 Respondents attitude to TDE

Debriefing of respondents has shown that a large fraction of them accept or even prefer to use TDE rather than shuttle forms. The reason may be that they feel that they are working in a modern way and avoid additional work. If they do the editing well there will be no need for re-contacts to explain and correct. To correct errors immediately is less demanding than to make it one or two weeks later and repeat the whole procedure. The reporting person often has to collect information from others and even make some calculations. To register the information by TDE or in a mail form is only a minor part of the response burden.

4.4 Improved control of the surveys

Except for the savings there are a number of advantages in the TDE part of a sample. There are possibilities to monitor it better and even to increase the response rate. When there is a strong demand for the results to be punctual and the data collection period is rather short, an efficient field work is extremely important. Given that the respondents register the information the same day, independently of the mode they prefer, TDE will reach the statistical agency at least one day faster than mailed responses. As the producer can learn almost instantly if the sampled object has responded by TDE or not there will be no unnecessary reminders in this group as can happen when a filled-in questionnaire is still in the mail.

A contact person is registered for each company/municipality. In many cases, although not in all, the same person reports the answers during several reporting occasions. The production staff will repeatedly get into contact with many of the responding persons and will learn about their attitudes and reporting styles - for example if one has to wait for their answers and if some kind of personal reminder will be necessary. This is, together with an inflow chart of responses per day, very helpful to develop an efficient plan for remainders. There is no need for further registration and less need of editing or contacting respondents again as they mostly already have edited their responses. The possibilities to use production statistics to identify disturbances and to monitor the data collection are still only partially exploited.

References

Clayton, R. (1989). Developing a Cost Model for Alternative Data Collection Methods, Mail CATI and TDE. JASA proceedings

Lindström, H. L. (1995). The Touch-tone Data Entry Experiment (TDE) in the Swedish Producers Price Index (PPI) in 1993. Proceedings of the International Conference on Survey Measurement and Process Quality, Bristol, April 1-4. Contributed Papers Volume, pp. 163-167

Phipps, P.A. and Tupek, A. R. (1990). Assessing Measurement Errors in a Touch-tone Recognition Survey. Survey Methodology, 17, pp. 15-26

Weeks, M. F. (1992). Computer-Assisted Survey Information. A Review of CASIC Methods and Their Implications for Survey Operations. Journal of Official Statistics, 8, pp. 445- 466

A Survey on Telephone Coverage in Finland

VESA KUUSELA

Abstract: *Telephone coverage and telephone ownership in Finland were explored by a specific survey in January 1996. The results show that telephone coverage, in terms of conventional phones, has decreased steadily from 94% in early 90s to 86%. Simultaneously mobile phone penetration has increased from 7% to 42% and therefore the combined telephone coverage has increased. Nearly all households with a stable position in society have a conventional phone and also households who live outside cities. Households with unstable position in society or people who have fallen outside the processes and institutions of the society have much fewer conventional phones. One member households had fewer phones than other households. Especially young men living alone had more frequently mobile phones than conventional phones. Because of the rapid increase of mobile phones, some areas of survey practise have to be reconsidered.*

Keywords: *telephone coverage, mobile phone coverage, undercoverage*

1 Introduction

Telephone coverage has an effect on the nonresponse in surveys in two ways: in telephone surveys it directly sets the limits to which part of the population may be interviewed.[1] In face-to-face interviews, the initial contact and appointment are often done by a telephone. Telephone coverage is usually defined as the proportion of households that have a conventional telephone.

The actual telephone coverage in Finland was questioned in the early 90s when the search of telephone numbers of the sampled persons was commissioned to telephone operators (see Kuusela 1996). The search yielded a telephone number to only 85% of a random sample from the Population Registry, although the official telephone coverage was 94%.. The telephone coverage in Finland, like in many other countries, has been traditionally

---

1 Conceptually we should rather speak of undercoverage. However, telephone coverage may be regarded as nonresponse, as well, because both effect the sample realised in the same way and both may be corrected in the same way, as well.

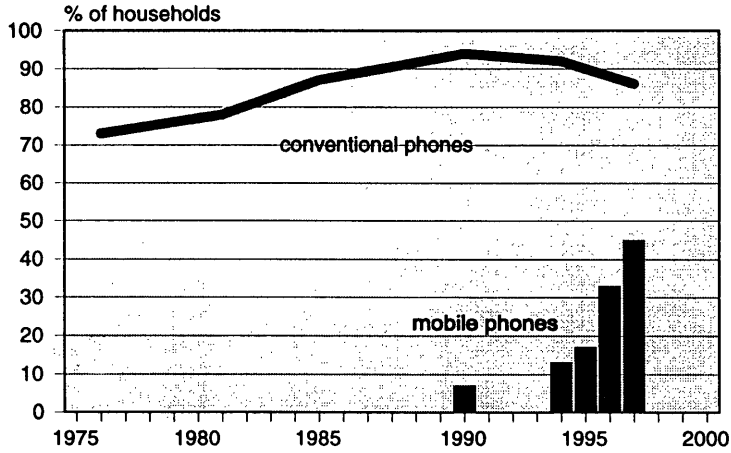
estimated on the basis of general purpose household surveys. Nonresponse in these surveys has been relatively high (about 30% in Finland) and the questions concerning telephones were kept fairly simple and they were asked in the middle of questions about other equipment in the household. This brought up the question whether the actual telephone coverage was something else.

In January 1996 a survey focused on telephone ownership was conducted to obtain a more accurate estimate of telephone coverage in Finland. Another aim was to get information about the nontelephone households. Sample size was 4,838 and the nonresponse rate was 10.3%. Interviews were carried out on the telephone when possible and face-to-face interviews otherwise.

2 Telephone coverage in Finland

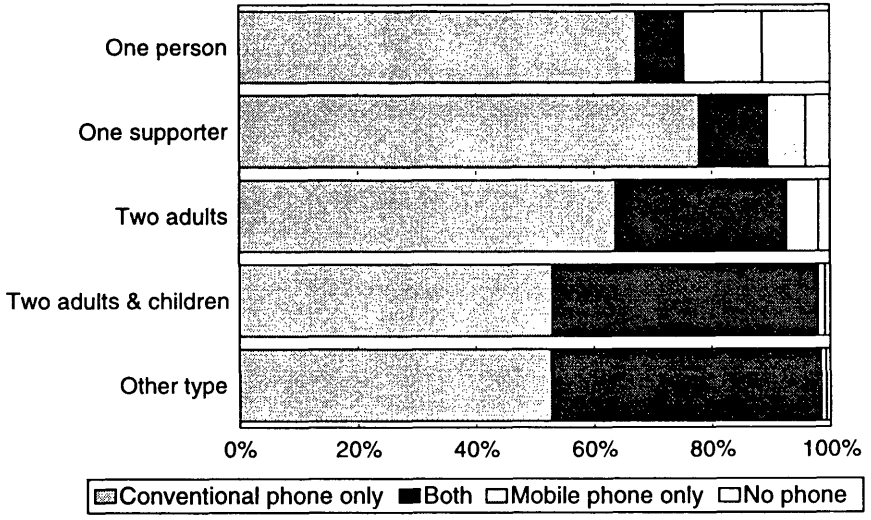
Telephone coverage in terms of conventional phones (that is phones connected with a line to a plug) fell from 94% in 1990 to 88% in 1996 (see Figure 1). Simultaneously, the coverage of mobile phones saw a rapid increase. In January 1996, the coverage of mobile phones was 33% of the households, but during the next three months the number of mobile phones increased by nearly 25%. In another survey (see Nurmela 1997) carried

Figure 1: The change in telephone coverage in Finland from 1975 to 1997. The coverage of conventional phones is presented by the line and the coverage of mobile phones by the columns



out at the end of the same year, 86% of households had a conventional phone and 42% had at least one mobile phone. (The number of mobile phones has grown even more rapidly since then). In January 1996, the coverage of these two amounted to 95% and ten months later the combined coverage was 98%. In other words, telephone coverage increased but the increase was due to the increase in mobile phones. So, their relative amount has increased very rapidly. Already, the number of mobile phones in Finland is so great that they cannot be ignored in survey practice. In particular, if RDD is used mobile phones should be included in the algorithm.

Figure 2: Combined telephone coverage (%) in Finland (1996) by the type of the household

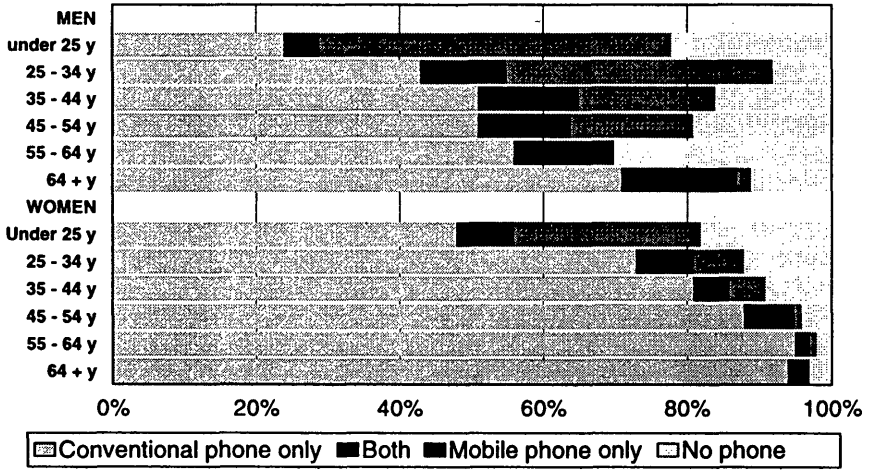


Telephone coverage (conventional phones) varied considerably both geographically and from the socio-economic point of view. The size and the structure of household had an exceptionally strong effect on telephone ownership (see Figure 2): only 75% of one-person households had a conventional phone. In all other types of households, telephone coverage was clearly more than 90%. If there were more than two children in the household, there was practically always a conventional telephone in the household, as well. Other types of households with high coverage of conventional phones were those of farmers, pensioners, and of households who had lived in the same dwelling at least three

years, especially if the dwelling was in a single family house. In all these groups the coverage was almost 100%. If the head of the households was an entrepreneur or a white collar worker, the telephone coverage was also high.

Telephone coverage was low in one member households, in particular if the member was a young man, who was a student or unemployed. Telephone coverage among females was distinctly higher in all population segments. For instance, only 45% of unemployed men had a telephone but 73% of unemployed women had one. The low telephone coverage of the one member households partly explains the decrease in the total telephone coverage, because in Finland the proportion of the one-member households has increased for many years now.

Figure 3: Combined telephone coverage of conventional and mobile phones in one-person households in Finland



One reason why people live without telephones is, of course, lack of money, but an even more important reason is their life style. An indication of this is that mobile phones were most common in those households which less often had conventional phones. That is, the mobile phones had replaced conventional phones. For instance, only 29% of young men living alone had a conventional phone. However, 54% of them had a mobile phone and 49% had only a mobile phone. Another subgroup with low telephone coverage was made up of people who might be described as outsiders with a weak attachment to society and

its processes and institutions (for example single or divorced middle-aged, possibly unemployed, men). Smith (1990) made the same observation in United States. Typical to the latter group is that they do not have many mobile phones either.

Telephone coverage was higher in country areas than in towns. In many towns, less than 85% of the households had a conventional phone. (In Helsinki, telephone coverage was 90%.) The further the household was from a population centre the more probably there was a telephone.

87% of the conventional phones were listed in the printed telephone directory under the name of someone living in the same household and 5% of the numbers were not listed but they were not secret, either. Slightly more than 4% of the households had a secret telephone number and the rest of the telephones were listed by a person who did not live in that household or by the name of the employer. Secret numbers were more frequent in big cities, especially in the capital area. In the countryside, secret numbers were quite rare.

3 Discussion

In terms of conventional phones there are some clearly detectable population segments regarding telephone coverage: Practically all households with a stable position in society (ie. head of the household has a (good) job; families with children; household had stayed for some time in the same dwelling, etc.) had a conventional telephone. On the other hand, households with unstable positions in society have far fewer conventional phones. The unstable position may stem from, for instance, moving from parents' home, studying, unemployment, or change of dwelling. The most problematic subgroup, in many ways, is composed of those people who have fallen outside the processes and institutions of the society.

In Finland, the situation in telephone coverage is changing very rapidly due to the rapid increase in the number of mobile phones. Consequently, the proportion of households that have only a mobile phone is increasing. Characteristically, the mobile phones are more common in that type of households where the conventional phones are less frequent. For instance, many smaller households have only mobile phones. Two main processes are causing this: some of these people have given up the conventional phone and bought a mobile phone instead because it fits better in their life style; some of them bought a mobile phone as their first phone, and probably they have not acquired a conventional phone.

However, it is not clear what the future of telephone ownership will look like. The transition is partly connected to the change of household structure in the Finish society.

The proportion of one-member households has been increasing for a long time (at the moment, more than 35%) and especially young people living alone are most frequently acquiring mobile phones instead of conventional phones. This development would increase the amount of those households that have only a mobile phone. On the other hand, along with the growth of the Internet, the number of conventional phones may increase, because a mobile phone does not serve internet purposes well and a part of the same group of people, who now have only the mobile phone, are heavy users of the Internet. Another thing difficult to predict is what will happen when young people get married (or move together) and have children. Will they have only mobile phone(s) or will they have a conventional phone, as well?

From survey organisations' point of view, the transition from the conventional phones to mobile phones has two sides: The combined telephone coverage has increased and now those population segments which earlier were hard to reach have always a phone in their pocket; on the other hand a mobile phone is not as good for interviewing as the conventional phone: long interviews are not possible and respondents may be virtually anywhere, possibly in a place where he or she cannot be interviewed. The increasing number of mobile phones changes the costs on interviewing because it is much more expensive to call to a mobile phone, and, moreover, occasionally a part of the expenses goes to the respondent.

Another point of view, at least in Finland, is the change in sampling procedures. At the moment, mobile phones are not listed in a catalogue as frequently as conventional phones, thus introducing a new source of bias in telephone surveys (see also Brick et al. 1995, Keeter 1995). In Random Digit Dialling-procedures mobile phones cannot be ignored. This, in turn, leads to a new problem: there are households with more than one mobile phone, hence all households do not have equal selection probability.

The increase of mobile phones forces us to think over what telephone coverage means. In Finland, quite a few households have already more than one mobile phone and the number of those households is increasing. Even now, in some larger households, every member has one of his/her own. In the future, the mobile phone might become more like a personal phone (like a wrist watch) and the conventional phone will then assume the role of a household appliance (like a clock on the wall). Consequently telephone coverage will be more than 100% if the number of private phones is calculated in respect to number of households. Nevertheless, there will be households without any telephone and many households without a conventional phone. The old definition of telephone coverage will not give a precise description anymore and it cannot be used as a measure of the accessibility.

References

Brick, J.M., Waksberg, J., Kulp, D. and Starer, A. (1995). Bias in List-Assisted Telephone Samples. Public Opinion Quarterly, 59, pp. 218-235

Keeter, S. (1995). Estimating Telephone Noncoverage Bias with a Telephone Survey. Public Opinion Quarterly, 59, pp. 196-217

Kuusela, V. (1996). Telephone Coverage and Accessibility by Telephone in Finland. In: Laaksonen, S. (ed.). International Perspectives on Nonresponse. Proceedings of the Sixth International Workshop on Household Survey Nonresponse. Statistics Finland, 1996.

Nurmela, J. (1997). The Finns and Modern Information Technology. Statistics Finland, Reviews 1997/12

Smith, T.W. (1990). Phone Home? An Analysis of Household Telephone Ownership. International Journal of Public Opinion Research, 2, pp. 369-390

Is it True that Nonresponse Rates in a Panel Survey Increase when Supplement Surveys are Annexed?

MALKA KANTOROWITZ[1]

Abstract: *It is usually assumed that there will be an increase in nonresponse, especially for refusals, when a greater burden is imposed on the respondents. The paper deals primarily with the issue relating to how nonresponse rates of a current panel survey are affected by the additional burden arising from appending supplement surveys. A long-term analysis of the nonresponse rates in the Israeli Labour Force Survey (LFS), by main types of nonresponse, is used for this study. At the same time, some other effects are studied, such as that of the mode of data collection and of the season of the year. The assumption of a decrease in the response rates with the number of rounds is examined to a limited extent.*

Keywords: *annexed survey, interviewing mode, interviewer workload, nonresponse, panel survey, respondent burden, supplement survey*

1 Introduction

In general, the assumption is that when a greater burden is imposed on the respondents, there will be an increase in the nonresponse rates, especially for refusals. The longer the questionnaire and the more the number of interviews in a panel or a longitudinal survey, the greater is the burden. In order to reduce potential nonresponse bias, there is a tendency to lessen the burden, despite the advantages of conducting a survey with a longer interview, or with more repeated interviews.

However, there is no conclusive evidence in the literature to support this assumption. Bogen (1996) gives a comprehensive review of the existing literature, dealing mostly with mail surveys, though covering a few studies relating to other types of surveys. She quotes Berdie (1973), who earlier made a review of the research done on this issue, and concluded that "surprisingly few studies actually have examined correlations between length of questionnaires and rate of response, and those studies that have done so generally have yielded confusing results". Bogen also finds that, the results are still not

[1] Many thanks to Haim Fraid for his assistance in the data processing.

conclusive: "they are confusing and contradictory... and designers still aim for shorter questionnaires with little more justification than the logical assumption that longer interviews will result in higher nonresponse".

A specific case of a greater burden is when a current rotating panel survey, like many Labour Force Surveys (LFS's) over the world, is used as a convenient tool for conducting supplement surveys (though only with a reasonable extra burden). These may be annexed to one or more rounds, of all or some of the panels, either occasionally or on a regular basis.

The primary advantage of conducting a supplement survey appended to a current survey is that of cost. It will be less than for a separate survey, since no expense of separate sampling and of organizational infrastructure will be incurred. Also, the additional interview will be shorter, because some of the data collected in the current survey will anyway be required for the supplement survey. Beside cost considerations, further gains can be achieved in improving data processing, imputation and weighting in the supplement survey.

Against these advantages, there is a concern of unfavourable effects on the response rate, especially for refusals, in both surveys. This may be even more serious when the supplement survey is prone to a relatively high refusal rate. Because usually preference is given to the current survey, the annexed questions are left until the end of the interview. When the survey is annexed to one of the intermediate rounds of the current survey, there is concern that this may harm the following interviews. Therefore, there is a tendency to hold heavy supplement surveys in the last round. Furthermore, because there is a fear of harming the current survey, sometimes a separate survey is conducted, although it would have been better to annex it to the current survey.

With supplement surveys, there may be either direct or indirect impact on the interviewer's workload (Marquis (1979)). Usually, the interviewer's quota is about the same for all rounds, and they have to cope with the additional time required for the extra interview. So, less time can be allocated to deal with hard-to-get respondents, thus may affect negatively the response rate of the current survey. Also, it could be that because interviewers expect a negative reaction for a long interview, they do not try as hard as they could.

Supplement surveys may have different degrees of burden on the respondent and on the interviewer workload. These depend on the length of the additional interview, on the type of the questions and on the mode of data collection, e.g. a personal interview of all or some of the household members, or the use of a proxy.

Surveys that can also be considered as supplement surveys, though of a different type, are subsequent surveys. These use rotating-out panels, so that in practice, the households have an additional interview (all of them, or only a certain sub-population, for which the screening is done by the current survey). Harm to the current survey is avoided, but, there is the fear that the subsequent survey may suffer from more refusals due to a further round of interviews. Thus, such surveys are often avoided, despite the advantages. For the same reason, the number of rounds in a current survey is limited to avoid negative effects on the response rate, although in some cases more rounds would much better fulfill the targets of the survey.

In this paper, the effect on the nonresponse of the current survey from the additional burden that is imposed by conducting supplement surveys, is examined. At the same time, some other effects are studied, such as that of the mode of data collection and of the season of the year. The assumption of a fall in the response rates with the number of rounds is examined to a limited extent. A long-term analysis of nonresponse rates in the Israeli LFS is used for this study. This survey, briefly described in section 2, is suitable since it is a panel survey and is used as a vehicle for other surveys relating to various subjects with different degrees of burden. The analysis is done for Total Nonresponse, as well as for four types of nonresponse: Refusals, Absentees, Interviewing Difficulties and Other Nonresponse. The data and the model that were used for the analysis are described in section 3. The results obtained are presented in section 4 and a summary is given in section 5.

2 The Israeli LFS

Current LFS's (and similar surveys) in many countries are conducted as rotating panel surveys. In general, the rotation system is chosen to obtain estimates of changes between successive periods (e.g. months, quarters or years) with sampling variances as low as possible, together with periodic cross-sectional estimates. This is done by considering the burden on the respondents, thus, with a limited number of repeated rounds to avoid increasing refusal rates. Different rotation systems with different numbers of rounds are used. For example, the CPS in the USA (U.S. Census 1973) is a monthly survey with eight rounds (four successive months and the same four months in the following year); the Canadian LFS (Statistics Canada 1997) is also a monthly survey with six successive rounds; and, the British survey (Steel 1997) is quarterly with five successive rounds.

In Israel, the LFS is carried out currently as a quarterly rotating panel survey with four rounds, i.e. four quarterly surveys are conducted every year, each comprising four panels. So, each quarter includes one panel investigated for the first time and one each for the second, third and fourth time. Each dwelling unit (with a few exceptions in the small localities) participates in the sample for two consecutive quarters and, after a break of two quarters, for two additional consecutive quarters (Israel CBS 1995).

The mode of data collection in LFS's is not necessarily the same in all rounds of the same panel. Usually, the first interview is face-to-face and some or all of the following interviews are by telephone (where possible). For example, in the Canadian and the British LFS's all interviews, except the first, are conducted by telephone. In the USA CPS, besides the first interview, also the fifth interview (conducted eight months after the fourth) is face-to-face, but all other interviews are by telephone. In the Israeli LFS, in general, the first interview is face-to-face, the second and third are by telephone, but for the fourth round, the interviewer visits the household for the Income Survey, regularly appended to this round.

The sample for the LFS is selected once a year for all four panels with equal final probabilities. Each panel is a sub-sample of about 2,700 dwelling units, distributed evenly over the 13 weeks of the quarter. Large localities are included with certainty and small localities, after stratification, are sampled with PPS. Within each locality, the dwellings are sampled from the municipal tax file. A sample of new dwellings is currently assigned to all panels, thus starting their investigation only in the advanced rounds of their panel.

Some dwellings may be out-of-scope (vacant, demolished, used as a business, etc.) in some or all rounds. The LFS is based on repeated investigation of the same dwelling units rather than of the same households. Thus, it is not necessarily true that the same household participates in all the eligible rounds and households that do not live in the same dwelling for the whole duration of the panel have less rounds. If there is more than one household in a dwelling, each will have at least one "blank" round. Further, new dwellings are introduced into the sample only at an advanced round, and so all rounds prior to the round in which they were inhabited are considered as "blanks". Because of two exceptional events in Israel (the Gulf War in the first quarter of 1991 that lasted for 6 weeks and, a 3-week strike of all government employees in the third quarter of 1993), the samples of two quarters had to be reduced, thus more "blanks" were introduced into the sample. On the average, in the years 1988-96, of all households, 63% were in the sample in 4 rounds, 9% were with 3, 15% with 2 and, 13% with 1.

Whenever an interview has not been conducted, a Non-Interview Form has to be filled in by the interviewer who has to state the reason (if necessary, neighbours are asked or enquiries are made locally, like at the nearest post office or grocery). This is done first, to distinguish between out-of-scope and nonresponse cases. Then, for each nonresponse, further enquiries are made to find out a detailed reason (e.g. Refusals, by who refused on which visit; Absentees, by length and reason; Interviewing Difficulties, according to language barrier, illness, incapable of cooperating; and, Other Nonresponse, due to failed telephone interview, non-location, no attempt to interview, etc.). Thus, for each household for each of its four rounds there may be one of these possibilities: (1) out-of-scope; (2) "blank"; (3) nonresponse, by detailed reason; and, (4) response.

The overall Total Nonresponse rate in 1988-96 was 10.5%. Out of this, 33% were Refusals, 43% Absentees, 12% Interviewing Difficulties and, 12% Other Nonresponse. The nonresponse rates were not the same for households with different number of eligible rounds. For the 4-round households, the Total Nonresponse rate was lower (8.1%), and out of this, the proportion was higher for Refusals (42%), lower for Absentees (36%) and Other Nonresponse (9%), and it was almost the same for Interviewing Difficulties (13%).

The nonresponse rates were higher for the 12 panels of the earlier part of the period (group {a}) than for the 14 panels of the later period (group {b}). For the 4-round households, the average rate was 9% for group {a} and 7.6% for group {b}. This could be due to difficulties in conducting the survey at a time of exceptionally large waves of immigrants arriving in the country at the early part of the period. However, the decrease in the nonresponse rates may be also due to improvements in the field work.

As in similar surveys in other countries, so too in Israel it is the practice from time to time to annex different supplement surveys to the LFS. These surveys are conducted periodically or only once, except for the Income Survey that is regularly annexed to the last round of each panel. Whenever a supplement survey is conducted, notification about it is given in the advance letters that are sent to the households for the LFS. Nevertheless, the supplement questions are always asked after the interview for the LFS, to avoid harming the LFS. Some of these supplement surveys are easy, such as the Membership in Youth Movements Survey, Kindergarten Attendance Survey, and the Multiple Jobs Survey. Others are heavier, for example, the Income Survey, the Household Equipment and Living Conditions Survey, the Use of Health Services Survey and the Victimization Survey (in the latter two, every adult member of the household has to be interviewed).

3 The data and the model used for the analysis

The data used for the analysis are from all the rounds of 26 panels that participated in the Israeli LFS in the period between 1988 and 1996. Most localities with at least 2,000 inhabitants (comprising about 80% of the sample) took part in this study. Other localities were excluded, primarily because they were rarely covered in the supplement surveys.

Up to the present, only the 4-round households, i.e. residing in the same dwelling during the whole cycle of a panel, participated in the analysis. The inclusion of households with less rounds would cause complications in the analysis. For example, it would be necessary to determine for each household if it was in the current survey both in rounds with a supplement survey and in subsequent quarters. Nevertheless, it is intended to extend the study to include households with less eligible rounds.

Altogether, 42,300 households participated, each household with its four eligible rounds - a total of nearly 170,000 observations, though not all independent.

The analysis of the effects on the nonresponse rates was done for Total Nonresponse and separately for Refusals, Absentees, Interviewing Difficulties and for Other Nonresponse, by means of a regression model:

$$p_i = \alpha + \Sigma_k \beta(k) X(k)_i + \varepsilon_i$$

The explained (dependent) variable p_i (i= 1,...,104) is the nonresponse rate for each of the various types of nonresponse, for 26 panels each with 4 rounds dependent on each other. $X(k)_i$ are the explanatory (independent) variables, $\beta(k)$ are the corresponding regression coefficients, α the intercept and, ε_i is the random deviation, namely the residual not explained by the regression model (where the expected value of ε_i is zero).

The explanatory variables are all dummy variables, i.e. with a value 1 if panel-round **i** has the characteristic, and 0 otherwise:

- Panel group: **X(1)** for the 14 panels in the later part of the period - group {b}.
- Quarter of the year: **X(2)** for quarter II (April-June), **X(3)** for quarter III (July-September), and **X(4)** for quarter IV (October-December).
- Round: **X(5)** for the second round, **X(6)** for the third round, and **X(7)** for the fourth round. Since in this survey, interviews in both the second and the third rounds are by telephone and, in both the first and the fourth rounds they are face-to-face, the effect of the number of rounds can be distinguished only between rounds of the same interviewing mode.

- Supplement survey is annexed: **X(8)** when easy, and **X(9)** when heavy. For this study, each supplement survey was determined to be easy or heavy according to several judgemental criteria, e.g. the length of the annexed questionnaire, the sensitivity of the questions, the possibility of getting the required data by proxy against personal interview.
- Quarter following supplement survey, for a given panel: **X(10)** when following an easy survey, and **X(11)** when following a heavy survey. This variable was included to find out if there is any effect on the nonresponse rate, resulting from the reaction of the respondents to the burden imposed on them in the previous interview.
- The intercept α corresponds to group {a}, in quarter I, in its first round, with no supplement survey (and, by definition, cannot be after a supplement survey).

The simple regression model can be presented as $\mathbf{P=XB+E}$, where, **P** is the vector {104*1} of $\mathbf{p_i}$ corresponding to each panel-round **i**, **X** the matrix {104*12} of $\mathbf{X(k)_i}$, each row for each panel-round **i**, **B** the vector {12*1} of $\beta(\mathbf{k})$ of the explanatory variables **k** and the intercept, and, **E** the vector {104*1} of the ε_i for each panel-round **i**. With this, unbiased estimates of the coefficients can be obtained. But, in order to test their significance, in the case of dependence between the observations, it is necessary to use adjusted variances of the regression coefficients. This is done by means of $\mathbf{Var(B)=(X'X)^{-1}X'VX(X'X)^{-1}}$, where, $\mathbf{V=Var(E)}$ is a non-diagonal variance-covariance matrix of the random deviations (for example, see Greene 1993, Ch. 15).

For a preliminary view of the various effects, a simple regression model was used, where all the explanatory variables participated and, the adjusted variances of the regression coefficients were used to test their significance. However, to obtain an appropriate model, based only on explanatory variables with significant coefficients (for a given level), stepwise procedures can be used. In our case, because of the dependence between observations, a simple stepwise regression procedure is not suitable. Thus, a generalized stepwise regression was used, where at the outset, the dependence between repeated investigations of the same panel is taken into account. The generalized regression is performed using the following transformation (Draper and Smith 1966), by means of the matrix **V**:

$$\mathbf{V^{-1/2}\,P = V^{-1/2}\,X\,B + V^{-1/2}\,E}$$

Using the generalized stepwise regression, the "best" model was determined for each nonresponse type separately, based only on explanatory variables with 5% significant coefficients (a 10% level did not make any difference). The models were similar but not the same for all nonresponse types. For the sake of consistency, the estimates of the

regression coefficients for the nonresponse types should sum to those corresponding to the Total Nonresponse. Therefore, one common model, that contained all explanatory variables which participate in at least one stepwise model, was used for all nonresponse types. Then, the regression coefficients for the common model were estimated by using simple regressions. With generalized regression it is possible to predict nonresponse rates for a given panel-round, according to its characteristics and its previous round (this is not, however, the primary aim of this study). To obtain estimates of the effects on nonresponse (which is the goal of this study), a simple regression is more suitable. To examine the significance of these estimates, standard errors were computed, separately for each type of nonresponse, with the required adjustment for the dependence between the observations.

4 The results

In the common model, as explained in section 3, only explanatory variables that were significant for at least one nonresponse type were incorporated, namely, those representing:

- The burden arising from adding a heavy supplement survey.
- All the rounds.
- Only quarter III.
- Panels of the later part of the period - group {b}.

The other explanatory variables, which were not significant for any of the nonresponse types, did not participate, representing:

- The burden arising from adding an easy supplement survey.
- The effect of any supplement survey on following investigations, even if heavy.
- The quarters, except from quarter III.

The results for each nonresponse type, as obtained by the simple regressions based on the common model, are presented in Appendix I. Since a common model was used, not all effects are significant for each type separately. The unshaded β's are significant at 5% level (most of them with a p-value close to zero). A few with lighter shading are hardly significant at $5\% < p \leq 10\%$ level and, the rest (most of them with a very high p-value) have dark shading.

It should be mentioned that the $\mathbf{R}^2$ (measuring the fit of the model) are quite high: 0.69 and 0.75 for Total Nonresponse, with and without Other Nonresponse, 0.55 for Refusals, 0.67 for Absentees, 0.73 for Interviewing Difficulties and, for Other Nonresponse, it is only 0.39.

Comparisons between the predicted rates, as derived from the regression model, for each of the four types under the conditions of panel group {b}, are presented in Appendix II. It should be emphasised that these estimates are based on all variables that took part in the common model, thus not all the differences between the predicted rates are significant. Even so, they give some idea of the general behaviour of the different types of nonresponse.

4.1 Refusals

The overall refusal rate for the whole period was 3.4%, higher (3.6%) for the earlier panels and lower (3.3%) for the later panels, though the difference is not significant.

- Contrary to common belief, the results show no effects on the refusal rate in the LFS due to the burden stemming from supplement surveys, even if heavy. The same is true for cooperation in the following investigations (as in Sharp and Frankel 1983).
- Further, the results show no increase in the refusal rate with the number of round for a given mode of interviewing - face-to-face or telephone. On the contrary, the refusal rate in the third round is significantly lower than in the second (β=-1.1 and β=-1.4), and no significant increase is found in the fourth round from the first (there may even be a slight fall). Thus, from this study, there is no evidence to support the assumption that a greater number of rounds leads to a higher refusal rate. However, this is for a limited number of rounds and may not be true for more rounds.
- There are no significant differences in the Refusals rate between the quarters, as would be expected.
- The results show considerable effects for the second and the third rounds. Compared to the first round, where the average refusal rate is about 4.0% (in group {b}), there is a reduction of 28% in the second round and of 35% in the third. More refusals in the first round can be explained by the difficulties associated with introducing a new sample, so that the interviewer invests less effort to persuade the more stubborn refusals. However, no significant effect is found for the fourth round vis-a-vis the first. Thus, it is reasonable to believe that the effects are of the mode of data collection. More persistent attempts can be made by telephone to gain an interview, as it is much easier and cheaper than conducting more visits to the respondent's house. A reduction of refusals rates by telephone interviews may, however, be true only when the first contact with the respondent is done by a face-to-face interview.

4.2 Absentees

The overall Absentees rate for the whole period was 3.0%. A significant difference is found between the earlier and the later parts of the period: 3.4% for group {a} and 2.6% for group {b}. As was pointed out in section 2, this can be explained, either by the special difficulties in conducting the survey at the time of very large waves of immigration, or by improvements in the field work. In both cases, this is more likely for Absentees rates than for Refusals, as the results show.

- There is room for thinking that there will be some effect on Absentees rate from the heavier workload of the interviewers arising from the extended interviews required for a heavy supplement survey. This may not allow them to devote sufficient time for repeated visits to absentees, since they have to complete their quota in the same time as in other quarters. The results indicate some increase (β=+0.3), though not significant.

- Although it would be reasonable to assume that Absentees rate would increase in the summer months, the results show only a small and hardly significant regression coefficient relating to quarter III (β=+0.3 with **p=7%**). In Israel, besides the summer holiday, there are two other holiday periods each year. Thus, there is the tendency to spread vacations over the year and vacations are usually not that long. Because the holidays are determined by the Hebrew calendar, one is usually in quarter II but may be sometimes in quarter I and, the other may be either in quarter III or IV. In addition, in the Israeli LFS, it is permitted to make use of a proxy, to postpone the interview to a subsequent week and the interviewer leaves a questionnaire for absentees to return by post. Thus, for Israel, the results obtained are reasonable, but it could be that in other countries absentees rates would be higher in the summer months.

- It is not surprising that Absentees rate in the telephone rounds is significantly less than in face-to-face rounds (β=-1.4 for the second round and β=-1.7 for the third). The results show that the interviewing mode is more meaningful for Absentees than for Refusals: a reduction of about 40% in the second round and of about 50% in the third, as compared to the other two rounds, where the average rate is 3.3% (in group {b}).

4.3 Interviewing Difficulties

Interviewing difficulties are mostly due to problems in conducting the interview either due to language difficulties or to physical or mental illness. Although these stem from limitations of the respondents, the resulting nonresponse will depend also on the interviewer effort invested to hold an interview. The overall nonresponse rate was 1.0% for the whole period, and, similar to Absentees, a significant difference is found between the earlier and the later parts of the period: 1.1% and 0.9%, respectively.

- This is the only nonresponse type, for which "almost" a significant effect of heavy supplement surveys can be detected (β=+0.2 with **p**=7%). Maybe, since these cases are the most difficult to interview, the interviewer gives up when he is too busy.
- The results show a significant effect of quarter III for Interviewing Difficulties (β=+0.1), whereas for Absentees there is only some indication (not definitely significant). The explanation for this lies possibly in the extra workload of the interviewer which cannot be reduced by postal questionnaires or postponements, as for Absentees.
- For this nonresponse type, the fall in the two telephone rounds is most substantial (β=-0.9). When compared to the first round, this is a reduction of about 70%. Language problems can be solved in telephone interviews by using interviewers able to converse in the language of the immigrants, whereas in face-to-face interviews this would considerably increase the cost, because the immigrants are spread all over the country. As for the other reasons of this nonresponse type, more contacts may yield a successful interview with another member of the household who would be capable of providing the required information, and, this is easier by telephone.

4.4 Other Nonresponse

This nonresponse type is different from the others in that it is hardly dependent on the respondents or the interviewers. Part stems from unsuccessful attempts to interview by telephone (mostly for technical reasons), part from non-location arising from deficiencies in the sampling frame and the rest from organizational problems. The overall rate was 0.7% with only a slight decrease (not significant) in the later period. Except for the rounds, none of the other effects are significant.

- In contrast to all the previous types of nonresponse, where there is a substantial decrease in the rates in the telephone rounds, for Other Nonresponse there is a significant increase in both the second and the third rounds (β=+0.4).

- Against all nonresponse types, for Other Nonresponse, a significant decrease is found in the fourth round (β=-0.3). This is probably since there are no telephone problems in the fourth round and, many of the non-location problems have already been solved in the meantime.

Thus, the effects of the rounds follow a completely different pattern for this nonresponse type: it is the highest in both the second and the third rounds (1.0%), it is less in the first (0.6%), and the smallest (0.3%) in the fourth round.

4.5 Total Nonresponse

The overall Total Nonresponse rate was 8.1% for the whole period, 8.9% for the early period and 7.5% for the later period, and the difference (β=-1.4%) is significant. This is due mostly to Absentees and to Interviewing Difficulties.

- The effects of the second and the third rounds for all the nonresponse types, except for Other Nonresponse, are all significant and of the same direction. Thus, a considerable fall is observed in the Total Nonresponse for telephone interviews with a larger decrease in the third round than in the second (β=-3.0 and β=-3.7, respectively). When Total Nonresponse without Other Nonresponse is considered, the effects are even greater (β=-3.4 for the second round and β=-4.1 for the third), due to the rounds effects for Other Nonresponse being in the opposite direction. This constitutes about a 40% and 50% reduction of the Total Nonresponse in the second and the third rounds, respectively, as compared to average rate in the first and the fourth rounds (8.6%, as in group {b}).

- For all the nonresponse types, the effect of quarter III is small, and is significant only for Interviewing Difficulties, hardly for Absentees and not significant for the other two types. For Total Nonresponse, there is some indication of a small increase in quarter III, though not significant (β=+0.6 with **p**>16%).

- For Total Nonresponse, as for all nonresponse types separately, both the effects of burden from a heavy supplement survey and of the fourth round are not significant. It should be noted that these effects are in opposite directions and cancel out (however, only to some extent, when Other Nonresponse is excluded), probably because of the high correlation between them due to the Income Survey.

5 Summary

The aim of this study was primarily to examine if there is an increase in nonresponse of a current panel survey, when supplement surveys are annexed:

- For Refusals, contrary to common belief, the results show no effect whatsoever from the extra burden, even when adding a heavy survey, nor on cooperation in investigations that follow supplement surveys.
- Despite the negative effect that may be expected because of the interviewer's additional workload, no significant effect is found for Absentees and, only a very small effect (hardly significant) is detected for Interviewing Difficulties.

The issue of the burden on respondents arising from repeated interviews and the effect on the nonresponse rate, was examined, however, to a limited extent:

- Empirically, for the case of four rounds, no evidence is found to support the assumption of increasing refusal rates with a greater number of rounds. On the contrary, refusal rates decrease significantly in the third round vis-a-vis the second, and a slight fall is indicated for the fourth round against the first.
- Although not discussed in this paper, it should be mentioned that, the same applies to subsequent surveys that make use of the rotating-out panels. For various subsequent surveys in Israel, the refusal rates were similar to that of the LFS.

As for the expectation of more Absentees in the summer months:

- There is some indication of an increase in the Absentees rate in quarter III, though not significant. This may be specific to Israel and may not be true in other countries.

The outstanding finding of this study, is the substantial differences in the nonresponse rates, of all types, between telephone rounds and face-to-face rounds:

- For all nonresponse types, except for Other Nonresponse, the rates are considerably lower in the telephone rounds. For Other Nonresponse they are higher in the telephone rounds, as explained above.

Although the primary purpose of this paper was not to study the effects on the response rate of telephone vis-a-vis face-to-face interviews, the findings suggest quite strongly that nonresponse, of all types, are more affected by the way the interviews are conducted. The ability of the interviewers to cope with the extra workload and their constraints of time and cost to make more visits to hard-to-get respondents, probably play an important role, especially when face-to-face interviews are conducted.

Finally, it should be mentioned that the results obtained from this study may not be conclusive and may not give a complete picture:

- The findings brought here relate only to the 4-round households that did not change their place of residence for the entire duration of the panel. As mentioned in section 2, higher nonresponse rates are associated with households with less eligible rounds and, besides, their nonresponse patterns may be substantially different from that of the 4-round households (Kantorowitz 1994). Therefore, it is planned to extend the study to include households with less than four eligible rounds, although this may cause some complications in the analysis. As for the exclusion of small localities (with less than 2,000 inhabitants) as explained in section 3, this should not influence the main conclusions derived from this study.

- This study was based on a macro analysis of the nonresponse rates. A micro analysis, in addition, could contribute to a better understanding of the effects on nonresponse. Such analysis would be based on households, taking into account also their characteristics and also their responding profiles for the whole duration of the panel, i.e. always respond, respond but not always and, never respond, by the reason for not responding.

- Further analysis of the nonresponse in supplement (or subsequent) surveys is required for a more conclusive policy whether they should be annexed to a panel survey, or conducted separately.

References

Berdie, D.R. (1973). Questionnaire Length and Response Rate. Journal of Applied Statistics, Vol. 58, No. 2

Bogen, K. (1996). The Effect of Questionnaire Length on Response Rates - A Review of the Literature. Proceedings of the Section on Survey Research Methods. American Statistical Association, Chicago, Illinois

Draper, N. and Smith, H. (1966). Applied Regression Analysis. New York: Wiley

Greene, W.H. (1993). Econometric Analysis. New York: Macmillan

Israel Central Bureau of Statistics (1995). Labour Force Surveys. Publication No. 1057

Kantorowitz, M. (1994). Adjustment for Nonresponse in Household Panel Surveys. The 5th International Workshop on Household Survey Nonresponse, Ottawa, Canada

Marquis, K. (1979). Survey Response Rates: Some Trends, Causes and Correlations. In: L. Reeder (ed.): Health Survey Research Methods. Second Biennial Conference, Williamsburg, Va. Washington D.C.: U.S. Goverment Printing Office

Sharp, L.M. and Frankel, J. (1983). Respondent Burden: A Test of Some Common Assumptions. Public Opinion Quarterly, Vol. 47

Statistics Canada (1997). Guide to the Labour Force Survey

Steel, D. (1997). Producing Monthly Estimates of Unemployment and Employment According to the International Labour Office Definition. Journal of the Royal Statistical Society, Series A, Vol. 160, No. 1

U.S. Bureau of the Census (1978). The Current Population Survey - Design and Methodology. Technical Paper No. 40

Appendix I

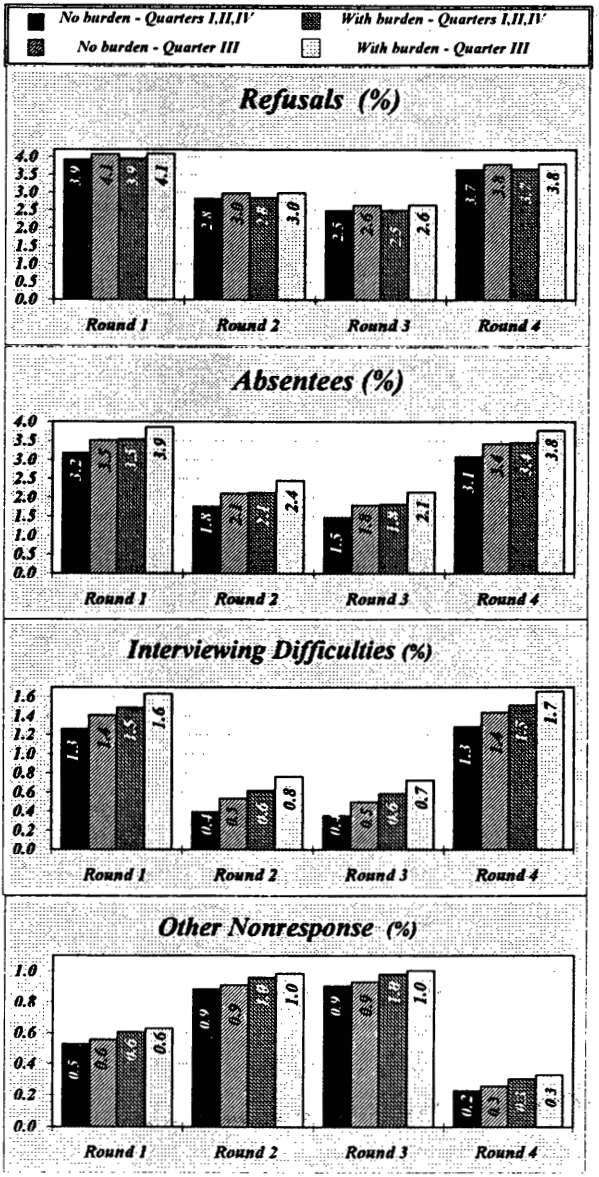
Regression Coefficients for the Common Model

(percentages)

Explanatory Variables				β	s.e.	p-value	β	s.e.	p-value
Panel group	Quarter	Round	Supplement survey	Refusals (R^2=0.55)			Absentees (R^2=0.67)		
(a)	I	1		4.29	0.27	0.000	4.00	0.22	0.000
(b)				-0.35	0.26	0.181	-0.81	0.18	0.000
	III			0.14	0.21	0.510	0.32	0.18	0.066
		2		-1.10	0.17	0.000	-1.41	0.16	0.000
		3		-1.44	0.18	0.000	-1.71	0.17	0.000
		4		-0.28	0.54	0.607	-0.08	0.52	0.871
			Heavy	0.01	0.41	0.971	0.35	0.36	0.332
Panel group	Quarter	Round	Supplement survey	Interviewing Difficulties (R^2=0.73)			Other Nonresponse (R^2=0.39)		
(a)	I	1		1.46	0.07	0.000	0.62	0.05	0.000
(b)				-0.19	0.05	0.000	-0.09	0.05	0.063
	III			0.14	0.06	0.016	0.02	0.04	0.562
		2		-0.87	0.07	0.000	0.35	0.04	0.000
		3		-0.91	0.07	0.000	0.37	0.05	0.000
		4		0.02	0.21	0.968	-0.30	0.11	0.008
			Heavy	0.22	0.12	0.072	0.07	0.11	0.524
Panel group	Quarter	Round	Supplement survey	Total Nonresponse Incl. Other Nonresponse (R^2=0.69)			Total Nonresponse Excl. Other Nonresponse (R^2=0.75)		
(a)	I	1		10.37	0.60	0.000	9.76	0.55	0.000
(b)				-1.44	0.61	0.018	-1.36	0.55	0.014
	III			0.63	0.47	0.184	0.60	0.43	0.159
		2		-3.03	0.33	0.000	-3.38	0.30	0.000
		3		-3.68	0.38	0.000	-4.06	0.36	0.000
		4		-0.64	1.22	0.602	-0.34	1.13	0.767
			Heavy	0.66	0.93	0.480	0.58	0.83	0.483

Appendix II

Predicted Nonresponse Rates in Group {b}, by Rounds, Quarters and Burden due to Heavy Supplement Survey



How Many Mailings Are Enough?

VASJA VEHOVAR AND KATJA LOZAR

Abstract: *This paper addresses the balance between costs and errors in mail surveys. Two alternatives are compared: a larger sample with less follow-up mailings and a smaller sample with more follow-up mailings. The comparison involves a detailed elaboration of mean squared errors and cost functions. Based on the model, the key variables are discussed. The empirical example refers to the percentage of companies with access to the Internet. It is shown that the above elaboration can be helpful for practical decisions.*

Keywords: *survey costs, nonresponse, mail surveys*

1 Introduction

In survey research, we often discuss various procedures for improving the quality of data but rather rarely do we discuss survey costs. However, when quality improvement efforts are discussed in such an isolated form a heavy mismatch between theory and practice may occur (Groves 1989, vi, vii).

In this paper, we pose the following practical question: What is the optimum balance between the errors and the costs of a mail survey?

We use the general understanding of survey errors and survey costs (Kish 1965, Groves 1989). Specifically, we narrow our analysis down to the issues of sample size and number of contacts. Both parameters, large initial sample size and large number of follow-ups, lead to smaller error, but at the same time, they also produce higher costs. We search for the precise balance between survey costs and errors.

Empirically, we present the case of the Total Design Method (TDM) mail survey (Dillman 1978) with three follow-ups (the third conducted by telephone). The key question is whether or not to use the third follow-up. The dilemma can be expressed in the following question: Is it better to have a small initial sample with three follow-ups, or a large initial sample with only two follow-ups? We explore the following factors: the expected nonresponse conversion rate after the 3rd follow-up, the relative costs for the 3rd follow-up and the relative bias after the 2nd follow-up. We also outline some general guidelines for finding the optimal number of contacts.

2 Errors

The components of survey errors have already been well-elaborated (Groves 1989, 3). However, in this paper we limit ourselves to the sampling error and to the component that belongs to the nonresponse bias. We use the standard form of the mean square error (Kish 1965):

$$MSE\,(p_c) = E\{p_c - P\}^2 = E\{p_c - E(p_c)\}^2 + \{E(p_c) - P\}^2,$$

where P is a true population value and p_c is a sample estimate where c runs across all possible samples. The first part represents the variance part and the second part represents the bias part.

The sampling error can be regulated with an increase/decrease of the sample size. The (nonresponse) bias component of the MSE can be, at least in this context, reduced with additional contacts. We will assume a simple random sample (SRS). We will further assume that we are dealing with a simple population parameter - the population percentage P. In our specific example the root mean square error (RMSE) is based on the sum of variance and squared bias:

$$RMSE(p) = \sqrt{MSE(p)} = \sqrt{Var(p) + Bias(p)^2}\ .$$

We have an estimate of $RMSE(p)$ for a population percentage P:

$$rmse(p) = \sqrt{\frac{p(1-p)}{n} + (P-p)^2}\ ,$$

where we omitted the finite population correction. Of course, in the case of continuous variable y we have:

$$rmse\left(\bar{y}\right) = \sqrt{\frac{s^2(y)}{n} + \left(\bar{Y} - \bar{y}\right)^2}\ .$$

The above definition of RMSE(p) is not taking into account the different number of follow-ups. In the case of *v* follow-ups we have[1] the following expression for *rmse(p)*:

$$rmse(p)=\sqrt{\frac{p(1-p)}{n}+(P-p)^2}=\sqrt{\frac{1}{\sum_{i=0}^{v} n_i'}\left[\frac{\sum_{i=0}^{v} p_i n_i'}{\sum_{i=0}^{v} n_1'}*\left(1-\frac{\sum_{i=0}^{v} p_i n_i'}{\sum_{i=0}^{v} n_1'}\right)\right]+\left[P-\frac{\sum_{i=0}^{v} p_i n_i'}{\sum_{i=0}^{v} n_i'}\right]^2}.$$

Here we have n'_i as the achieved sample size in the *i*-th follow-up, p_i as the estimate at the *i*-th follow-up, where i=1... *v*. We have n_i' as the function of the initial sample size n^* and the completion rate at the i-th follow-up, $CR_i = n_i'/n^*$. The RMSE function can be thus rewritten:

$$rmse(p)=\sqrt{\frac{1}{n\sum_{i=0}^{v} CR_i}\left[\frac{\sum_{i=0}^{v} p_i * CR_i}{\sum_{i=0}^{v} CR_i}*\left(1-\frac{\sum_{i=0}^{v} p_i * CR_i}{\sum_{i=0}^{v} CR_i}\right)\right]+\left[P-\frac{\sum_{i=0}^{v} p_i * CR_i}{\sum_{i=0}^{v} CR_i}\right]^2}.$$

RMSE is thus a function of the population percentage *P*, nonresponse bias, completion rates, initial sample size n^* and number of follow-ups *v*. We have - at least in this context – no influence on the value of the population percentage *P* nor can we regulate the sample estimate *p* (i.e. the bias). We also have no influence on the completion rate at the i-th follow-up (*CRi*). On the other hand, we can regulate the initial sample size n^* and the number of follow-ups *v*. In the expressions above we have assumed that all the previous nonrespondents have been included in each additional follow-up.

3 Costs

As we have already mentioned, the changes in the sample size and the number of follow-ups influence the survey costs. The cost function in our example is thus not a continuous one as in the case of one initial contact. In a simplified form it can be written:

$$T = K + C_v = K + (K_v + C_{vn}) = K + (v+1)*A + n^* \sum_{i=0}^{v+1} (1 - CC_i) * c_i \ ,$$

[1] Since *v* is the number of follow-ups the first contact has no follow-up, therefore v = 0.

where the total costs (T) of the survey consist of constant costs (K) and variable costs (C_v). The constant costs include: design of the survey, construction of a questionnaire, data management, data analysis, preparation, printing as well as circulation of survey reports, overhead costs etc. The variable costs have two components:

- the costs K_v that vary with the number of follow-ups (but not with the initial sample size). We assumed that these costs are the same for each follow-up, therefore we can express K_v as the product of number of phases in the survey research process and a certain level of administrative costs $((v+1)*A)$.
- the variable costs C_{vn} that are proportional to the initial sample size n^* and to the variable costs at the i-th follow-up. These costs are the function of the cumulative contact rate (CCi) and ci - the costs per unit within each follow-up[2]. These costs include the costs of paper, envelopes, printing, and a part of the administrative work that depends on the number of mailings (packing etc.). Of course, $CC_0 = CC_1=0$, since all units receive the initial mailing and the reminder. The last component $(i=v+1)$ does not refer to any follow-up but includes only the costs of data entry. Therefore, CCi takes value $CC_{v+1} =(1- n/n^*)$, so that $n^*(1- CC_{v+1})$ gives the number of all responding units which are denoted as n.

4 Optimisation

The aim of the optimal design may be stated in two alternative ways: achieving minimum MSE for fixed costs, or achieving minimum costs for fixed MSE. Both principles would generally lead to the same solution (Kish 1965, 263-264). However, unlike with the standard sampling theory, it is difficult to find the analytical solution when the variable to optimise is a discrete one – the number of contacts v.

In the case of the above two equations (costs, RMSE) we have calculate costs and RMSE for each value of v and then compare the values.

We will concentrate on the optimisation of the RMSE for fixed costs. It is possible to increase the sampling error (with a decrease of n^*, initial sample size) and simultaneously reduce the nonresponse error (with an increase of v, number of contacts), or the other way around, but the total costs must remain the same. Typically, we can have a large initial sample and a small number of follow-ups, or the opposite, a small initial sample and a large number of follow-ups.

[2] At this we have in mind the initial sample since survey costs comprise costs for respondents and nonrespondents.

5 Example

In 1996, a mail survey on the use of Internet among Slovenian companies was conducted as a part of a larger research project *Research on Internet in Slovenia* at the Faculty of Social Sciences (http://www.ris.org). The design was a standard TDM design. The envelopes containing a cover letter, a questionnaire and a return envelope were sent to 4,698 companies. A week later a reminder was sent. After two weeks, the second follow-up with a replacement questionnaire and a return envelope was sent to the nonrespondents. Three weeks later CATI (computer assisted telephone interviewing) follow-up was performed among the nonrespondents. The response rate after three follow-ups was 67.9% and the completion rate was 60.0%. In reality, only 10% of the nonrespondents were contacted by telephone. However, in the calculations below we assumed that all the nonrespondents were surveyed. Without this simplification the model would become much more complicated as we would introduce another variable in order to optimise - the sub-sampling rate for the third follow-up.

The key population parameter in the survey was a percentage of companies with the access to Internet. Since the true population value was unknown we have assumed that the true value was the value achieved after the last follow-up. The bias after the 3rd follow-up was therefore automatically set to zero.

Table 1: Optimum design in a mail survey

No. of follow-ups	estimate p of P	costs for a unit of init. sample	cumul. costs for a unit of initial s.	% of total costs	cumul. % of total costs	Cumulat-ive compl. rate	initial sample size	final sample size (resp.)	rmse(p) - estimate of RMSE
0	18.0%	47.5	47.5	23.6	23.6	18.2%	19,880	3,618	0.040506
1	16.6%	26	73.5	12.9	36.5	26.5%	12,848	3,405	0.026770
2	14.1%	35.5	109	17.7	54.2	44.2%	8663	3,829	0.005712
3	14.0%	92	201	45.8	100.0	60.0%	4698	2,819	0.006535

We can observe that after three follow-ups 14.0% of companies had access to Internet. Without the telephone follow-up the estimate would be 14.1%. Of course, these are cumulative percentages; the percentage for the 3rd wave respondents alone is lower than 14.0%. If we performed only one follow-up the estimated percentage would be 16.6%, and without any follow-ups the estimate would be 18.0%. Obviously, a strong nonresponse bias exists, however the decision concerning the optimal number of contacts

is not that obvious at all, unless we perform some calculations.

All designs in Table 1 assume the same fixed budged. For the budget needed for three follow-ups and an initial sample of 4,698 units, we could omit the telephone follow-up and enlarge the initial sample size to 8,663. If we omitted the 3rd and the 2nd follow-up, we could have an initial sample of 12,848 companies. In case of no follow-up we could send the mailing to 19,880 companies. In all these cases the bias and the sample variance vary considerable. However, the optimum for this fixed budget is achieved in the case of two follow-ups (the smallest RMSE).

6 Generalisation

The example above was, no doubt, a very specific one. However, a general solution can not be derived analytically, so we would like to find the basic principles by varying the parameters in the above example. For this purpose we will alter one of the variables while keeping the others constant. We will concentrate on the decision whether to use the 3rd follow-up or not. We will thus compare the situation after the second and after third follow-up. Three key parameters are important in this process:

1. relative costs of the third follow-up in comparison to the costs of the first two follow-ups,
2. bias after the 2nd follow-up,
3. the nonresponse conversion rate in the 3rd follow-up.

A) The costs for the 3rd follow-up

How high can the costs of the 3rd follow-up be (in relation to the costs of previous contacts), so that the use of 3rd follow-up would reach the optimal RMSE?

If we have the same fixed costs for both situations, with two and with three follow-ups, this will obviously create a difference in the initial sample size. Therefore, the important variable here is the sample size, which is fixed at 4,698 for two follow-ups. That is why the corresponding RMSE line is **constant** for two follow-ups. Of course with a 3rd follow-up the RMSE changes according to the relative costs of the 3rd follow-up. From Figure 1 we can see that RMSE is smaller for 3 then for 2 follow-ups only if the costs for the 3rd follow-up are less then 40% of the costs for two follow-ups. The 3rd follow-up is thus optimal only if its costs will not increase the previous costs for more then 40%[3].

[3] In our case two follow-ups was the best decision, because costs for the 3rd follow-up presented more then 84% of the costs for first two follow-ups.

Figure 1: Influence of the relative costs of the 3rd follow-up[4]

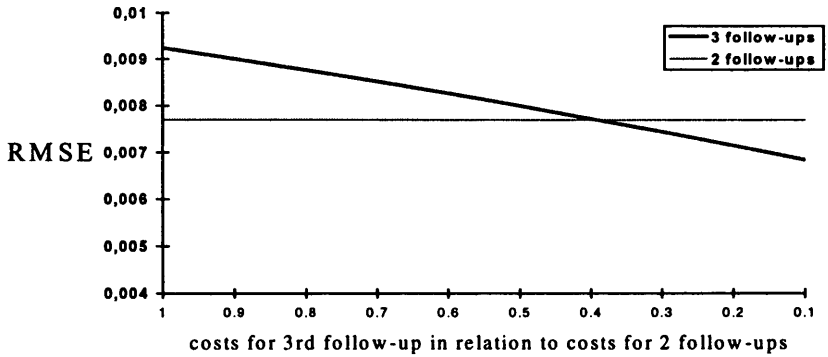


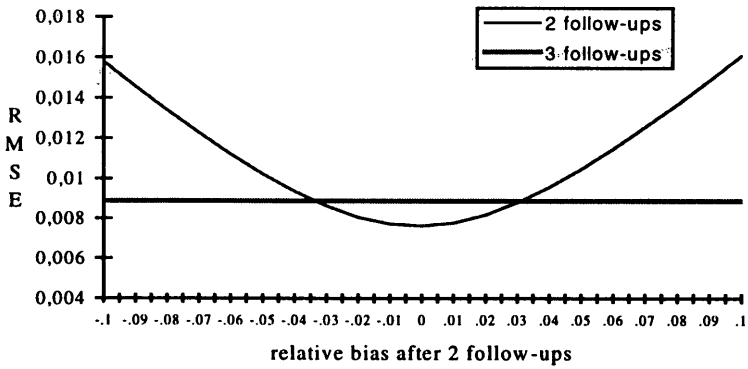
Figure 1 presents only the relationship when a relative bias after the 2nd follow-up takes a specific value of 1%. If the relative bias after the 2nd follow-up is larger, the line presenting the corresponding RMSE will be higher. In such a case the 3rd follow-up can be reasonable also in the case of higher relative costs.

B) The bias after two follow-ups

How large should the relative bias be after two follow-ups in order to justify the use of an additional contact? In this situation the RMSE for the 3rd follow-up is **constant**, as we assume no bias after the last follow-up.

[4] The constant factors in this case are: costs for two follow-ups - 109 SIT, estimated percentage after 2nd follow-up - 0.141, estimated percentage after 3rd follow-up - 0.140, relative bias after 2nd follow-up - 0.7%, completion rate after 2nd follow-up - 0.442, completion rate after 3rd follow-up - 0.600.

Figure 2: Influence of the relative bias after two follow-ups[5]



The variable factor is the relative bias after two follow-ups. It changes from -10% to +10%. We can observe that RMSE is smaller for three then for two follow-ups if the relative bias after two follow-ups is larger than 0.035[6].

Again, the above figure shows only a specific case where the costs for the 3rd follow-up represent 84% of the costs for previous follow-ups. If the costs for the 3rd follow-up would be smaller, the initial sample size for 3 follow-ups could be larger - the horizontal line would be lower, with different intercepts.

C) The nonresponse conversion rate after the 3rd follow-up

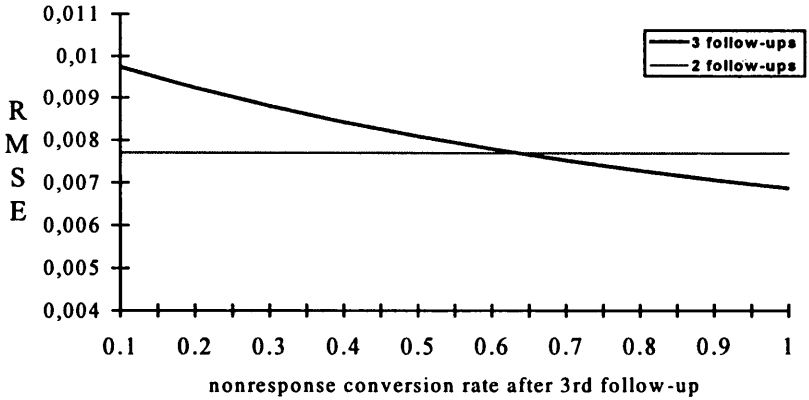
Another factor that can influence the decision as regards the 3rd follow-up is the nonresponse conversion rate after the 3rd follow-up. How large should this conversion be in order to use the 3rd follow-up?

---

[5] The constant factors in this case are costs for two follow-ups - 109 SIT, costs for three follow-ups: 210 SIT, initial sample size for two follow-ups – 4698, initial sample size for three follow-ups - 2548. In this case the total costs for 2 or 3 follow-ups are the same. The completion rate after the 2nd follow-up - 0.442, completion rate after the 3rd follow-up - 0.600, estimated percentage after 3rd follow-up - 0.141, relative bias after 3rd follow-up - 0.

[6] In our case the relative bias after two follow-ups was less then 1%, so the 3rd follow-up was not needed.

Figure 3: The impact of the expected nonresponse conversion rate[7]



The variable factor is the nonresponse conversion rate after the 3rd follow-up. This rate influences the total completion rate and therefore also the final sample sizes. The larger the conversion rate, the larger the final sample and smaller sampling variance and RMSE. Of course, this variable factor has no impact on the completion rate in case of two follow-ups, so this line is a **constant**.

We can see that RMSE is smaller for three then for two follow-ups when nonresponse conversion rate after the 3rd follow-up is larger then 60%[8].

7 Conclusion

We have demonstrated the impact of various factors effecting the decision of whether or not to use a third follow-up contact in a mail survey. The same principles apply to any previous or additional follow-up. Of course, the relationship between the parameters involved is complex and depends on many specific circumstances. Other factors also may

[7] The constant factors in this case are costs for two follow-ups - 109 SIT, costs for three follow-ups - 210 SIT, initial sample size for two follow-ups – 4698, initial sample size for three follow-ups – 2548, estimated percentage after 2 follow-ups - 0.141, estimated percentage after 3 follow-ups - 0.140, completion rate after 2 follow-ups - 0.442. In this case total costs for 2 or 3 follow-ups are the same.

[8] In our case the nonresponse conversion rate after the 3rd follow-up was 24.8% what increased the completion rate from 44.2% to 60%. As the nonresponse conversion rate was small, the 3rd follow-up was not worthwhile to use.

play an important role, such as time constraints or low quality of late responses[9]. However, when faced with a clear dilemma between sampling error and nonresponse bias the above results can be useful.

For the simultaneous understanding of all three factors together (bias, costs, response) a multivariate presentation in a three-dimensional space may be helpful. In such a space a sort of pyramid can be drawn. Only within the body of such a pyramid can the parameters take on values that may justify the use of the third follow-up.

It is somewhat difficult to perform the above calculation in practice. One obvious complication is the case of different sub-populations which behave differently. Another obstacle may be that we have no information about the bias and the nonresponse rates. There may even be difficulties with the accurate anticipation of the costs. Of course, in such situations a good decision cannot be reached. However, it is reasonable to make certain estimates from previous surveys or, at least an educated guess. It is also possible to make estimates from earlier stages of the same survey. Based on these assumptions we can - with the aid of the above-described model - obtain a better understanding of the interaction between costs and errors in mail surveys.

References

Dillman, D.A. (1972). Increasing Mail Questionnaire Response in Large Samples of the General Public. Public Opinion Quarterly, 36, 2, pp. 254-257

Dillman D.A. (1978). Mail and Telephone Survey. The Total Design Method. New York: John Wiley and Sons

Dillman, D.A. (1989). Our Changing Sample Survey Technologies. Choices, 4, 4, pp. 40-41

Dillman, D.A. (1991). The Design and Administration of Mail Surveys. Annual Review of Sociology, 17, pp. 225-249

Dillman, D.A., Clark, J.R. and Treat, J.B. (1994). Influence of 13 Design Factors on Completion Rates to Decennial Cencus Questionnaires. Paper presented at the 1994 Annual Research Conference of the U.S. Bureau of the Cencus, Arlington, Virginia, March 21, 1994

Green, K.E. (1991). Reluctant Respondents: Differences Between Early, Late, and Nonresponders to a Mail Survey. Journal of Experimental Education, 59, 3, pp. 268-276

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: John Wiley and Sons

[9] Several authors recognise that late respondents are less interested in the survey topic and less willing to cooperate in the survey (Green 1991, 268-276; Kojetin et al. 1993, 838-843).

Kish, L. (1965). Survey Sampling. New York: John Wiley and Sons

Kojetin, B.A., Borgida, E. and Snyder, M. (1993). Survey Topic Involvement and Nonresponse Bias. Proceeding of the Section on Survey Research Method, Volume II., American Statistical Association, pp. 838-843

Martin, C.L. (1994). The Impact of Topic Interest on Mail Survey Response Behaviour. Journal of the Market Research Society, 36, 4, pp. 327-338

Pearl, D.K. and Fairley, D. (1985). Testing for the Potential for Nonresponse Bias in Sample Surveys. Public Opinion Quarterly, 49, 4, pp. 553-560

Roeher, G.A. (1963). Effective Techniques in Increasing Response to Mailed Questionnaires. Public Opinion Quarterly, 27, 2, pp. 299-302

Improving Advance Letters for Major Government Surveys

AMANDA WHITE, JEAN MARTIN, NIKKI BENNETT AND STEPHANIE FREETH

Abstract: *This paper reports the latest results in a project examining the role of the advance letter in major government surveys. Results of cognitive interviews with members of the public showed that subjects preferred short letters. They wanted the letter to include the purpose of the survey, the name of the survey organisation, a simple statement about confidentiality, details of the interviewer's visit, and past survey results. Interviewers' views were sought using focus groups. Results from these stages were used to redesign the advance letters of the surveys. A three way split sample experiment was carried out on the Family Resources Survey, original letter versus new letter versus new letter plus leaflet. Preliminary results showed that response rates were slightly higher for the new letter compared to the original, but the new letter plus the leaflet did not show much of an increase on the original.*

Keywords: *advance letters, response rates, cognitive interviews, focus groups, split-sample experiment*

1 Introduction

This report presents further findings from a project looking at the role of advance letters in the survey process that was first described at the 7th International Workshop on Household Survey Nonresponse (White and Freeth 1996). The project was set up because of concerns over falling response rates on several of the large scale continuous government household surveys carried out by Social Survey Division (SSD) of the Office for National Statistics (ONS). The aim of the project was to evaluate the current advance letters used on seven of our major continuous surveys[1] which are sent to all sampled addresses; to use information from the evaluation process to design improvements to the letters; and to test the improved letters designed to encourage people to co-operate with the survey request.

---

[1] The National Travel Survey (NTS), the General Household Survey (GHS), the Family Expenditure Survey (FES), the Family Resources Survey (FRS), the National Food Survey (NFS), the Survey of English Housing (SEH), and the Omnibus Survey.

Response rates have been falling in spite of extra expenditure on measures to increase response, such as reissuing addresses not contacted during the original field period or extending the field period, and mirrors experiences in other survey organisations both in Britain and abroad (Steeth 1981; Lievesley 1986; Bradburn 1992). Trends in living patterns and attitudes have increased the problems faced by research agencies in obtaining acceptable levels of response in two main ways. It has become more difficult to find people at home, and it has become more difficult to persuade people, once contacted, to take part.

There are a number of ways of making contact with potential respondents on personal interview surveys to provide information about the survey and to persuade them to take part which depend partly on what information is available about the selected units. Most surveys carried out by ONS are based on a sample of addresses with no easy means of finding out the names or telephone numbers of respondents in advance so the main decisions are about what to send in writing before the interview, what the interviewer should cover verbally and what should be provided in writing when the interviewer calls.

There are three main reasons for sending a letter in advance of the interviewer's visit:

- to announce the visit of an interviewer;
- to give information about the survey;
- to stimulate the willingness of the respondent to participate in the survey.

When telephone calls are not feasible, only advance letters can cover the first of these. Information about the survey can be provided in a number of ways, so the issue here is what information should an advance letter provide and in what detail. The main means of persuading people to co-operate on surveys is personal contact with an interviewer. An advance letter can facilitate this by ensuring the respondent answers the door, listens to what the interviewer has to say and is positively disposed to take part. In addition, there is an SSD requirement to state formally to respondents that the survey is voluntary and that the information collected is confidential; in recent years it has been understood that such assurances must be provided in writing.

In SSD advance letters have been used routinely on all major surveys since 1987. Prior to this, experiments were conducted (Clarke et al. 1987) which showed that sending an advance letter improved response and was popular with interviewers. However, there has been little methodological investigation since into their effect on response, or into how the letters are viewed by either respondents or interviewers. Systematic evaluations of the effect of advance letters can be expensive to set up as they often require large split sample designs, with the sample being matched on variables such as area type, interviewer characteristics, respondent characteristics etc. Even then it is often very difficult to

attribute a difference in response rate to the effect of the letter. For this reason it was decided to use a variety of different approaches to take an in-depth look at the role of advance letters in the survey process.

Various criteria were used to evaluate the current letters used on the surveys:

- conformance with compliance principles known to encourage response
- intelligibility to respondents
- respondents' and interviewers' preferences on content and style
- effect on response

The project was carried out in a number of stages using different methods in combination:

i) Asking respondents a short series of structured questions about the advance letter on a quantitative survey
ii) Content analysis of current advance letters
iii) Cognitive interviews with members of the public to evaluate new letters against old
iv) Focus group discussions with interviewers comparing new and old letters
v) Formal field test of new letters

The previous paper presented at the last nonresponse workshop described stages i) and ii) of the project. In summary, the findings from the Omnibus work suggested that respondents do not want a letter any longer than the current one; any redesigned letter should aim to explain the purpose of the survey more clearly, including more about the sorts of information collected, and the confidentiality statement should be made more convincing in some way. The content analysis (stage 2) revealed gaps and some inconsistencies in the letters currently in use and little explicit referral to compliance principles, described in detail in the previous article. This information served as a useful aid when drafting alternative versions of the letter to use in the cognitive test stage of the project.

This paper describes the findings from the cognitive interviews with members of the public (stage iii) and the focus group discussions held with interviewers (stage iv). These and earlier findings provided useful information for redrafting advance letters on the major government surveys conducted by SSD. The formal field test to evaluate the effect of the new letters is currently in progress.

2 Cognitive interviews with members of the public

2.1 Method

Luppes (1994; 1995) also used cognitive interviews to help analyse and improve the design of advance letters and our approach was based on his. The method can be used to investigate:

- reactions to a letter
- interpretation of a letter
- whether the intended message has been understood
- what information people would like to be provided
- the linguistic style people would prefer

Cognitive interviewing techniques are not designed for the simultaneous assessment of multiple versions of letters from a large number of surveys so it was decided to assess the letter from one survey only and use the information gained to aid the redesign of letters for the other surveys. The survey chosen was the National Travel Survey (NTS). This survey requires all members of the household to be interviewed and to keep a one week travel diary; all members must co-operate fully for the household to be counted as responding so anything that will encourage participation is of potential interest. Unlike the Family Expenditure Survey, which also involves diary keeping, no monetary incentive is offered. The original NTS advance letter is shown in the Appendix (letter A).

As well as using results from the content analysis and the structured interviews, the design of the alternative letters to be used in the experiment was influenced by a number of other issues. Informal discussions with field supervisors and interviewers revealed that most letters were thought to be too long, were not direct enough about the purpose of the survey and tried to give too much information. The general view was that the letters should provide the minimum necessary to encourage respondents to open the door to the interviewer; anything else was better covered by the interviewer on the doorstep or in a leaflet providing further information. However, because of the importance attached to the confidentiality statement, we were interested in exploring further what people understood by it and whether they believed it.

With these issues in mind two alternative versions to the original NTS letter, a short and a long version were designed (Appendix letters B and C). The short version was designed to cover the absolute minimum information needed in an advance letter and used a direct style with short sentences. The long version, while covering the same topics, went into more detail, particularly about confidentiality. It also included some previous results from the survey on the back of the letter. The alternative versions were designed to be distinctly different in style from the current letter, drawing on the findings of the earlier work and appealing to compliance principles. Figure 1 summarises the key features of the three letters.

Figure 1: Key features of the letters used in the advance letters cognitive test

Original NTS letter	Short version	Long version
Topics covered:		
	• survey organisation	• survey organisation
	• survey sponsor	• survey sponsor
• survey objective	• survey objective	• survey objective
• information to be collected	• information to be collected	• information to be collected
• confidentiality	• confidentiality	• confidentiality
• participation is voluntary	• participation is voluntary	• participation is voluntary
		• past survey results
		• the survey is used by organisations outside government
• sampling procedure		
Compliance principles and tendencies invoked:		
• helping tendency	• helping tendency	• helping tendency
• request comes from an authority	• request comes from an authority	• request comes from an authority
	• feeling of uniqueness	• feeling of uniqueness
	• reward is given or perceived	• reward is given or perceived
Distinguishing features:		
• long sentences	• short	• detailed explanation
	• direct style	• inclusion of past results
	• large print	• detailed confidentiality statement

Eighteen individual interviews were carried out by a researcher trained in cognitive interviewing skills. Subjects were recruited into three groups with similar age and sex profiles. To ensure fair assessments, each group was given a different one of the three versions first. They were asked to read the letter and were asked questions about their comprehension using the 'retrospective probing' method (Willis 1994). They were then asked questions covering the following issues:

- overall impression of the letter
- perceived tone of the letter
- the information the letter was thought to provide
- the information the letter should provide
- use of additional material such as survey results

Finally the subjects were shown the two other versions of the letter and asked to compare all three. Interviews lasted about an hour and a half, and were tape recorded. Analysis was carried out of the transcriptions.

2.2 Results of cognitive interviews on NTS letters

Understanding of the letters

All three letters were understood by the subjects who could remember the main points; misinterpretations were rare and minor, although one person thought the letter was from the Department of Transport, one thought he had to ring the ONS number to make an appointment and one thought she was being asked to become an interviewer (but soon realised the misunderstanding).

Overall impressions

Both the original letter and the long new version were thought too long; the short one was about right. The salutation "Dear Residents" was not liked as it suggested the letter was a circular. The original letter was the least popular. Subjects thought it too long and too officious, giving the impression the survey was compulsory. The opening sentence referring to the media suggested that they ought to have heard of the NTS when in fact no-one had. Several commented that telephone numbers were given in the wrong place. Subjects thought the confidentiality pledge was too wordy, raising rather than alleviating concerns about information disclosure. They thought the letter implied, the first contact by the interviewer would be by letter or phone rather than in person.

Although the short letter was thought to contain sufficient information and was preferred to the original letter, subjects disliked the short sentences and paragraphs, preferring the style of the long letter. This was the most preferred letter; subjects thought it clear and well laid out and liked the use of bullet points. They also liked the inclusion of past results. However, they did think it was too long and the print too small. Several people said they would like to select the best features of each letter and combine them to form their ideal letter.

Positive and negative expectations

Advance letters can raise both positive and negative expectations about taking part in a survey so it is important to cultivate the former and reduce the latter. The cognitive interviews indicated that subjects had a number of concerns which would influence whether they took part in the survey. These are summarised below, with suggestions of how they might be addressed.

Allaying concerns:

- Concerns about ONS - give information about its functions that respondents will relate to (e.g. the Census, registration of births and deaths).
- Concerns about the sponsoring government department - explain how the results will be used, particularly to make improvements that might affect them.
- Concerns about confidentiality - need a statement that identifiable information will not be released but not as long winded as in the original letter or in the long version. Subjects understood that data would be published in aggregate form but few had thought about the release of anonymised microdata so they did not understand the reference to this in the long version; it raised more concerns than it allayed.
- Concerns about intrusion - provide details about the interviewer's visit. Subjects wanted a clear idea of what was involved in taking part in the survey. It should be made clearer that an interviewer will call in person rather than telephoning first.

Increasing positive expectations

The interviews also revealed ways in which people's positive expectations could be increased in order to raise positive feelings about participating in the survey:

- Emphasise the purpose and interest of the survey in a way that will appeal to respondents. Make it clear that everyone's experiences and views are important - so people who don't travel much feel it is still relevant to them.
- Appeal to compliance principles[2], particularly altruism. However, the scarcity principle - saying 'you are representing thousands' - could backfire as people did not think they were capable of doing so or thought they were unlike anyone else. Something like: 'You are representing thousands of households with similar characteristics to yours' might be better.
- The ONS logo was important to give authority, together with the address and a contact number. But only one number should be given and placed in the body of the letter. (Problems were caused by using standard SSD letter paper which has a general inquiries number at the bottom.)

[2] These are described in detail in the previous paper (see White and Freeth 1996).

- Voluntary nature of the survey - this has to be stated and most subjects thought this important.
- Signature on the letter - subjects thought the letter should be signed by someone at headquarters (not the interviewer). The term 'survey manager' was preferred to 'field officer' which had associations with the army or with the countryside.

What should letters include?

When asked what information they thought their ideal letter should include, the most common items mentioned were:

- the name of the organisation carrying out the survey;
- the purpose of the survey;
- details about the interviewer's visit;
- the information required;
- a statement saying the information would be treated 'in confidence';
- past results from the survey showing how the information is used.

2.3 Conclusions from the cognitive testing

From the above results we concluded that the ideal letter should be quite short, conveying the main items of information required to allay concerns and encourage participation. Inclusion of past results was popular but these need not be included in the letter - a separate leaflet is a possibility.

The main recommendations were:

- Name the survey organisation and give a brief description of its responsibilities.
- Specify the purpose of the survey. If appropriate, raise respondents' interest by specifying improvements planned or made as a result of the survey.
- If appropriate, mention that the survey results can be used by organisations outside government.
- Avoid complicated confidentiality statements which refer to local authority areas.
- Provide details about the interviewer's visit.
- Appeal to respondent's helping tendency by suggesting that their participation would benefit society at large rather than the government.
- Give respondents the feeling that they are in an unique position by using statements such as "You are representing thousands of households with similar characteristics to yours".
- Provide one contact telephone number rather than two or three.
- Use direct line telephone numbers rather than the switch board number.
- Change the job title of the Field Work Co-ordinator to "Survey Manager".

It is important to recognise that although the views of potential respondents are important, they are not the sole determinants of what an advance letter should cover. The letters are only one aspect of the process of securing co-operation and the subjects interviewed just about the letters were not in a position to judge what should be left to interviewers to convey on the doorstep rather than in a letter. Interviewers' views about what the letters should include in order to facilitate their task on first contacting the address are also important, so the next stage of the project was to consult interviewers.

3 Focus groups with interviewers

3.1 Method

While the preceding stages of the project were in progress, suggestions generated at regular interviewer support meetings had been passed to the research team. Together with the findings from the cognitive testing, these influenced the design of two further letters (D and E in the Appendix) which were sent out to nine interviewers who were invited to attend a focus group.

The first version (D) was a basic redesign of the letter to be sent out by headquarters. The second (E) consisted of two elements: a short letter to be sent out by HQ in advance of fieldwork and a longer letter to be sent out by interviewers at the start of the field period.

3.2 Results from the interviewer focus groups

Although interviewers thought both letters were an improvement on the current advance letters, they had further comments and suggestions for improvement. The main comments and recommendations that emerged from the focus groups were:

- Both versions D and E were considered too long, too ponderous and too officious.
- Version D could be much shorter by omitting how and by whom the results are used; the interviewer can provide this information on the doorstep.
- Version E: many interviewers deliver their own personalised notes which they adapt to specific circumstances; some give their telephone numbers so the informant can contact them. On balance, it was felt that the current system worked well and the proposed letter in version E was rather inflexible.
- The inclusion of the crown on the headed paper was much appreciated.
- The sentence about 'you can help us to produce reliable statistics which everyone can use to understand what is going on in our society' was liked and interviewers felt it should be used in the new letters.

- Details of how addresses are selected should be left to interviewers to deal with as this could easily be misunderstood by informants.
- It is important to make an impact in the first few sentences as many people don't read the letters beyond a certain point.
- Keep the letter short and simple as this provides less for people to take exception to and call HQ to refuse.
- The key points to mention in the advance letter are:
  - who ONS is (and mention the Census);
  - what the survey is about (briefly);
  - introduce the interviewer's visit;
  - mention the interviewer's ID card;
  - ask them to show the letter to others in the household;
  - stress that their participation is important;
  - the survey is confidential (but keep the statement simple);
  - the survey is voluntary (this must be included);
- Leave people wanting to know more; allow the interviewer to play a bigger role in gaining response - people are less likely to turn down a friendly face than a detailed letter.
- Avoid using the term 'survey' too often and replace with 'study' or 'research'.
- Use even bigger typeface.
- Remove 'The Resident' from the address label to encourage people to open the envelope.

Prior to the focus group a couple of the interviewers due to attend had drafted their own versions of advance letters. These were discussed by the group and several features of these letters were adopted. In particular, interviewers wanted a statement to help authenticate their visit, mentioning that their ID card carries their photograph and the ONS logo. They also wanted to reassure people that if they were not in when the interviewer first called or were busy, the interviewer would be happy to call again. This would help emphasise the importance of co-operation.

Survey leaflets

Interviewers expressed mixed views about whether we should enclose leaflets with advance letters giving more detail about the survey and presenting some results. This procedure is currently used on the Labour Force Survey. Many felt that the leaflets are a useful tool for them to have in their armoury that helps them gain response. A leaflet is often left by the interviewer at a reluctant household and he/she will return later to try to gain response. In addition, the leaflet is often left at addresses where no contact has been made together with a note to say that the interviewer has called and that he/she will return. Some interviewers were sceptical about whether people would read the leaflet if it was

sent out with the letter. Others felt that a letter and leaflet may overload the informant and they may be more inclined to refuse to HQ. However, this has not been the experience on the LFS. We therefore need to do some methodological work in the field to explore the effect of sending leaflets with the letters on interviewers' behaviour and on response.

4 Field trials

4.1 Method

Based on all the preceding work a final version of the letter (F in the Appendix) was drawn up to test in the field. Ideally we would have liked to carry out a split sample test of this letter against the original letter on the National Travel Survey, on which so much of the work had taken place. However, for a variety of reasons it was not possible to use this survey for an experiment so the new letter was adopted without further testing.

On the basis of the research described here advance letters for all the other major surveys were redesigned and in general there was little enthusiasm for allowing experimental comparisons with the original letters. However, we have done a number of things to evaluate how well the letters appear to be working in the field. We have obtained feedback from interviewers about how well they think the new letters are working and are monitoring the number and nature of calls to HQ in response to the letters as well as overall response rates.

It was agreed to conduct a three-way split sample trial of the advance letter on one survey, the Family Resources Survey. The three elements are:

- original letter
- new letter
- new letter plus survey leaflet

Under the first two conditions interviewers have copies of the survey leaflet to use at their discretion when they visit the address; under the third condition a leaflet is sent with the letter in advance of the interviewers' visits. The split sample trial is being carried out for three months, June - August 1997, covering a sample of around 2,000 households per month.

4.2 Results from the field trials

The field trials started in June 1997 so final results are not yet available. Results to date are presented below.

General observations

Feedback from interviewers

All interviewers working on surveys which introduced new advance letters in June or July 1997 were asked to complete an evaluation form giving their views on the letter. Views were obtained from over 230 interviewers. Interviewers reported that the public had few specific comments to make about the letters. Of the few comments received most were positive, remembering the letter, particularly the payment for taking part in the FES. The interviewers' overwhelming impressions were that the new letters were much improved on the old. They particularly liked the short length, the clear, concise nature, and the overall 'official' look created by the crest and the ONS logo at the top of the paper. They did not think that the letter had actually improved response directly but felt it had paved the way for them to approach the household and gain co-operation.

We deliberately involved interviewers in this project; we listened to their criticisms and ideas and sought their opinions in redesigning the letters. Although it is somewhat intangible to measure, the motivation of interviewers does seem to have improved as revealed by their favourable comments. They feel that staff in HQ have listened to their problems and done something about them. One should not underestimate the effect that this can have on their enthusiasm and performance in their jobs and hence on obtaining high response.

Telephone calls to HQ

One criticism of using advance letters is that they give survey respondents the opportunity to refuse direct to HQ by telephone before the interviewer ever has a chance of a face-to-face contact where she/he can try and persuade those that are reluctant. Thus it was important to monitor telephone calls to HQ when the new letters were introduced to ensure that there was no dramatic rise in this type of refusal. There has been no evidence to date that either the number or nature of the calls were any different to previously.

Response rates

Obviously the new letters were designed to improve response rates. Although it is virtually impossible to attribute any change in response directly to the letters we wanted to make sure that response had not changed dramatically in a negative direction since their introduction. To date we only have final response rates available for June which show little difference from those obtained throughout the last 12 months. Interestingly the only survey which showed any substantial difference was the NTS: the response rate for June 1997 was 73% which was the highest since September 1996 (75%) and at least 3 percentage points higher than the previous eight months (range 67% - 70%). It is too early to draw any firm conclusions other than the new letters do not appear to be having any adverse effect.

Split sample trial on the FRS

Again response rates for only one of the three months of the split sample trial are available. Response rates for June are shown in Table 1.

Table 1: Response rates on the Family Resources Survey by type of advance letter

	June 1997		
	Original letter	**New letter**	**New letter + leaflet**
	%	%	%
Response	69	71	69
HQ Refusals	2	2	2
Refusals to interviewer	26	24	26
Total Refusals	28	26	28
Non-contact	3	3	3
Base	*967*	*949*	*913*

As can be seen from Table 1 the sample that received the new letter had a slightly higher overall response rate than either that which received the original letter or the new letter plus the survey leaflet, 71% compared with 69%. This difference was due to a lower level of refusals to the interviewer. It is too early to draw any general conclusions but again the new letter does not appear to be having a detrimental effect.

Interviewers were asked how they felt about a leaflet, explaining the purpose of the survey and giving some results, being sent out with the advance letter. Interestingly, about half the interviewers felt it was a good idea but a substantial minority (a quarter) thought it was best left for the interviewer to handle and give out when she/he felt it was most appropriate, usually at the end of the interview.

5 Conclusions

The results so far indicate that the interviewers are certainly pleased with the new letters much preferring them to the old versions and hence the letter has had the effect of motivating them in their job of obtaining high response to government surveys. As yet no detrimental effect on overall response rates has been observed but we must still monitor results over the next few months. There is a question mark over whether to send a leaflet

of results in advance with the letter, interviewers working on the FRS have divided opinions over this and we must await the remaining response results on the split sample trial.

References

Bradburn, N.M. (1992). What have we learned? In: Schwarz, N. and Sudman, S. (eds.): Context Effects in Social and Psychological Research. New York: Springer Verlag, pp. 315-323

Clarke, L., Phibbs, M., Klepacz, A. and Griffiths, D. (1987). General Household Survey advance letter experiment. Survey Methodology Bulletin 21 (Sep 1987). London: OPCS

Lievesley, D. (1983). Reducing unit nonresponse in interview surveys, unpublished paper, London: SCPR

Luppes, M. (1994). Interpretation and evaluation of advance letters. Paper presented at the Fifth International Workshop on Household Nonresponse, Ottawa, 1994

Luppes, M. (1995). A content analysis on advance letters from expenditure surveys of seven countries. Journal of Official Statistics, 11, 4, pp. 461-480

Steeth, C.G. (1981). Trends in nonresponse rates 1952-1979. Public Opinion Quarterly, 45

White, A. and Freeth, S. (1996). Improving advance Letters. Paper presented at the 7th International Workshop on Household Survey Nonresponse, Rome, 1996

Willis, G.B. (1994). Cognitive interviewing and questionnaire design: a training manual. National Center for Health Statistics Cognitive Methods Staff Working Paper Series, No. 7

Annexes: Copies of letters

Letter A	Original NTS letter
Letter B	New short version letter for cognitive interviews
Letter C	New long version letter for cognitive interviews
Letter D	Short letter for focus group
Letter E	Long letter for focus group
Letter F	Final letter for field trials

Letter A Original NTS letter

OFFICE FOR NATIONAL STATISTICS

Social Survey Division

Our Ref: NTS

Date:

Dear Resident(s)

You may have read in the newspapers or heard on the radio or television references to the National Travel Survey. This is one of the major sources of information used by the Department of Transport when considering national policy and use of resources in such fields as public transport, road building and railways. It is a survey carried out every year to provide the Department with facts and figures about travel in Britain.

This month your address is one of those selected randomly from the Post Office's list of addresses. I am therefore writing to ask for your help.

We are interested in such things as whether or not people travel, which methods of transport they use and how much it costs them.

Within the next few weeks, one of our interviewers will call on you. The interviewer will show an official identification card, explain the survey to you in more detail and ask to talk to each adult in your household. If you happen to be busy when the interviewer calls he/she will be happy to call again.

The information you give is treated in confidence. The Office for National Statistics does not release this information in any way in which it can be associated with your name or address. No identifiable information about you or your household will be passed to other government departments, local authorities, members of the public or press.

We rely on people's voluntary co-operation, which is vital if the survey is to be successful. By taking part in this survey you will be assisting the Department of Transport and the Transport and Road Research laboratory which use the results. We are very grateful for your help.

Yours sincerely

Tom Tattan

Tom Tattan
Field Officer
National Travel Survey

For further information please contact me at the ONS address below or telephone me on 0171-533-5427 (Direct Line) or 0171-233-9233 Ext 5427 (ONS Switchboard). Work on this survey will start on or after the 1st of this month.

NTS-E8.96

1 Drummond Gate, London, SW1V 2QQ
Social Survey Division Enquiries (0171) 533 5500 Fax (0171) 533 5300

Letter B New short version letter for cognitive interviews

Social Survey Division

Our ref:

Date:

Dear Residents,

NATIONAL TRAVEL SURVEY

I am asking for your help on the National Travel Survey. This is being carried out by the Office for National Statistics (ONS), the official government research organisation.

The information on this survey is collected for the Department of Transport. The Department of Transport is responsible for:

- travel safety (including pedestrians)
- providing money for road building, railways, and other types of transport (buses, cycling)

We are interested in how much travelling people do, the cost of travel and how they get to work, to the shops or to school. By helping our interviewer, you will be assisting the Department of Transport and the Transport And Road Research Laboratory to understand and improve travel provision and safety.

You are representing thousands of other households, therefore your participation is extremely important. One of our interviewers will contact you and will be pleased to answer questions about the survey.

The information you give will be treated in strict confidence. Your answers will only be used for statistical purposes , so no one can identify you or your family.

The National Travel Survey asks about your own experiences. People who take part in this voluntary survey find it enjoyable and interesting. By taking part, you will be making a contribution which will benefit everyone.

If you want to know more about the survey, please call me on 0171-396-2283 (direct line).

Thank you very much for your help.

Yours sincerely,

Tom Tattan
Survey Manager, National Travel Survey

1 Drummond Gate, London, SW1V 2QQ
Social Survey Division Enquiries (0171) 533 5500 Fax (0171) 533 5300

Letter C New long version letter for cognitive interviews

Social Survey Division

Our ref:
Date:

Dear Residents,

NATIONAL TRAVEL SURVEY

I am asking for your help on the National Travel Survey. This is being carried out by the Office for National Statistics (ONS), the official government research organisation.

The information on this survey is collected for the Department of Transport. The Department of Transport is responsible for:

- travel safety (including pedestrians)
- providing money for road building, railways, and other types of transport (buses, cycling).

We are interested in how much travelling people do, the cost of travel and how they get to work, to the shops or to school. By helping our interviewer, you will be assisting the Department of Transport, the Transport And Road Research Laboratory, and other non-government organisations who use your results to understand and improve travel provision and safety.

People's travel varies according to their age, occupation, where they live etc. So we need to talk to all sorts of people therefore your participation is extremely important. One of our interviewers will call at your address during the next 8 weeks. The interviewer will:

- show you ONS's official identity card
- answer any questions you have about the survey
- call during the evening if you are not at home during the day

We never reveal the names and addresses of participants to anyone outside the research teams in ONS working on this survey. No survey results that record people's local authority area, and which might therefore stand some small chance of being identified with them or their households, are ever made available to local councils, the press, or members of the public. Such results are only available to the research staff of government departments and reputable research institutes who need to use statistics from the survey. Again I should emphasise that we pass on such results only for genuine research purposes and without names or addresses.

The National Travel Survey asks about your own experiences. People who take part find the survey enjoyable and interesting. Although the survey is voluntary, your help is very important to us. By taking part, you will be making a contribution which will benefit everyone.

Some of the results of previous National Travel Surveys are presented on the back of this letter. If you want to know more about the survey, please call me on 0171-396-2283 (direct line). Thank you very much for your help in this important survey.

Yours sincerely,

Tom Tattan
Survey Manager, National Travel Survey

1 Drummond Gate, London, SW1V 2QQ
Social Survey Division Enquiries (0171) 533 5500 Fax (0171) 533 5300

Reverse side of Letter C

Results from previous National Travel Surveys

- Nowadays people travel nearly four times as far, on average, as they did in 1950.
- On average people spend about as much on travel within Great Britain, as they do on housing.
- About one in five pensioners say that they have difficulty using buses.
- The average distance travelled by children between home and school is 1.9 miles.
- In 1994, 67% of car mileage was completed by cars using unleaded petrol or diesel fuel, compared with 53% in 1992.
- Sixty six per cent of workers in Great Britain usually travel to work by car, but only 18% of those working in Central London do so.

Letter D Short letter for focus group

Crown

Office for National Statistics
1 Drummond Gate
LONDON SW1 2QQ
Tel: [field office direct line no.]

Our ref: XXS

The Residents
XX Any Road
Any Town
Any County
Post code
extra line for address
extra line for address
extra line for address
extra line for address

Dear Sir/Madam,

National Travel Survey

I am writing to ask for your help with the National Travel Survey. By collecting information on people's own experiences, this voluntary study provides important facts and figures about travel in Britain. The findings are used widely - e.g. by government, charities, independent researchers, students in schools and universities - and are often commented upon in the press and on television.

We are the government office that carries out social research amongst the population as well as being responsible for organising the Census and the registration of births, deaths and marriages. Like the Census, the information you give will be treated in strict confidence. Your answers will be used only for statistical purposes, so no one can identify you or your household.

You are representing thousands of similar households so your participation is extremely important. By taking part you can help us to produce reliable statistics which everyone can use to understand what is going on in our society.

In the next few weeks one of our interviewers will write to tell you more about the study. In the meantime if you want to talk to somebody about the survey, please call me on 0171 533 XXXX (direct line).

Thank you in advance for your help.

Yours sincerely,

Field Officer's name
Survey Manager

Letter E Long letter for focus group

Crown

OFFICE FOR NATIONAL STATISTICS

Office for National Statistics
1 Drummond Gate
LONDON SW1 2QQ
Tel: [field office direct line no.]

Our ref: XXS

The Residents
XX Any Road
Any Town
Any County
Post code
extra line for address
extra line for address
extra line for address
extra line for address

Dear Sir/Madam,

National Travel Survey

I am an interviewer from the Office for National Statistics (ONS) - the government office that carries out social research amongst the population as well as being responsible for organising the Census and the registration of births, deaths and marriages. A few weeks ago, Tom Tattan from my office wrote to invite your household to take part in the National Travel Survey (NTS). Now I would like to tell you more about how you can help with this voluntary study.

The NTS provides important facts and figures about travel in Britain which are used, for example, by the Department of Transport (DoT), charities, independent researchers, and students in schools and universities. The results have been used to:

- plan cycle lanes and safety facilities for pedestrians
- identify the sorts of vehicles that cause the most pollution
- monitor the take up of concessionary fares (59% of elderly people currently have a bus pass)
- find out about why people used (or don't use) public transport.

I will be visiting you in the next few days to make an appointment to talk about your own travel experiences. Just in case you are not at home when I call, please show this letter to other people in your household so they know I will be coming.

Everything you tell me will be treated in strict confidence. Your answers will be used only for statistical purposes by researchers at ONS and the DoT who sponsor the study. No identifiable information about your household will be passed to another government department, local councils, the press or members of the public.

By taking part you can help us produce reliable statistics which everyone can use to understand what is going on in our society.

Thank you in advance for your help. I look forward to meeting you.

Yours sincerely,

[Interviewer to insert name and authorisation number here]

Letter F Final letter for field trials

Social Survey Division
Office for National Statistics
1 Drummond Gate
London SW1V 2

THE COTTAGE
1 SMITH DRIVE
JONESTOWN
KENT
BR33HR

January 1997

NTS 12345 678

Dear Resident,

You have been chosen to take part in the **National Travel Survey** - a survey about people's own experience with travelling in Great Britain. This research is being carried out by the **Office for National Statistics** for the Department of Transport.

Your address has been selected at random and one of our interviewers will contact you in the near future to explain the study in more detail. You may want to show this letter to other people in your household just in case the interviewer calls when you are not at home. If you are busy when the interviewer calls he/she will be happy to call again.

All our interviewers carry an official identification card which includes their photograph and the National Statistics logo as it appears at the top of this letter. Everything you tell us will be treated in confidence.

It is important to the success of this study that everyone chosen takes part. We rely on people's voluntary co-operation to produce official statistics to help everyone understand what is happening in our society. For further information please call 0171 533 5427 (direct line). Work on this survey starts on or after the 1st of *[month]*.

Thank you in advance for your help.

Yours sincerely,

Tom Tattan
Survey Manager

The Office for National Statistics is the government office that organises the Census and conducts surveys on behalf of Government Departments and other public bodies.

Fighting Nonresponse in Telephone Interviews; Successful Interviewer Tactics[1]

JOOP HOX, EDITH DE LEEUW AND GER SNIJKERS[2]

***Abstract**: In telephone interviews interviewers have far less time to persuade sampling units to cooperate than in a face-to-face interview. Furthermore, they also have less information to tailor their behaviour due to the limited channel capacity of the telephone. Nevertheless, experienced telephone interviewers have a 'tool box' of tactics, which they apply expertly. In this paper we describe the tactics to fight nonresponse as reported by experienced telephone interviewers at Statistics Netherlands.*

***Keywords**: survey participation, cooperation, telephone interview, persuasion, interviewer role, concept mapping*

1 Introduction

Telephone survey interviews have become more and more popular in the last thirty years. Main advantages of the telephone interview, as compared with the face-to-face interview, are less costs and stricter interviewer control (De Leeuw 1992). Like all data collection methods, telephone surveys also suffer from nonresponse, which forms a serious threat to the quality of the data.

To successfully fight nonresponse, knowledge about causes of survey (non)participation is needed. In their comprehensive theoretical review on survey participation Groves, Cialdini and Couper (1992) stress the importance of the interviewer-respondent interaction. During the initial moments of contact the interviewer is the initiator and dominant actor in this interaction, and much depends on the interviewer's ability to persuade the potential respondent. Empirical research shows that there is a considerable variation in response rates between telephone interviewers (Lyberg and Dean 1992; Collins et al. 1988; Oksenberg and Cannell 1988).

---

[1] The views expressed are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

[2] The authors gratefully acknowledge the assistance of the field department of Statistics Netherlands. They sincerely thank the interviewers and supervisors of the telephone unit for their enthusiastic help and the opportunity to learn from their experiences.

While nonresponse research in telephone surveys has focussed more on the technological and administrative aspects (e.g., optimal timing, the challenge of the answering machine), research regarding the face-to-face interview has centered on the human interaction and persuasion strategies. For instance, Morton-Williams (1993) analyzed tapes of 'door-step' interaction in relation to respondent cooperation. Interviewers who deviated from the prescribed 'script' were more successful in persuading reluctant respondents than interviewers who rigidly followed the introductory script. Morton-Williams emphasizes the importance of social skills to perceive and adapt to individual doorstep situations. In their theory of survey participation, Groves et al. (1992) emphasize the importance of 'tailoring' this is the use of different approaches - in words, behaviour and strategies - for different sampling persons. They also highlight the concept of 'maintaining interaction.' Maintaining interaction means that successful interviewers avoid a hard refusal by stepping back and keeping the opportunity open to contact the respondent again. There is some empirical evidence of the importance of these factors (cf. Campanelli et al. 1997; Groves and Couper 1994). Snijkers, Hox, and De Leeuw (1996) identified eight factors that experienced interviewers use to obtain cooperation in a face-to-face survey. Among these were, projecting a positive image by social skills, tailoring the introduction, and maintaining communication.

Telephone interviews and their introductions differ on important points from the face-to-face interview. First, face-to-face interviewers have more opportunities to collect the information about the respondent that is necessary for successful tailoring. They can use both the visual and auditive channels of communication (cf. De Leeuw 1992), both before the attempted interview (neighbourhood, type of housing) and during the introduction (appearance of respondent, body language, and verbal cues) (cf. Couper and Groves 1996). In telephone interviews only the auditive channel is available, limiting the interviewer to receive and transmit information using verbal and paralinguistic cues only. It all depends on what is being said and how it is said (e.g., the tone of voice). Second, telephone interviewers have far less time to convince a reluctant sampling unit. Typically, in face-to-face interviews initial interaction is completed within five minutes, while in telephone interviews the majority of the decisions to cooperate or refuse, are made within one minute (Groves 1992).

In sum, telephone interviewers have fewer clues, fewer means, and less time to tailor than face-to-face interviewers. Still, there is variation in response rate between telephone interviewers. Several studies have attempted to identify characteristics of successful interviewers. Oksenbergh and Cannell (1988) address the limited channel capacity of the telephone interview, focussing on interviewer vocal characteristics. They found that interviewers rated as speaking rapidly, clearly and loudly and perceived as sounding confident and successful, had lower refusal rates. Others focussed on the verbal respondent-interviewer interaction. For instance, Maynard, Schaeffer, and Cradock (1993) used conversational analysis on telephone introduction and discovered that refusals occur when

interviewers fail to address questions adequately. Houtkoop-Steenstra and Van den Bergh (1994) found in an experiment that a 'conversational' introduction in which interviewers were allowed to use their own words, produced fewer refusals than scripted standard introductions. Both studies suggest that adequate tailoring may work in telephone introductions. This was confirmed by Couper and Groves (1996), who showed that tailoring increased the likelihood of cooperation in telephone surveys. They also present evidence that after a negative statement of the respondent tailoring may increase the likelihood of cooperation.

Pondman (1998), following Smit and Dijkstra (1991), takes this one step further and concentrates on what elements make for successful tailoring. She identifies four rules: (1) avoid asking 'why', (2) avoid repeating the refusal, (3) offer to call back when refusal states lack of time, and (4) react to other refusals by giving positive, relevant information about the interview.

In this study we broaden the approach and portray the tactics used by experienced telephone interviewers. Our main goal was to identify successful strategies that would be trainable to new, inexperienced interviewers. We used a highly structured interviewer debriefing study to draw upon the knowledge and wealth of experience that interviewers have (cf. Campanelli, Martin and Rothgeb 1991). In addition, we wanted to contrast these strategies with strategies described by successful face-to-face interviewers (Snijkers et al. 1996), thereby providing more insight in the special nature of telephone introductions.

In the next section we first give a short description of the group of experienced telephone interviewers who acted as informants and we outline the procedures used in concept mapping. We continue with the major results and end with a summary in which the similarities and dissimilarities of successful strategies for telephone and face-to-face interviews are discussed.

2 Method

2.1 Group studied

During the months March-May 1996 a field experiment was carried out at Statistics Netherlands using mixed-mode computer assisted data collection. This experiment was part of a larger implementation study for the redesign of the continuous survey on living condition (POLS).

Thirteen very experienced CATI-interviewers were selected for this task. Selection criteria were among others, good social skills, research minded, a generally high response rate and good interviewer performance as evaluated by their supervisors (cf. De Leeuw et al. 1996).

The interviewers were specially trained for the POLS study, but no special training in nonresponse reduction and persuasion of respondents was given. After completing the field experiment, the interviewers took part in a special evaluation and debriefing study.

2.2 Procedure

Part of the debriefing study was a focus group on successful tactics to obtain cooperation in a telephone survey. The knowledge of interviewers and the information on what defines successful strategies is often rather diffuse and unstructured. Therefore, to obtain structured and usable information we used the technique of 'concept mapping'. Concept mapping is a qualitative, but highly structured method to extract information from a group. A comprehensive system for concept mapping has been developed by Trochim (1989). The major advantage of this method is that it quickly leads from fuzzy knowledge to an interpretable conceptual framework, in our case on interviewer tactics to persuade the potential respondent. Furthermore, this framework can be expressed in a graphical representation, which shows all major ideas and their interrelationships. For an introduction on concept mapping see Trochim (1989).

Concept mapping in focus groups consists of five steps: (1) preparation and *developing the focus*, (2) *statement generation* by the group, (3) *statement structuring* and *rating* by the group, (4) statistical analysis and *statement representation* as a cluster tree and concept map, and (5) *interpretation* of the results by the group.

Step 1 or the preparation phase should result in two separate products: the primary focus or domain of interest for the brainstorming session with the focus group, and the rating scale needed for the structuring of statements in step 3. We decided on the following focus for the brainstorming session: "What is effective to obtain cooperation in a telephone survey: What can YOU do as an interviewer, Which tactics work, What can we as Bureau do to help you". The rating focus concerned the effectiveness of the tactics, and was stated as follows: "For each tactic mentioned, give a rating of its effectiveness. Use the following response categories:

- 1 This tactic could backfire
0 This tactic probably has no effect
+ 1 This tactic works a little
+ 2 This tactic works well
+ 3 This tactic works very well
+ 4 This tactic works almost always"

Step 2 or statement generation. During a one hour brainstorming session statements were generated with the members of the focus group. The focus statement described above constituted the prompt for the brainstorming. An informal introduction stated "you are all very experienced interviewers, what would you advice a novice to reach high response rates, what is the golden tip". This was added to compensate for the rather abstract formulation of the focus and to emphasize the practical applicability of tactics and strategies we wanted to elicit.

The usual rules for brainstorming applied, such as, encourage lots of statements, and emphasize the importance of no criticism or discussion during the generation of statements. The statements were recorded on a whiteboard by the moderator. The wording was checked with the group members, and if necessary the text was adjusted. The final text was entered into a laptop computer by one of the team members, who was seated behind a one-way mirror. Participants knew that this was happening and the moderator openly addressed the 'recorder' at certain moments to make sure that the statements were entered correctly.

In step 3, structuring, the individual participants were instructed to sort cards on which the statements were printed into different piles ' the way it makes sense to you'. Restrictions were: each statement can only be placed in one pile, all statements may not be put in one large single pile, and all statements may not be put into a pile of one, although a small number of piles of one statement are allowed. After this card sort the individual participants were asked to rate the statements as to effectiveness, using the six-point rating scale described above.

Step 4 is the analysis or 'statement representation' phase. The individual sorts were combined into a group similarity matrix. This similarity matrix is the input for a multidimensional scaling procedure and cluster analysis. The two-dimensional plot of points created by the MDS may be viewed as a representation of the 'emerging concepts' of group knowledge, hence the name concept mapping. The cluster solution is superimposed on the map of points to facilitate interpretation by the group members. Furthermore, the mean group ratings for each statement are computed. It is possible to overlay the ratings onto the concept map.

Step 5 is again a group activity. The participants discussed possible meanings and acceptable names for each cluster of statements. This last step attempts to identify relations between tactics in the form of a group-approved map.

3 Results

3.1 Generated statements

The brainstorming resulted in 37 different statements. Each statement was thought to be an effective tactic by at least one group member. Each statement was individually rated on effectivity to gain cooperation. Table 1 lists the statements in order of perceived effectiveness.

Table 1: Most effective interviewer tactics

Given is average group rating (scale:-1,0,1,2,3,4), text of generated statement and number of order in which statements were generated.

3.23	use practical arguments why survey is important (23)
3.15	quiet work environment (better attention, faster reaction) (34)
3.15	if hesitant because of privacy indicate such questions may be skipped (37)
3.08	voice friendly and with much intonation (16)
3.08	good ergonomic work environment to stay attentive and enthusiastic (35)
3.00	be convincing (19)
3.00	more good background information about CBS[1] (14)
3.00	know the topic of the study well (13)
3.00	tell how much time is needed for interview (5)
2.92	advance letter should give more and better information (31)
2.85	use simple concepts in introduction that are close to the language of the respondent (15)
2.85	CBS should inform the public what they do with the data (30)
2.69	react to respondent (11)
2.69	project enthusiasm (18)
2.69	show understanding (2)
2.62	assure resp. that it is no trouble to call back later (3)
2.62	use information from background based on initial reaction respondent and experience (22)
2.54	react to the type of person that is on the phone (10)
2.54	react/use what respondents say (1)

[1] CBS is the abbreviation of Central Bureau voor de Statistiek (Statistics Netherlands)

2.54 in a short interview, mentioning the duration, prevents the need to make appointments (they answer directly) (6)
2.46 figure out who (what kind of person) you are talking to (e.g., elderly) (9)
2.46 it ONLY takes five minutes (way you stress the short duration) (7)
2.33 have toll-free telephone number available for information (32)
2.33 specific questions stimulate break off (e.g., date of birth, age is easier, less official/threatening) (36)
2.31 reassure, remove concerns about government and misuse of information (big brother) (28)
2.23 the words you use in the introduction are especially important (choosing the right words, e.g, 'ONLY a few questions') (8)
2.23 make clear that you too believe in the study (21)
2.23 CBS should be better known by public (29)
2.23 if they did not receive advance letter, start with giving them information yourself, reassure that letter was only a short announcement (33)
2.15 reassure you are NOT selling, remove concern about commercial interest (27)
2.00 give everyone the feeling they are the first and very special (17)
2.00 offer opportunity for 'time-out' (step back) (4)

1.85 keep the conversation going with open questions (12)
1.85 communicate from person to person (25)
1.85 'may I start asking the questions' works better than 'do you want to answer the questions' (24)
1.38 if respondent is reluctant draw them out with specific remarks (that is exactly what we are interested in, would be a pity if person like you..) (26)
1.18 response depends also on interviewers mood (20)

When we look at Table 1, we should remember that interviewers were asked to mention successful tactics to gain cooperation. Every statement is therefore successful in the opinion of at least one experienced interviewer. This does not mean that everybody completely agrees on every statement, there is some variance among the interviewers. When we look at the total range of the effectiveness ratings and the standard deviation, we notice that experienced interviewers strongly disagree on certain statements. Prime examples are statement number 20 ('influence own mood'), and number 36 ('question stimulates break-off').

On the other hand, there were tactics that every interviewer rated as either works well or works very well. Examples are statement number 5 (inform about time needed), number 13 (know the topic of the study well), 16 (voice), and 34 (work environment/concentration).

3.2 Interrelationship of statements

Analyses based on the similarity matrix of sortings resulted in ten clusters. These clusters were discussed and named by the group. Table 2 lists the statements grouped by named cluster; for each cluster the average cluster rating on effectiveness is given in parentheses. The clusters are ordered in descending effectiveness.

Table 2: Statements grouped by cluster

Cluster names and average effectiveness ratings based on interviewers opinion. Most effective clusters are named first.

Cluster 9: *Work environment (3.12)*

34 quiet work environment (thus better attention, faster reaction)
35 good ergonomic work environment to stay attentive and enthusiastic

Cluster 6: *Persuasion (2.95)*

23 use practical arguments why survey is important
13 know the topic of the study well
22 use arguments from background information based on initial reaction

Cluster 5: *Audibly positive (2.75)*

19 be convincing
16 voice friendly and with intonation
18 project enthusiasm
21 make clear that you too believe in the study

Cluster 4: *Time (2.67)*

6 in a short interview mentioning the duration prevents need for appointments
5 tell how much time is needed for interview
7 stress that it takes ONLY a few minutes

Cluster 7: *Public relations (2.67)*

14 more and better background information from CBS
31 advance letter should give more information
30 CBS should inform public what they do with the data
32 have toll-free telephone number available for information
29 CBS should be better known by public

Cluster 3: *adapt language (2.59)*

15 use simple concepts in introduction
2 show understanding
8 using the right words is especially important in the introduction

Cluster 8: *Legitimacy (2.54)*

37 if hesitant because of privacy indicate such questions may be skipped
28 remove concerns about government or misuse of information (big brother)
27 remove concerns about commercial interests (no selling)

Cluster 10: *(2.33)*

36 specific questions stimulate break-off

Cluster 1: *Keep interaction going (2.11)*

11 react to respondents
1 react to what respondents say
10 react to the type of person that is on the telephone
9 figure out what kind of person you are talking to
33 if they did not receive the advance letter, start with giving general information
12 keep the conversation going with open questions
26 is respondent is reluctant draw them out with specific remarks
20 the response depends also on the interviewers mood

Cluster 2: *Establish a relationship (2.06)*

3. assure respondents that it is no trouble to call back later
4. offer opportunity for 'time-out' (step back)
17. give everyone the feeling they are the first and very special
24. 'may I start asking the questions' works better than 'do you want to answer the questions'
25. communicate from person to person (avoid feeling of institute/agency calling respondent, give feeling of person (interviewer) communicating with respondent

If we concentrate on the most effective clusters, we see that besides a good work-environment, the use of adequate arguments in reaction to the respondent and the emphasis on using voice characteristics are seen as the most effective strategies. Also, when applicable, the short time needed to complete the interview was emphasized. The interviewers are well aware of the advantages and disadvantages of the telephone as 'medium' and try to use this medium as effectively as possible. They recognize the need to concentrate on what the respondent says and react to it. The emphasis on an ergonomic work environment probably also reflects the 'sub-optimal' telephone facilities at the time of our data collection. There are now new and pleasant facilities for telephone interviewing. Nevertheless, ergonomics

remains very important, a happy interviewer is more motivated and projects a positive image. Also, in order to 'tailor' the introduction, an interviewer should be able to concentrate on the respondent. Ergonomic adaptation to the work-floor, such as noise absorbing isolation, good chairs, etc, will help interviewers in achieving a good performance.

4 Summary and discussion

We identified ten clusters of response improving techniques. Some clusters (i.e., cluster 4, 5, 8, and 9) were typical for telephone interviews. Cluster 4 (time) exploits the advantage of telephone interviews; arguments emphasizing the short duration are used to persuade reluctant sampling units. Clusters 5, 8, and 9 all center round the limited channel capacity of telephone interviews.

Cluster 5 (audibly positive) stresses the extra effort telephone interviewers have to make to compensate for the absence of the visual channel of communication. They cannot use smiles or gestures, they have to **sound** enthusiastic and convincing. What is interesting to note is that face-to-face interviewers when asked by Snijkers et al (1996) named a cluster of strategies that was also focussing on the projection of a positive image, with statements such as be likeable, friendly, project enthusiasm, etc. This also relates to the 'social skill' mentioned by Morton-Williams (1993) as necessary for successful interviewers. The same concepts are used by both face-to-face and telephone interviewers in their introduction. They only differ in the way they implement the resulting strategies, with telephone interviewers of necessity strongly focussing on the voice as sole means of communication. Although a completely different research method was used, these results partly replicate and underscore the importance of the findings of Oksenberg and Cannell (1988), who pioneered research into the effect of voice characteristics on nonresponse.

Cluster 8 (legitimacy) centers around the special problems telephone interviewers have to establish that they are conducting a legitimate survey, that they are not selling anything, and that individual information will not be misused. In face-to-face surveys interviewers have more means to establish legitimacy and do use those. They show or hand-over their official ID, have copies of introductory letters or leaflets to show and can project by nonverbal means a non-threatening and reassuring image (cf. Morton-Williams 1993, chap 7).

Connected with the need to establish legitimacy is the emphasis on good public relations of the agency. Although a good P.R. was also mentioned by the face-to-face interviewers investigated by Snijkers et al, the telephone interviewers in this study not only rate P.R. in general as more effective and helpful, they also name more P.R.-related strategies. A main concern of the telephone interviewers was a well-known and positive image of the agency, in

combination with informative (advance) letters and a widely advertised toll-free telephone number for information.

Cluster 9 (work environment) is typical for the telephone situation and highlights the importance of an ergonomical work environment with a low noise level. In order to adequately react to respondents and tailor their arguments, interviewers have to concentrate on slight changes in the tone of voice of respondents, and pick up general para-linguistic signals. In cluttered and noisy surroundings a good auditive communication of interviewer and respondent is jeopardized, and unnecessary break-offs or refusals may result.

In face-to-face interviews the importance of 'tailoring' and 'maintaining interaction' are emphasized (Morton-Williams 1993; Groves et al. 1992). Snijkers et al. (1996) replicated their findings for Dutch interviewers, using a different research method. The interviewers investigated were not specially trained in doorstep techniques. The same is true for the experienced telephone interviewers in this study. However, also these telephone interviewers named strategies for tailoring and maintaining interaction in the clusters 3 (adapt language), cluster 6 (persuasion), and cluster 1 (keep interaction going). The main difference between the telephone and the face-to-face interviewers is that the telephone interviewers not explicitly mentioned the strategy to offer to call (come) back, when time-problems were mentioned by the respondents. Afterwards, we explicitly asked our telephone interviewers why they had not referred to this strategy. Their main reaction was surprise and they told us that offering to call-back and making appointments are basic strategies. It is mentioned as one of the first things in their training, and every interviewer knows this. During this debriefing session the interviewers again stated that mentioning the short duration (cluster 4) often prevents the need to make appointments. They stressed that this only works with really short interviews and that they often start with offering a call back, but mention in the same breath that it will only take.

This study replicates the effectiveness of important theoretical concepts about nonresponse reduction: tailoring and maintaining interaction (cf. Campanelli et al. 1997; Groves et al. 1992; Groves and Couper 1994; Morton-Williams 1993; Snijkers et al. 1996). This is now replicated across countries (USA, UK and Holland), across research methods (interaction coding, focus groups, interviewer questionnaires and concept mapping), and across interview modes (face-to-face and telephone), which gives rise to great trust in the utility of 'tailoring'. Professional competence, tailoring of introduction, and maintaining the interaction are key concepts for a successful doorstep approach in surveys AND for a telephone survey introduction.

5 References

Campanelli, P., Sturgis, P. and Purdon, S. (1997). Can you hear me knocking: An investigation into the impact of interviewers on survey response rates. London: SCPR

Campanelli, P.C., Martin, E.A. and Rothgeb, J.M. (1991). The use of respondent and interviewer debriefing studies as a way to study response error in survey data. The Statistician, 40, pp. 253-264

Collins, M., Sykes, W., Wilson, P. and Blackshaw, N. (1988). Nonresponse, the UK experience. In: R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II and J. Waksberg: Telephone Survey Methodology. New York: Wiley

Couper, M.P. and Groves, R.M. (1996). Introductory interactions in telephone surveys and nonresponse. Paper presented at the Annual Meeting of the American association for Public Opinion Research, Salt Lake City, USA

Groves, R.M. (1992). Invited lecture on nonresponse for Statistics Netherlands. Voorburg, The Netherlands

Groves, R.M. and Couper, M.P. (1994). Householders and interviewers: The anatomy of pre-interview interactions. SMP-working paper 011, Ann Arbor: University of Michigan

Groves, R.M., Cialdini, R.B. and Couper, M.P. (1992). Understanding the decision to participate in a survey. Public Opinion Quarterly, 56, pp. 475-495

De Leeuw, E.D. (1992). Data quality in mail, telephone, and face-to-face surveys. Amsterdam: TT-publikaties

De Leeuw, E.D. et al. (1996). Procedures voor werving en selectie van meta-interviewers en meta-consulenten [Procedures for recruiting and selecting special interviewers]. Heerlen: CBS, In Dutch

Forsyth, B.H. and Lessler, J.T. (1991). Cognitive laboratory methods: A taxonomy. In: P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman (Eds.): Measurement Errors in Surveys. New York: Wiley

Lyberg, I. and Lyberg, L. (1991). Nonresponse research at Statistics Sweden. Invited paper presented at the Annual Meeting of the American Statistical Association, Atlanta, USA

Lyberg, I. and Dean, P. (1992). Methods for reducing nonresponse rates: A review. Paper presented at the Annual Meeting of the American Association for Public Opinion Research. St. Petersburg, Fl., USA

Maynard, D.W., Schaeffer, N.C. and Cradock, R.M. (1993). Declinations of the request to participate in the survey interview. Washington DC: US Bureau of the Census

Morton-Williams, J. (1993). Interviewer approaches, Aldershot: Darthmouth Pub.

Oksenberg, L. and Cannell, Ch. (1988). Effects of interviewer vocal characteristics on nonresponse. In: R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg: Telephone Survey Methodology. New York: Wiley

Pondman, L.M. (1998). The influence of the interviewer on the refusal rate in telephone surveys. Ph.D.-thesis. Vrije Universiteit, Amsterdam

Smith, J.H. and Dijkstra, W. (1991). Persuasion strategies for reducing refusal rates in telephone surveys. BMS, 33, pp. 3-19

Snijkers, G., Hox, J. and De Leeuw, E.D. (1996). Interviewers' tactics for fighting survey nonresponse. Paper presented at the 1996 International Conference on Social Science Methodology, Colchester, UK

Trochim, W.K. (1989). An introduction to concept mapping for planning and evaluation, Evaluation and Program Planning, 12, pp. 1-16

Trochim, W.K. (1993). The concept system (program manual). Ithaca: Concept systems software

The Effect of Interviewer Persuasion Strategies on Refusal Rates in Household Surveys

PATRICK STURGIS AND PAMELA CAMPANELLI

Abstract: *Our analysis focuses on the extent to which there is scope for reducing rates of refusal on large-scale household surveys below current standard levels. Our data consisted of over 300 tape-recorded doorstep interactions, drawn from substantive surveys at two different organisations. Tape-recorded interactions were classified in terms of the degree of reluctance expressed by the respondent and the consequent scope for the interviewer to deploy interpersonal skills and persuasion. Interactions where there was both reluctance and scope for persuasion were classified according to whether they resulted in cooperation or refusal and the interviewer tactics associated with the two types of outcome were compared. Conclusions are drawn about the prevalence of different types of interaction and the scope for reducing refusal rates through training interviewers to use techniques likely to minimise refusals on the doorstep.*

Keywords: *reluctance; tailoring; response rates; interviewers; persuasion.*

1 Introduction

It has long been recognised that the job of the survey interviewer is the vital nexus between the survey organisation and address residents and much has been written about survey interviewing (see, for example, Hyman 1954, 1975, Kahn and Cannell 1957, Sudman and Bradburn 1974, Dijkstra and Van der Zouwen 1982, Fowler and Mangione 1990). What has received less attention is the impact of the interviewer on survey nonresponse. A few studies have looked at the physical characteristics and attributes of interviewers and have suggested that under certain circumstances factors such as interviewer gender, manner of dress, and vocal characteristics can make a difference to how the respondent views the interviewer (see, for example, Oksenberg et al. 1986, Fowler and Mangione 1990, Morton-Williams 1993). Studies have also investigated the role of a number of different interviewer personality and attitude variables on both individual response rates and the interview data obtained. Singer and Kohnke-Aguirre (1979), for example, have shown that interviewer beliefs about item sensitivity can

significantly predict the likelihood of their obtaining or failing to obtain responses on those items. Perhaps the most popular approach of study, however, has been the consideration of interviewer behaviour during the initial request for survey participation. For example, early research experimented with varying the content of what the interviewer said 'on the doorstep' (Dillman et al. 1976, O'Neil et al. 1980, Couper and Groves 1991) and later research explored the applicability to survey research of the principles of the psychology of compliance, helping behaviour, and opinion change (Groves et al. 1992). Empirical data clearly suggest that interviewer response rates correlate positively with years in the job (Durbin and Stuart 1951, Colombo 1983, Lievesley 1986, Couper and Groves 1991). Although this finding is confounded with interviewers' self-selection to remain as interviewers, one inference that can be drawn is that experienced interviewers' success derives from their 'larger number of combinations of behaviours proven to be effective for one or more types of householders' (Groves et al. 1992, p 478-9).

This paper describes analyses conducted on a sample of tape-recorded doorstep interactions from two large face-to-face interview surveys. Transcribed interactions are coded to a 6 category coding frame in order to explore interviewer persuasion strategies and to assess what scope there is for reducing the refusal component of nonresponse through improved interviewer doorstep technique.

2 Methods

Our doorstep experiment involved 32 face-to-face interviewers from two different UK organisations: Social and Community Planning Research (SCPR) and the NOP Research Group. The interviewers were selected to allow for geographic spread (but excluding Scotland so as to minimise travel and hotel costs) and to allow for a range of experience levels. The interviewers in the two organisations were working on two rather different types of surveys. The NOP interviewers were working on the Political Tracking survey which is a face-to-face paper and pencil (PAPI) survey. In the Political Tracking survey, one person per household was interviewed as pre-selected from a probability sample of the electoral register. Thus, the NOP interviewers were after a pre-named individual. The SCPR interviewers were working on the Family Resources Survey (FRS), an extremely detailed computer assisted personal interview (CAPI) financial survey in which all adult members of the household participate.

Small portable tape-recorders were used for the taping. Interviewers were instructed to approach the household with the tape-recorder switched on and positioned on their clip board. In carrying it in this way, they neither concealed the tape-recorder nor drew the respondent's attention to it.

2.1 Data obtained

The average response rate for the selected interviewer areas was 61[1] percent for the Political Tracking survey and 72 percent for the FRS. For the Political Tracking survey 256 households were to have doorstep introductions recorded and at least one useable tape was received for 207 of these. The respective figures for the FRS were 192 and 146.

For households in the tape-record condition, interviewers were instructed to record all calls on the household until an interview was achieved. This resulted in 401 individual taped calls for the Political Tracking survey and 447 individual calls for the FRS. All tapes were then transcribed verbatim.

3 The coding frame for interaction types

The primary (but by no means only) measure of the efficacy of an interviewer's doorstep approach at a sample unit is whether it results in an interview or a refusal at the address. Non-contacts are excluded as they afford no chance for persuasion. Deadwood is also excluded. This, then constitutes the first basis on which transcripts were coded - *interview/refusal.*

Within each of these categories (interview/refusal) it was possible for there to have been an opportunity for the interviewer to use a persuasion strategy or not. So *'opportunity yes/no'* became the first sub-division within the superordinate category of interview/refusal. In defining what constitutes an 'opportunity' to persuade we decided (as we attempted to throughout the analysis) to adopt a liberal definition, erring on the side of false positives rather than false negatives. Thus anything that the address resident said during any of the interactions, across calls that could be interpreted as expressing some form of reluctance[2] to participate was coded as presenting an opportunity for the interviewer to persuade. Thus, even interactions in which address residents merely asked how long the interview would take before agreeing or said that they were too busy at present but were prepared to participate, were coded as presenting the interviewer with an 'opportunity' to persuade.

---

[1] The overall response rate was 67 per cent for Political Tracking and 69 per cent for the joint ONS/SCPR FRS.

[2] In this context, 'expressing some form of reluctance' does not include cases in which the reluctance expressed by the address resident is so strong that it does not provide the interviewer with any opportunity to persuade. For example an address resident slamming the door in an interviewer's face before the interviewer has completed their initial introduction, although a clear expression of reluctance, would not be coded as presenting an opportunity for the interviewer to persuade but would rather be coded as a refusal which presented no opportunity for the interviewer to persuade.

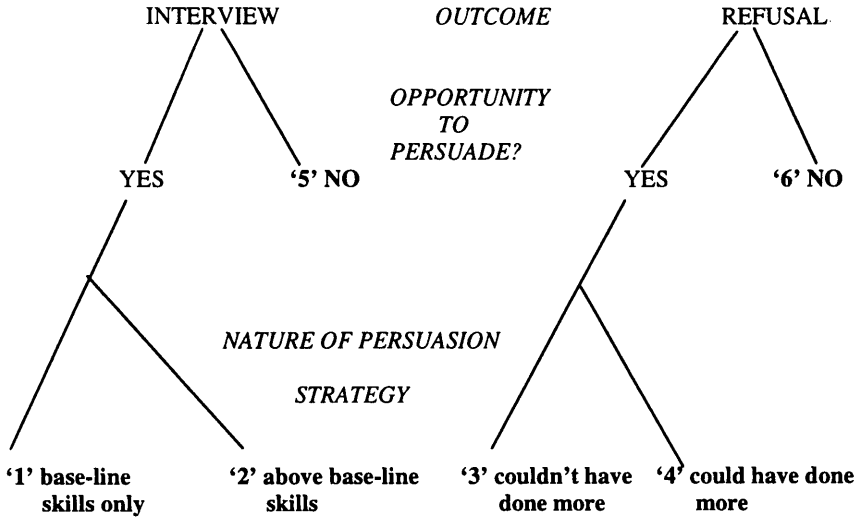
Within the sub-category of 'opportunity yes/no' it was also possible to subdivide cases on the basis of how the interviewer responded to the opportunity. So *'nature of persuasion'* became the sub-division within 'opportunity yes/no'. Four different codes were used to describe the nature of the persuasion. Which of these four codes applied was firstly dependent on whether the outcome at the address was an interview or a refusal.

Where the outcome was an interview: the application of the code depended on the level of persuasion skill that the interviewer had demonstrated in the interaction. At this point we also developed the idea of a 'base-line' skill level for interviewers. The basis of this idea is that there are certain fundamental elements of an interviewer's job which even very new and inexperienced interviewers would be expected to be able to execute. These include: being able to answer basic questions about the length of the interview; the topic of the survey; the selection procedure; arranging appointments etc. As we are primarily interested in studying the potential reduction in refusal rates *below current levels through improved interviewer persuasion technique*, it is not really in such 'base-line' skills that our interest lies. Rather, it is in assessing the extent of and examining the features that characterise interactions in which the skills demonstrated by the interviewer fall either above or below 'base-line' levels.

Therefore, if the interviewer had demonstrated only 'base-line' persuasion skills in obtaining an interview, the transcript was coded '1'. If, however, the persuasion skills contained elements over and above base-line levels and an interview was achieved, the transcript was coded '2'. Where a code '1' is applied as opposed to a '2' it should not necessarily be taken to mean that there was a short-coming in the interviewer's approach but rather that 'base-line' skills were all that were required in order to 'persuade' the address resident to participate on that occasion.

Where the outcome of the call was a refusal: Rather than simply denoting whether base-line skills or above had been used by the interviewer during the interaction, in the case of a refusal the codes refer to whether the interviewer *could have done more* to persuade the address resident to participate. When the outcome at an address is a refusal then the code '3' means that the interviewer said and did all that would be expected of them in terms of standard interviewer training guidelines in attempting to avoid the refusal. The code '4' on the other hand denotes cases in which, again from the perspective of standard interviewer training guidelines, alternative/better persuasion strategies from the interviewer could have been used. Or, in other words, the '4' code applies to interactions in which the interviewer dropped below baseline skill levels. The basic structure of the coding frame is represented in Figure 1 below.

Figure 1: Diagrammatic representation of coding frame for interaction types



This yields *six* distinct categories into which interactions can be coded (indicated in bold in Figure 1). Incorporating an 'other' category for transcripts where there is insufficient information for the transcript to be coded gives a final total of seven. [3]

3.1 Descriptions of interaction types

Each of the interaction types shown in Figure 1 is described in more detail below.

Interaction type 1 (Interview - baseline skills only): These interactions are characterised by address residents who require only a small degree of persuasion before agreeing to be interviewed. Typically such interactions consist of an interviewer's initial introduction followed by the address resident either:

- expressing a wish for time delay - they are too busy to give an interview at the present moment but suggest, or at least do not discount the possibility, that an interview will be given at some future point in time.

[3] Transcripts were coded 'other' when all or significant sections of interviewer-address resident interactions were not recorded - providing insufficient information to determine the correct code.

- asking for further information about the nature of the request being made of them (interview length; survey topic; sponsoring organisation; confidentiality).

Type 1 interactions conclude in an interview after the interviewer either arranges an appointment for another occasion (and gets the interview then) or answers the address resident's request for further information etc. and gets the interview immediately.

Interaction type 2 (Interview - above baseline skills): These interactions represent situations in which the address resident demonstrates at least a fairly high degree of reluctance but nevertheless agrees to participate because the interviewer responds effectively to their concerns over participation. Had the interviewer been unable to respond so effectively, the expressed reluctance of the address resident may have led to a refusal.

Interaction type 3 (Refusal - couldn't have done more): Interactions of this type are characterised by address residents who express reluctance to participate and despite relevant and competent attempts by the interviewer to persuade them, nevertheless refuse to be interviewed. Expressions of reluctance in this category are more often characterised by a general reluctance to participate (too busy in general; not interested in general; not interested in topic; often multiple combinations of aforesaid) than by time-delay statements and requests for more information. Address residents in this category are similar in terms of reluctance to those in category 6 (see below) but are perhaps more concerned to be polite in refusal without ever having a significantly greater intention of participating.

Interaction type 4 (Refusal - could have done more): Type 4 interactions are situations in which the address resident demonstrates a similar level of reluctance as described in type 2 interactions (above) but with the difference that the interviewer fails to adequately allay the address resident's concerns and consequently loses the interview. It is likely that this type of interaction represents a slightly higher degree of address resident reluctance than is found in type 3 interactions, as although it is possible to identify shortcomings in the interviewer's persuasion strategies within type 4 interactions, we cannot automatically assume that the address resident would have given an interview had the shortcomings not been present.

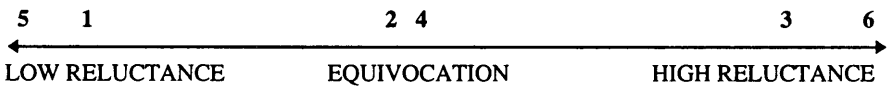
Interaction type 5 (Interview - no opportunity to persuade): These interactions are characterised by address residents who do not really require any degree of persuasion at all. They express no reluctance to participate, not even requiring information about interview length or topic etc. but simply agree to be interviewed more or less immediately after the initial introduction - sometimes even making a positive comment about the survey request. Often the interviewer provides the respondent with additional information

about what they are required to do *after* they have agreed to participate but before the interview has begun. They constitute what could be termed the 'hard-core' of respondents.

Interaction type 6 (Refusal - no opportunity to persuade): These interactions are characterised by address residents who are very brusque, sometimes rude, and always adamant that they do not want to even discuss the possibility of giving an interview. This type of interaction is perhaps what constitutes the real 'hard-core' of nonrespondents. Thankfully, this type of interaction is comparatively rare.

In summary then, beyond the final outcome (interview/refusal), the interaction type at a particular address is determined by two inter-related factors, the degree of reluctance of the address resident and the nature of the persuasion done by the interviewer. Figure 2, below shows the hypothesises positions of the six types of interaction on a dimension of 'address resident reluctance'.

Figure 2: Interaction types on dimension of address resident reluctance



The interaction type associated with the least reluctance on the part of the address resident is number 5 (interview - no opportunity to persuade). Next least reluctance is found in interactions of type 1 in which the interviewer need only employ 'baseline' skills in order to achieve the interview. At the opposite extreme of the reluctance dimension are interactions of type 6 - where address residents refuse without even presenting the interviewer with an opportunity to persuade them otherwise. Also at the high reluctance pole, are interactions of type 3. These interactions are characterised by respondents who are perhaps about as reluctant as those in type 6 interactions but are somewhat more willing to state their reasons for refusing.

Now, when assessing the scope for refusal reduction through improved interviewer doorstep technique, none of the interaction types described so far offer much hope. The two types at the low reluctance pole of the dimension, by definition, offer no scope for refusal reduction as they constitute only households which required at most baseline skill levels to achieve an interview. Likewise, the two types at the high reluctance pole of the dimension, offer little scope as they are defined as situations in which there appears to be no objective shortcomings in the interviewer's doorstep technique. Thus the real scope for

refusal reduction is located at the middle of the dimension, where the eventual outcome at an equivocating household may be determined by the way in which the interviewer responds to any reluctance exhibited by the address resident. The extent to which refusals may be reduced on a particular survey will be determined by the extent to which interviewer technique is able to 'shift' households from being type 4 interactions (refusal - could have done more) to being type 2 interactions (interview - above baseline skills).

3.2 The data used for classifying interaction types

There were 207 addresses at which at least one call was tape-recorded for the Political Tracking data. Of these, 33 had too much of the interaction missing (not recorded) to be reliably coded (coded 'other')[4] and four tapes were unfortunately lost during the transcription process. This resulted in 170 which were transcribed and coded. Of the146 tape-recorded addresses for the FRS, 15 contained insufficient data to be coded reliably (coded 'other'), and 131 were transcribed and coded.

The transcripts were all coded by one researcher. When the coding was complete, a second researcher was given a random selection of transcripts from each category. In all, fourteen transcripts were coded by the second researcher, covering a minimum of two from each category. On ten of the fourteen transcripts (71 percent), the same code was applied by both researchers. Of the remaining four discrepant transcript codes, all were resolved after discussion (three in favour of the first researcher's original coding and one in favour of the second researcher's original coding).

4 Results of the interaction type classification

The results are summarised in Figure 3. The most notable features of Figure 3 are that, in both surveys, the highest percentages are found in category 5 (where the respondent has simply invited the interviewer in without requiring any persuasion) and category 1 (where the interviewer needed only baseline skills to achieve the interview). The lowest proportions were found in category 6, where the respondent refuses and gives the interviewer no opportunity to persuade them. This is clearly very encouraging from the perspective of survey practitioners as it indicates both that the majority of people contacted are generally receptive to being interviewed and that the amount of people who are totally adamant that they will not co-operate is rather insignificant.

4 Some of these may actually be non-contacts. The unclear cases typically show a series of call attempts and have contact with another household member, but it is unclear whether we have the complete set of calls. In addition, there is no detailed outcome code available on the Political Tracking dataset to verify against.

Figure 3: Proportions of interactions within each code across surveys

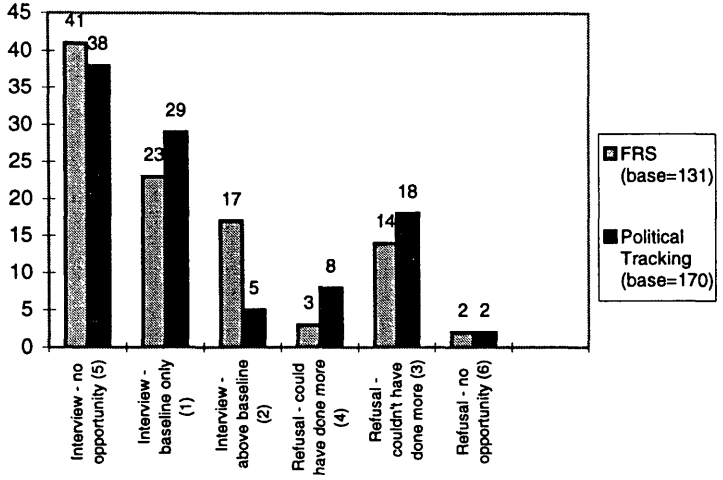


Figure 3 also shows that the distribution across the two surveys is fairly similar, the main differences being apparent across categories 1 - 4. Categories 2 and 4 are of particular interest as these are the ones where the skill of the interviewer can be the crucial factor in determining the final outcome at a particular address (this is discussed further below). It can be seen that the FRS data contains a higher proportion of interviews achieved by 'above baseline' skills (category 2) and a lower proportion of interactions where they could have done more to avoid a refusal (category 4). This is reflected in the overall difference in *co-operation rates* for the two surveys (successful interviews / (successful interviews + refusals)). Based on the figures in this analysis, the rates are 81 percent for the FRS and 72 percent for Political Tracking. These *co-operation rates* differ from the response rates of 72 and 61 percent respective, because non-contacts have been excluded, but also because cases in the 'other' category were excluded. For the FRS, 53 percent of these coded in 'other' were productive, while the corresponding figure for Political Tracking was 33 percent. The difference between these two percentages, however, is probably not as big as it may seem because the 'other' cases in the Political Tracking data may contain some non-contacts. Although 53 and 33 percent are distinctly lower than the co-operation rates for the coded cases, the percentages of 'other' cases are relatively small. There is also the broader issue of those cases which should have been taped but were not. These were slightly less likely than the taped cases to have been productive (for

example, 43 percent of this taping shortfall was productive on the FRS). We believe the overall impact of these cases of non-observation is relatively minor. Nonetheless, some caution should be used in generalising from the estimates based on this analysis to other studies, especially if these have different types of target populations or employ different types of interviewers.

Despite these caveats about the representativeness of the estimates, the relative differences between the FRS and Political Tracking studies remain. These are probably best explained by a few key survey differences. The FRS employs an advance letter and, as far as possible, a dedicated field-force who work on the FRS month in month out. Most interviewers working on the FRS have worked on it at least once before and all have received general and survey specific training. This means that it is likely that the FRS interviewers have had time to work out the best approach for them on this survey and are unlikely to hear novel objections from address residents to which they do not have a 'ready-made' response. There is also an element of self-selection occurring in the construction of the field force which is likely to have a beneficial effect on response rates. By this we mean that interviewers who do not like the survey will drop out while those who are comfortable with it (and thereby achieve higher response) will be keen to work as many months as possible. The positive effects on response that these types of organisational procedures can have is reflected in the fact that SCPR's response rate on the FRS has shown a steady increase since 1993. Although this effect incorporates the impact of a declining non-contact rate, refusal reduction has nevertheless played a significant role.

In contrast to this, Political Tracking began in 1994 and is scheduled to take place once every other year. Therefore due to its shorter history and less frequent administration, the interviewers working on the Political Tracking survey are likely to be less familiar with it and therefore less able to persuade address residents to participate when compared to the interviewers working on the FRS. The self-selection factor is also likely to play a far less significant role.

In summary then, despite the fact that, on the face of it, people are less likely to agree to give interviews on details of their income and expenditure than on their political attitudes, the FRS still achieves a higher response rate than the Political Tracking survey. This we feel can be attributed, at least in part, to the superior doorstep technique of the FRS interviewers revealed by this analysis. Under this model of survey response, many of the organisational/procedural survey effects on nonresponse are ultimately the same as those argued to derive directly from the interviewers' doorstep technique. This is because, if the macro level structural influences that differentiate the two surveys actually do affect response, then these must ultimately be manifested through individual interviewers on the

doorstep. By this reasoning the survey differences outlined above do not influence response rates *directly* but do so *indirectly* through enabling interviewers to improve their doorstep technique for each particular survey. Thus, the fact that the FRS is a long-standing (compared to Political Tracking), monthly, continuous survey with a 'core' field force *indirectly* affects response by influencing the *direct* effect of individual interviewers' doorstep techniques. In this light, it is important to note that in saying that the FRS interviewers have displayed superior doorstep technique in this study, we are not attributing this to the interviewers per se, nor simply to differences in training practices across organisations (although the SCPR interviewers who conduct the FRS do receive specific training in doorstep techniques) but predominantly, to specific differences in the natures of the two surveys under analysis.

Thus, the key objective from the perspective of minimising refusals must be to *simulate* these indirect influences by developing training schedules which provide interviewers with the same or similar effects as the ideal organisational/procedural ones. This is not to deny that there are issues common to most or all surveys that interviewers should be equally familiar with in order to accomplish a satisfactory doorstep technique on any given survey. Such 'pan-survey' issues include the subject of respondent confidentiality; the ethical implications of survey work; and the way individuals and households are selected. What must be achieved is an integration of this more general survey knowledge with an understanding of each specific survey. This is important as the extent to which the more general issues come up on the doorstep will be directly related to the nature of each individual survey. For example privacy/confidentiality is likely to be at its most pertinent when the survey deals with issues of income and expenditure as in the case of the FRS.

The lesson to be learned here, we feel, is that, in practical terms, the more familiar an interviewer is with the particular survey they are working on, the better their doorstep technique (and consequently their response rate) is likely to be on that survey. In training terms, the key implication is that interviewer briefings should concentrate, as much as possible, on familiarising interviewers with the nature of the survey they will be administering and to ensure that interviewers are aware of, and able to respond satisfactorily to, the most common questions that will be asked of them when they are trying to persuade reluctant respondents on the doorstep. Information on the most common 'pan-survey' respondent questions and statements can be acquired through reference to previous research in this area (chapter 4 of Campanelli et al. 1997, Morton-Williams 1993 and Couper 1995). Information on the more 'survey-specific' respondent behaviours can be obtained from interviewers who have experience of working on the particular survey (or on similar surveys if it is a one-off).

5 Conclusions

The main aims of this research have been to develop a typology of interviewer-address resident interactions and, from this typology, to determine the scope for reductions in the refusal component of nonresponse through improved interviewer persuasion strategies. The construction of the typology has therefore been driven by the requirement that differences between 'types' of interaction should be defined primarily by the degree of reluctance expressed by the respondent (and therefore the degree of persuasion required on the part of the interviewer).

Clearly, when attempting to establish the scope for reducing refusal rates our interest lies neither in interactions where no persuading is necessary (the hard-core of respondents) nor in those where no persuading is possible (the hard-core of nonrespondents). As we are interested in assessing the scope for reductions in refusal rates below *current* levels, we are also not really interested in interactions where only the most basic levels of persuasion are necessary to achieve an interview (given that the level of persuasion skill required in these interactions would already be expected of even the least experienced interviewers in organisations like SCPR and NOP). Nor should we be particularly interested in interactions where, although reluctant address residents allow interviewers an opportunity to persuade them to participate, there is nothing that the interviewer identifiably said or did wrong or even could have done differently to avoid the eventual refusal. When interviewers do everything expected of them (and sometimes more) throughout an interaction with a address resident and still fail to achieve an interview, we must assume that such interactions do not constitute an area where there is genuine scope for reductions in refusal rates through improved doorstep technique alone.

This leaves only two remaining interaction types from our coding frame. These are types 2 (interview - above baseline skills) and 4 (refusal - could have done more). It can be seen that these represent the interactions in which it is assumed that address residents are maximally ambivalent about participation. The scope for reducing refusal rates lies in shifting as many sample units as possible from type 4 interactions to type 2 interactions. In the two surveys under analysis the joint proportions of sample units falling into these two coding categories are nearly equivalent - 13% (Political Tracking) and 19% (FRS), but below the figures one might have naively imagined, as these suggest that the professional interviewer (at least in terms of the verbal aspects of the doorstep interaction) is only having an impact on about 1 in 5 cases, a small but nonetheless important percentage. (One mustn't forget, however, the baseline skills that professional survey interviewers easily employ to gain interviews among type 1 interactions. These interviews might not be so easily gained in other types of interview situations.)

The results also suggest that the FRS data has far more type 2 interactions than type 4 interactions (17%:3%) while the Political Tracking data has more equal proportions (5%:8%).

The reason for this discrepancy was argued to be a result of the design differences between the two surveys. Because the FRS is a monthly, continuous survey with a 'self-selecting' core field force, interviewers working on it are more likely to be knowledgeable about the survey and to be prepared for the sorts of respondent comments most frequently encountered on the survey. In comparison to this, Political Tracking is a newer survey which is conducted only biannually. Thus, in comparison to the FRS, the Political Tracking interviewers are likely to be less familiar with the exact nature of the survey they are trying to 'sell' and with the most frequently encountered respondent doorstep comments.

This suggests that there is not much scope for improvement in response rates via reductions in refusals in the FRS but that by concentrating on improving interviewer approaches, the Political Tracking survey could make a significant improvement in response rates by reducing the frequency of refusals. This can be done by simulating some of the positive organisational qualities from surveys like the FRS in the training programmes of other surveys such that interviewers are as familiar with the survey and what people on the doorstep are likely to say about it as they would be had they been working on a continuous survey for several months. As one interviewer put it

'You have to sell it to us first, before we can sell it to others.'

In addition interviewers can be instructed in the value of maintaining interaction to probe for opportunities to persuade and avoiding immediate participation decisions. They can also be encouraged, through discussion groups, to share their own experience and 'favourite' strategies.

References

Campanelli, P., Sturgis, P. and Purdon, S. (1997). Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates. London: SCPR

Colombo, R. (1983). Patterns of Nonresponse. In: Hoinville, G. (ed): SSRC Survey Methods Seminar Series, 1980-83, London: SCPR

Couper, M. (1995). The Impact of Survey Introductions on Data Quality, paper presented at tl International Conference on Survey Measurement and Process Quality, Bristol, UK

Couper, M. and Groves, R. (1991). The Role of the Interviewer in Survey Participation, paper presented at the annual conference of the American Association for Public Opinion Research, Phoenix, Arizona

Dijkstra, W. and Van der Zouwen, J. (1982). Response Behaviour in the Survey Interview, London: Academic Press Inc.

Dillman, D., Gallegos, J., and Frey, J. (1976). Reducing Refusal Rates for Telephone Interviews, Public Opinion Quarterly, 40, pp. 66-78

Durbin, J. and Stuart, A. (1951). Differences in Response Rates of Experienced and Inexperienced Interviewers, Journal of the Royal Statistical Society, Series A, p. 114

Fowler, F. and Mangione, T. (1990). Standardized Survey Interviewing, Newbury Par California: Sage

Groves, R., Cialdini, R., and Couper, M. (1992). Understanding the Decision to Participate in a Survey, Public Opinion Quarterly, 56, pp. 475-495

Hyman, H. (1954, 1975). Interviewing in Social Research, Chicago: The University Chicago Press

Kahn, R. and Cannell, C. (1957). The Dynamics of Interviewing, New York: Wiley

Lievesley, D. (1986): Unit Nonresponse in Interview Surveys, JCSM unpublished paper, London: SCPR

Morton-Williams, J. (1993). Interviewer Approaches, Aldershot: Dartmouth Publishir Company Limited

Oksenberg, L., Coleman, L., and Cannell, C. (1986). Interviewer Voices and Refusal Rates Telephone Surveys, Public Opinion Quarterly, 50, pp. 97-111

O'Neil, M. et al. (1980). Telephone Interview Introductions and Refusal Rates: Experiments in Increasing Respondent Cooperation, proceedings of the American Statistical Association, Section on Survey Research Methods

Singer, E. and Kohnke-Aguirre, L. (1979). Interviewer Expectation Effects: A Replication ar Extension, Public Opinion Quarterly, 43, pp. 245-260

Sudman, S. and Bradburn, N. (1974). Response Effects in Surveys, Chicago: Aldine

Incentives in Two German Mail Surveys 1996/97 & 1997

JANET HARKNESS, PETER MOHLER, MICHAEL SCHNEID AND BERNHARD CHRISTOPH

Abstract: *Two mail surveys, each with three mailings, were carried out with nationally representative samples in eastern and western Germany in winter 1996/97 and spring 1997. Data is available on respondents and nonrespondents for both. In the 1996/97 study this is limited for nonrespondents to sex, age and regional location. In the second study rich data is available on everyone eligible. The first study used an enclosed 'near-to-cash' incentive (postage stamps), the second promised inclusion in a special lottery in return for participation. Detailed records of response and other reactions were kept. This paper presents the first results from the studies. Recent research on incentives and nonresponse in studies with more than one mailing is scarce in Germany. Our findings on incentive impact are in line with findings for North America, both with respect to the generally positive effect of using incentives and to differences in the impact of promised and enclosed incentives. Stamps seem to have been perceived as 'cash' gifts. Higher refusal rates among those sent stamps are seen as increases in 'reaction-response' and are discussed in terms of communicative obligations as well as cognitive dissonance. Age and sex differences in the two studies are discussed, as are costs and cost-effectiveness for the two kinds of incentive.*

Key words: *incentives, lottery, stamps, mail, response*

1 Introduction

In December 1996 ZUMA participated in an international research project on administration mode effects across countries and languages.[1] The project included an experiment designed to gather information on incentives in mail surveys in Germany. Immediately following this study, a second incentives experiment was conducted as part of the 1997 International Social Survey Programme (ISSP) study. This paper presents the first results from the incentive studies.

---

[1] The stamps study replicated the ISSP module of the same year, except that a postal administration replaced *self-completion with interviewer attending*. In fact, a third of respondents in the main ISSP study completed the questionnaire as a face-to-face interview and so the ISSP 1996 dataset has a variable: *mode*.

2 Method

Both experiments were part of almost identical survey designs.

- Nationally representative, randomly selected samples of adults living in private accommodation in eastern and western Germany;
- Names and addresses from the residents' registers maintained by municipalities and drawn for the 1996 ALLBUS sample (for details of the ALLBUS sample, see Wasmer et al. 1996);
- Mail surveys with two follow-up mailings;
- Module design, accompanying materials, and fielding designs were identical apart from the incentive type offered and a one-page report on 1996 findings included in the second mailing for one of the studies (see Table 1);
- Each module was an ISSP survey, thus comparable in tone, length (ca. 60 ticks), background variable questions, investigatory approach, appearance and respondent burden.

One important difference between the two studies is that the 1996/97 study used fresh addresses, that is, contacted people for the first time, whereas the 1997 study was a panel; the names and addresses were of people who participated in the 1996 ALLBUS. The other important difference was the type of incentive offered.

Incentives Offered

1996/97 STAMPS: The ISSP Role of Government III module was sent to 1,296 people. Fifty percent of these (648 people), selected on a random basis, were sent four normal letter rate postage stamps in their first mailing. The stamps were described as a thank-you gesture for their help.

1997 LOTTERY: The ISSP Work Orientations II module was sent to 3,711 people. Two-thirds (2,475), selected on a random basis, were given the chance of being one of three winners of DM 1,000 if they returned their completed questionnaire within a deadline. The offer was described as a thank-you gesture for their help.

Information on Targeted Respondents

STAMPS: Sex, age, regional location and size of community is available for non-respondents.

LOTTERY: Sex, age and regional location is available for 18-year olds who were too young to participate in the 1996 ALLBUS-ISSP studies but were sent questionnaires in the 1997 lottery study. Everyone else sent questionnaires participated in the 1996 ALLBUS study. In the analyses discussed here, the 18-year olds were excluded from both studies, since in the lottery study they were the only people not part of a panel.

Implementation

Table 1 provides an overview of sample information, mailing contents, fielding dates, etc.

Table 1: Overview of samples, mailings and fielding information

SAMPLE	STAMPS	LOTTERY
coverage	adult population in private accommodation	adult population in private accommodation (panel)
issued	1,296	3,711
eligible	1,206	3,540
representation east and west	over-representation eastern Germany: 888 west, 408 east	over-representation eastern Germany: 2,519 west, 1,192 east
incentive coverage	50% (648) offered stamps	2/3 (2,475) offered lottery chance
FIELD	STAMPS	LOTTERY
Dates	begin: 5/12/96 end: 3/2/97	begin: 20/2/97 end: 16/5/97
1st mailing	December 5	February 20
2nd mailing	December 17	March 6
3rd mailing	January 2	April 3
Mailing contents	STAMPS	LOTTERY
all mailings	personally addressed letters, institute contact person's name, address and telephone no.	as for STAMPS.
1st mailing	questionnaire, data protection information, pre-paid return envelope, letter, incentive/no incentive	as for STAMPS
2nd mailing	thank you-cum-reminder	as for STAMPS, plus report on 1996 data
3rd mailing	questionnaire, data protection information, pre-paid return envelope, letter	as for STAMPS, plus reminder mention of lottery to relevant split

Field Contact Records

In both studies, detailed records were kept of information on non-participation. This covered people reported to have moved, died, as too ill or absent from home or otherwise unable to participate, as well as details on active refusals for both treatment groups, transcripts of comments included on questionnaires and any other information obtained, such as people returning stamps, empty questionnaires, etc. A summary is provided in Table 2.

Table 2: Overview of field records kept

STAMPS	
Questionnaire	Completed Incomplete Empty Questionnaire 1 or 2 used Respondent comments on questionnaire
Field Record	Date received Moved, unknown, etc., (post office categories) Deceased Too ill to participate Currently abroad/away Other 'No Participation' Refused (details)
Stamps	Incentive/ No incentive Used for return mail (details) Returned in envelope
R comments	(available as file)
Other	(details)

LOTTERY	
Questionnaire	as for STAMPS
Reactions	as for STAMPS
Lottery chance	Incentive/No incentive Winners' reactions
R comments	(available as file)
Other	(details)

3 Incentives in mail surveys

Reviews and meta-analyses from other countries, in particular the United States, indicate that incentives generally improve response rates (e.g., Church 1993, Yammarino et al. 1991, Fox et al. 1988, Armstrong 1975, Gajraj et al. 1990, Brennan 1992, Brennan et al. 1991). Enclosed incentives have proved more effective than promised incentives (e.g., Goyder 1994, Church 1993, Gajraj et al. 1990, Kalafatis and Madden 1994, Yu and Cooper 1983) and cash more effective than other token gifts (e.g., Linsky 1975, Kanuk and Berenson 1975, Goodstadt et al. 1977, Hansen 1980). Research also indicates promised lottery participation (of various kinds) is less effective than cash (Gajraj et al. 1990, Hubbard and Little 1988, McDaniel and Jackson 1984) with varying results on whether it is more effective than other promised incentives or than no incentive (Gajraj et al. 1990, Hubbard and Little 1988, McDaniel and Jackson 1984). While Church (1993) suggests that promised inducements are basically not worth the effort, Gajraj et al. (1990) see promised lottery incentives as a cost-effective possibility for large populations.

Findings available for Northern America need not necessarily hold for other continents and countries (Brennan 1992, Eichner and Habermehl 1981, Heberlein and Baumgartner 1981, cf. Yammarino et al. (1991: 630) on intracultural differences). Attitudes and exposures to games of chance and lotteries, for example, clearly differ across cultures, as does exposure to mail inducements. Thus while research on inducements in Germany is relatively scarce and information is still needed on their (cost-)effectiveness, American researchers are already pondering the consequences of working with a population accustomed to inducements and aware of mechanisms to maximise their rewards as respondents.

3.1 Goals of the incentives studies

The two studies described here served multiple research purposes described elsewhere (Harkness, forthcoming). The incentive experiments themselves also served numerous purposes not dealt with here. The goals of relevance to the present paper are:

- to measure the effect of giving 'near-to-cash' stamps to respondents;
- to measure the effect of offering the chance of a lottery cash win;
- to compare and contrast these effects and their cost-effectiveness across the two studies.

3.2 Incentive costs and rationale

STAMPS: Treated respondents were sent 4 one deutschmark stamps - at the time sufficient for four letters or greeting cards. Stamps were used for a number of reasons. For example, they are often used in Germany instead of coins to discharge small payments by mail and are technically reimbursable at post offices. A Christmas motif was chosen for the stamps enclosed and the stamps on the mailings, in the hope of increasing their attractiveness (cf. Dillman 1978, Church 1993). Intuitively, the research team felt that enclosing stamps would be more acceptable to respondents than enclosing coins, but the (national) sample was too small to test this, too. Incidentally, two fairly large-sized coins would have been involved for this sum. Lastly, it was thought that the *perceived usefulness* of stamps could be heightened by virtue of the near-to-Christmas fielding dates.

LOTTERY: Treated respondents were offered the chance to be one of three winners of DM 1,000, a sum which amounts to roughly one quarter of the average monthly income in Germany. We refer to Gajraj et al. (1990) for several important differences between promising a monetary reward upon completion and offering the chance to win money.

The lottery in the ZUMA study was a special draw limited to participants in the survey and carried out by a notary public - a respected public figure in Germany. Thus while the sum to be won was a much smaller amount than the multi-millions in national lotteries, three respondents were certain to win. Moreover, the individual chances of winning were very much greater than in an national lottery, even if the letters mailed did not particularly stress this. Given the depressed economic climate in Germany at the time, and the general cultural acceptance of low investment lottery 'flutters', it seemed likely that a lottery could prove motivating.

4 Overview of findings

Table 3 provides a general outline of our results. It contains response data for both studies, and distinguishes between realised interviews, refusals, issued, eligible and ineligible and non-reactions. The reported percentage differences are significant on a 95 % level.

The stamps incentive resulted in almost a five percent increase in response rate among eligible targeted respondents offered the incentive compared to eligible targeted respondents not offered the incentive. The lottery offer resulted in almost a three percent increase in response among eligible targeted respondents who were offered the chance to win money and those eligible not offered the chance to win money. As described below, the percent difference between responses from incentive respondents and those from respondents without incentives differed over time for both studies. The refusal rate in the stamps study was higher in both treated and control group than the total refusal rate and both the incentive and no incentive group refusal rates in the lottery study. We consider possible reasons for this in the next section.

4.1 Refusal as response

As Table 3 reflects, not only did more treated people complete and return questionnaires than the control group in the stamps study, more of those receiving stamps actively refused to participate. In the lottery study, the incentive was a weaker inducement to participate, but with one age group exception (see 4.4), people were also not prompted to refuse actively. (For discussion of refusals and return-refusals beyond the scope of this paper, see Brennan et al. 1991, Brennan 1992, Brennan and Hoek 1992). A number of factors can contribute to these differences.

Table 3: Responses: stamps study and lottery study

Stamps	Realised	Ineligible	Refusals	No reaction	Issued	Eligible
Incentive	210	35	51	352	648	613
% of issued	32,4	5,4	7,9	54,3	100,0	
% of eligible	34,3		8,3	57,4		100,0
Control	174	55	31	388	648	593
% of issued	26,9	8,5	4,8	59,9	100,0	
% of eligible	29,3		5,2	65,4		100,0
% Difference stamps-control (eligible)	5,0		3,1	- 8,0		
% Difference stamps-control (issued)	5,5		3,1	- 5,6		

Lottery	Realised	Ineligible	Refusals	No reaction	Issued	Eligible
Incentive	1159	156	38	1122	2475	2319
% of issued	46,8	6,4	1,5	45,3	100,0	
% of eligible	50,0		1,6	48,4		100,0
Control	546	75	31	583	1235	1160
% of issued	44,2	6,1	2,5	47,2	100,0	
% of eligible	47,1		2,7	50,3		100,0
% Difference draw-control (eligible)	2,9		- 1,1	- 1,9		
% Difference draw-control (issued)	2,6		- 1,0	-1,9		

(Percentage values are rounded off to one decimal place)

First, the issued population for the lottery study were people who participated in the ALLBUS 1996 study, i.e., a panel. This means that the population consisted of people who, at least in 1996, had been willing to take part in a face-to-face study. Since older

women, for example, are often less willing to participate, older women were under-represented in the ALLBUS study and hence also in the issued cases for the lottery study. And since older women again proved less willing to participate in the lottery study, this group is even further under-represented in the lottery study than in the ALLBUS 1996 study.

On the other hand, the ALLBUS study was carried out without an incentive. Thus the realized sample were all people willing to participate without an incentive. Hence people who could be won over solely by the incentive would be under-represented in the issued population for the lottery study. We see in section 4.4 below that older women in the lottery study do in fact have different reaction patterns from older women in the stamps study. Admittedly, various differences, including mode differences, compound this issue, since interviewers might persuade in instances where mailings do not (cf. Hox and De Leeuw 1994).

Second, the stamps study was fielded over Christmas and New Year, at a time usually not recommended in Germany for studies apart from seasonal projects. This could have contributed to differences in response rate between the two studies. It is less clear how the fielding dates might contribute to the overall higher active refusal rate in the stamps study. Moreover, the lottery study ran through Easter, a time when many people take spring vacations and in Germany are busy to some extent with Easter gifts and celebrations.

Third, the studies differed in topic and topic saliency affects response. The stamps study was on attitudes to the state and government. The lottery study was about attitudes to work. People not in paid employment were filtered past quite a few questions. Arguably, people who had never or no longer worked for pay could feel themselves unaddressed and experience less cognitive dissonance or sense of obligation to react (along the lines of 'this doesn't apply to me anyway').

From a cognitive dissonance perspective (Hackler and Bourgette 1973, Furse and Stewart 1982), an unsolicited near-to-cash gift of modest dimensions ought to create enough cognitive dissonance to prompt participation in order to discharge the social obligation thus created. As Dillman (1978:16) argues, a modest cash gift is more likely to be perceived as a token of appreciation, whereas somewhat larger gifts run the risk of being judged in terms of their adequacy as renumeration.

We suggest a social discourse perspective can usefully be added to this model of social exchange. Interaction between researchers and respondents via instruments (+/- interviewer) has been seen from a communicative (Schwarz 1996) or discourse (Harkness 1996) perspective. There seem to be few good grounds for limiting this communicative perspective to respondents whose communicative activity consists of answering

questionnaires, while excluding those who communicate by indicating that they do *not* wish to comply. In other words, respondents who actively refuse can also be seen as following the logic and principles which govern our social communicative acts, in terms of, for example, speech act theory (Austin 1962) and the cooperative principle underlying social interaction through conversation identified by Grice (1975). By actively responding in the form of a refusal to the request to participate, these targeted respondents, just like respondents who participate, discharge the discourse (and social exchange) obligation they experience after the first mailing with its request and, in some cases, incentive. Follow-up mailings may strengthen respondents' sense of obligation to reply, by virtue of repeating the request, by virtue of reminding them indirectly that they have received an incentive, or through a combination of both, as the case and treatment may be. In this context, it is worth mentioning that a number of people who refused to participate in the stamps study returned the stamps, returned the questionnaire, and/or used the stamps (unnecessarily) to return the questionnaire. Others wrote strongly worded refusals. For those disinclined to participate, dissonance may arguably grow, while unwillingness to participate may persist. In discourse terms this can then result in an active refusal, rather than passive (nonresponse) non-compliance with the request to participate.

The promised lottery incentive increased response but had a weaker and uneven impact (see section 4). Since the incentive was only *promised,* it may also have prompted weaker perceptions of an obligation to react. From a discourse politeness standpoint - where discourse politeness is to be understood as 1) being driven by the need to save one's own and others' 'face' and 2) involving a wide-ranging set of strategies to do so (Brown and Levinson 1978), people uninterested in participating were under no social obligation to *say* they were uninterested. Nonetheless, repeated mailings could create the need to 'answer' (refuse) to get the mailings to stop. In the case of the oldest women, active refusals seem to have been felt necessary.

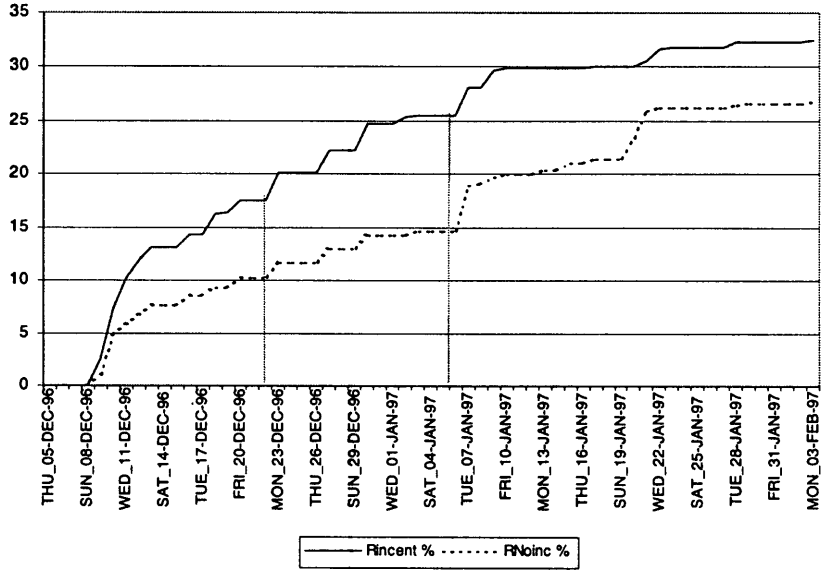
4.2 Response over time

In what follows, we distinguish between *responses*, *reactions*, *refusals* and *nonresponses*. Responses involve returned completed questionnaires, refusals are active indications that targeted respondents are not prepared to participate. These are most often verbally expressed refusals but empty returned questionnaires were also counted as refusals. Reactions include both responses and refusals. Communications indicating targeted respondents were, for example, dead, are included in the ineligibles count. No reaction of any kind from eligible targeted respondents are taken as nonresponses.

The impact of the incentives differed over time in both studies and the response advantage over the respective control group decreased in both studies over the data collection period. In each respect, however, the stamps study shows greater differences. Graphs 1

and 2 present the cumulative responses (percentages) over the three mailings on a daily basis for each study. The days of fielding are entered on the horizontal axis, the percentages on the vertical axis. The vertical lines in each graph mark the beginning of the return from second and third mailings, reckoned from the first reasonable return date for a respective mailing. The two horizontal lines represent the response over the fielding period; the broken line represents no-incentive response, the unbroken line represents incentive response.

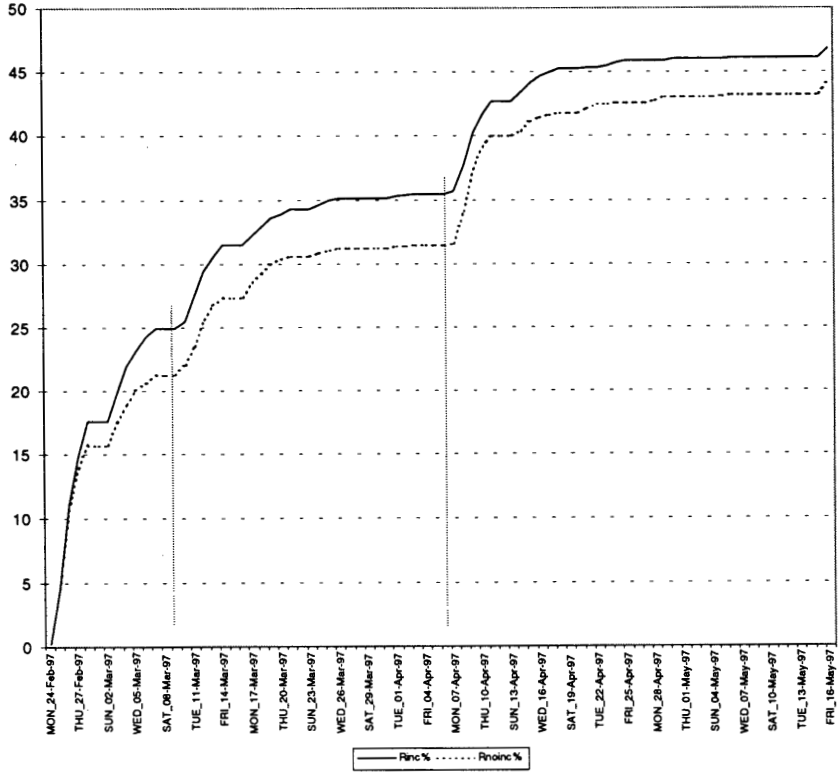
Graph 1: *Stamps* response in percent



Graph 1 shows how the stronger impact of stamps, which results in a 7 to 10 percent difference in response in the second mailing period between incentive and no-incentive groups, drops after the middle of the third to circa 5 percent. It seems likely that if only two mailings had been made, the total number of responses would have been smaller but the response difference between incentive and no-incentive groups would have been approximately double the 5 percent finally realised.

From Graph 2 we see that the difference between incentive and no-incentive response in the lottery study is consistently much smaller but also fluctuates less. In the second mailing period, it went from 2 to 4 percent and in the third, it stabilises around 3 percent. Thus if only two mailings had been involved here, the response difference between treatments is likely to have been about 1 percent higher than the 2,9 percent finally realised.

Graph 2: *Lottery* response in percent



4.3 Costs and cost-effectiveness

Table 4 details the costs involved in providing the incentives for each study. These are independent of all the other costs for the studies and they also exclude the re-contact expenses saved through receiving reactions before the third mailing. In the stamps study, the total number of increased responses over the three mailings (treatment versus control) was 36 (cf. Table 3), thus third mailing savings in handling and postage were minimal. For the lottery study, the number of ineligibles by the end of the first mailing period is uncertain, which affects calculations. As it is, some of the details of how and when mail was returned as 'undeliverable' raise questions about the accuracy of German post office information on returned mail.

The most central consideration in discussing the costs of the two studies is that the costs of providing a stamps incentive are directly dependent on the size of treated population, whereas the incentive costs of providing the lottery are independent of the number of people targeted. Thus a much larger targeted population could have been offered a lottery incentive for the same incentive costs, although all other costs would, of course, increase (handling, mailing, paper, data editing, etc.).

Table 4: Incentive costs

STAMPS 1996/97		LOTTERY 1997	
Stamps for 648 respondents (= 648 x DM 4,00)	DM 2,592	Lottery prize money	DM 3,000
Circa 6 hours student assistant help	DM 180	Notary Public	DM 200
		Institute handling of draw	compli-mentary
Total (spent at begin of fielding)	DM 2,772	Total (spent at end of fielding)	DM 3,200

Table 5 sets out these costs in relation to the total number of treated respondents, the realised interviews, and the costs per additional interview in comparison to the control group for each study.

Table 5: Costs and cost-effectiveness

	Stamps	Lottery
Response rate	34,3 %	50,0 %
Eligible	613 (issued 648)	2319 (issued 2475)
Total costs	2.772 DM	3.200 DM
Realized (n)	210	1159
Costs per realized interview	13,20 DM	2,76 DM
Costs per additional person gained	89,40 DM	47,76 DM
Incentive increase	5 % (n=31)	2,9 % (n=67)
Costs per 1 % increase	554.40 DM	1,103.45 DM

4.4 Age and gender differences in response and nonresponse

The reaction behaviour (responses, refusals and nonresponses) of men and women differed within and across the two studies. Tables 6 and 7 provide the details. Eighteen-year olds have been excluded from both samples since in the lottery study they were the only people not part of a panel.

STAMPS: In the stamps study, the issued population was made up of 52 % women and 48 % men, that is, there were 4 % more women. The realised sample, however, contains 6 % more men than women (53 % men and 47 % women). Not only did more women not respond than men, a greater proportion of women actively refused. 62 % of the active refusals came from women. Moreover, while the issued incentive and control groups were equally balanced with regard to the sexes, a greater percentage of women responded to the incentive than did men. The difference between incentive response and no-incentive response for men is 3,8 %. For women in the stamps study, the incentive versus no-incentive response difference is 9,3 %. While the case numbers involved are small, these differences in reaction behaviour also seem to be dependent on age. Women between 45 and 64, followed by the next oldest age group, 65-74, were more receptive to the incentive than other age groups of women. Women of 75 and over reacted least positively, both in terms of nonresponse (12 %) and in terms of active refusals (21 %).

With the exception of the 35-44 year olds, men in general were at least as likely to respond without incentive as women with incentive.

Table 6: Sex and age group differences in reaction and nonresponse: stamps study (N)

Stamps			Treatment							
			Incentive				Control			
				Reaction & Nonresponse				Reaction & Nonresponse		
			Sum Rows Incentive	Re-sponse	Re-fusals	Non-response	*Sum Rows Control*	Re-sponse	Re-fusals	Non-response
Men	Age	19-24	*18*	8	0	10	*20*	7	0	13
		25-34	*58*	21	4	33	*72*	26	2	44
		35-44	*59*	19	3	37	*51*	11	1	39
		45-54	*53*	18	4	31	*45*	16	1	28
		55-64	*47*	16	3	28	*38*	17	2	19
		65-74	*30*	17	3	10	*35*	16	1	18
		75+	*15*	9	2	4	*20*	7	4	9
Sum Columns Men			*280*	*108*	*19*	*153*	*281*	*100*	*11*	*170*
Women	Age	19-24	*14*	4	1	9	*14*	4	0	10
		25-34	*53*	15	1	37	*54*	11	1	42
		35-44	*48*	13	4	31	*42*	10	2	30
		45-54	*43*	16	3	24	*49*	13	3	33
		55-64	*65*	27	7	31	*52*	14	1	37
		65-74	*56*	18	8	30	*47*	10	7	30
		75+	*33*	4	7	22	*43*	8	5	30
Sum Columns Women			*312*	*97*	*31*	*184*	*301*	*70*	*19*	*212*

LOTTERY: As Table 7 following indicates, differences across age groups and between the sexes are less pronounced in the lottery study. The oldest group of women seem to react negatively to the incentive, as do women between 55-74 to lesser extent. The oldest group has a markedly lower response (21%) than other age groups with incentive, and a lower response than women of the same age in the control group (37%) and a high refusal rate (16 %). In the control group, differences between this age group and other ages are

less pronounced than in the incentive group. Active refusals are noticeably higher in the no-incentive group for the two oldest female age groups (11 % and 13 % respectively). In both incentive and no-incentive groups, the oldest women actively refused most. Women between 65-74 with incentive had, in contrast, few refusals.

Men in general were more likely to participate (52 %) than women (47 %) and were less likely to refuse. The incentive improved response rates (unevenly) across all the different age groups of men, with the exception of the group between 55 and 64.

Table 7: Sex and age group differences in reaction and nonresponse: lottery study (N)

Lottery*			Treatment							
			Lottery				Control			
				Reaction & Nonresponse				Reaction & Nonresponse		
			Sum Rows Incentive	Re-sponse	Re-fusals	Non-response	*Sum Rows Control*	Re-sponse	Re-fusals	Non-response
Men	Age	19-24	*83*	49	0	34	*65*	35	0	30
		25-34	*170*	87	1	82	*152*	74	0	78
		35-44	*175*	97	0	78	*176*	78	2	96
		45-54	*142*	79	1	62	*135*	72	3	60
		55-64	*154*	82	1	71	*161*	90	6	65
		65-74	*74*	40	3	31	*88*	44	2	42
		75+	*30*	16	2	12	*35*	15	2	18
Sum Columns Men			*828*	*450*	*8*	*370*	*812*	*408*	*15*	*389*
Women	Age	19-24	*63*	38	0	25	*53*	27	0	26
		25-34	*150*	75	0	75	*166*	84	2	80
		35-44	*166*	86	1	79	*145*	74	2	69
		45-54	*136*	74	1	61	*143*	69	3	71
		55-64	*152*	70	5	77	*151*	77	3	71
		65-74	*98*	45	1	52	*124*	36	14	74
		+75	*63*	13	10	40	*70*	26	9	35
Sum Columns Women			*828*	*401*	*18*	*409*	*852*	*393*	*33*	*426*

*Cases are weighted to adjust for the 2/3 to 1/3 distribution of incentive and control group

less pronounced than in the incentive group. Active refusals are noticeably higher in the no-incentive group for the two oldest female age groups (11 % and 13 % respectively). In both incentive and no-incentive groups, the oldest women actively refused most. Women between 65-74 with incentive had, in contrast, few refusals.

Men in general were more likely to participate (52 %) than women (47 %) and were less likely to refuse. The incentive improved response rates (unevenly) across all the different age groups of men, with the exception of the group between 55 and 64.

Table 7: Sex and age group differences in reaction and nonresponse: lottery study (N)

Lottery*			Treatment							
			Lottery				Control			
				Reaction & Nonresponse				Reaction & Nonresponse		
			Sum Rows Incentive	Re-sponse	Re-fusals	Non-response	*Sum Rows Control*	Re-sponse	Re-fusals	Non-response
Men	Age	19-24	*83*	49	0	34	*65*	35	0	30
		25-34	*170*	87	1	82	*152*	74	0	78
		35-44	*175*	97	0	78	*176*	78	2	96
		45-54	*142*	79	1	62	*135*	72	3	60
		55-64	*154*	82	1	71	*161*	90	6	65
		65-74	*74*	40	3	31	*88*	44	2	42
		75+	*30*	16	2	12	*35*	15	2	18
Sum Columns Men			***828***	*450*	*8*	*370*	***812***	*408*	*15*	*389*
Women	Age	19-24	*63*	38	0	25	*14*	27	0	26
		25-34	*150*	75	0	75	*54*	84	2	80
		35-44	*166*	86	1	79	*43*	74	2	69
		45-54	*136*	74	1	61	*49*	69	3	71
		55-64	*152*	70	5	77	*52*	77	3	71
		65-74	*98*	45	1	52	*47*	36	14	74
Sum Columns Women			***828***	*401*	*18*	*409*	***302***	*393*	*33*	*426*

*Cases are weighted to adjust for the 2/3 to 1/3 distribution of incentive and control group

In both studies, therefore, differences can be noted across age groups and across genders. In cross-cultural research in which questionnaires need to be translated, 'gendered' questionnaires have been proposed (Acquadro et al. 1996). In similar fashion, one might consider, whether 'gendered' and cohort-targeted incentive treatment could optimise incentive implementation and reduce costs. This would, however, require a better understanding of why, for example, the lottery incentive seems to have had a negative effect on older women.

5 Conclusion

The two studies can be seen as indicating that German targeted populations tend to react and respond in central respects in a fashion comparable to targeted respondents in Northern America. Promised rewards are less effective than delivered rewards, even when the promised reward could lead to considerable financial gain and the delivered reward amounts to only a small token (close to) cash gift. The lottery offer was in general a weakly positive incentive but seems to have had a negative effect on older women. Since the costs of a lottery can be made independent of sample size - and other models have involved considerably lower costs than our triple 'certain handout' study - the results suggest that this form of incentive is arguably useful and cost-effective given large sample sizes and a limited fielding period (cf. Gajraj et al. 1990).

In both studies, as reported elsewhere, older women prove less willing to participate in general. Since they were also more likely than other groups to refuse actively, their non-participation in these studies does not seem to be coupled to non-awareness of a social obligation. Lottery and trading organisations of dubious standing who conduct their business by mail and regularly offer pseudo-competitions and lottery inducements may affect perceptions. Certainly, among older women offered the lottery, fewer participated than among those not offered the lottery.

The cost of a stamps incentive is directly linked to the size of the treated population. The response enhancement achieved was fairly modest, if better than the lottery. If the costs per respondent could be reduced, it could be considered as a useful response enhancer, particularly if fielding is limited to one or two mailings. American studies have achieved positive results with very small amounts of cash. The 'lucky penny' enclosed by some German lottery firms and catalogue houses aspiring for mail trade might well disqualify this as a survey response incentive. It would be relevant to establish, however, whether enclosing one low-value stamp could be just as effective.

References

Acquadro, C., Jambon, B., Ellis, D. and Marquis, P. (1996). Language and translation issues. In: Spilker, B. (ed.): Quality of Life and Pharmacoeconomics in Clinical Trials (2nd ed.). Philadelphia: Lippincott-Raven

Armstrong, J.S. (1975). Monetary incentives in mail survey. Public Opinion Quarterly 39, pp. 111-116

Austin, J.L. (1962). How to do things with words. Oxford University Press, Oxford

Brown, P. and Levinson, S. (1978). Universals in language usage: politeness phenomena. In: Goody, E. (ed). Questions and Politeness: strategies in social interaction. Cambridge: Cambridge University Press

Brennan, M. (1992). The effect of monetary incentives on mail survey response rates: new data. Journal of the Market Research Society 34, 2, pp. 173-177

Brennan, M. and Hoek, J. (1992). The behaviour of respondents, nonrespondents and refusers across mail surveys. Public Opinion Quarterly 56, 4, pp. 530-535

Brennan, M., Hoek, J. and Astridge, C. (1991). The effects of monetary incentives on the response rate and cost-effectiveness of a mail survey. Journal of the Market Research Society 33, 3, pp. 229-241

Church, A. (1993). Estimating the effect of incentives on mail survey response rates: a meta-analysis. Public Opinion Quarterly 57, pp. 62-79

Dillman, D. (1978). Mail and Telephone Surveys. The Total Design Method. New York: Wiley and Sons

Eichner, K. and Habermehl, W. (1981). Predicting response rates to mailed questionnaires (Comment on Heberlein and Baumgartner, ASR, August 1978). American Sociological Review 46, pp. 361-363

Fox, R. J., Crask, M. R. and Kim, J. (1988). Mail survey response rate. A meta-analysis of selected techniques for inducing response. Public Opinion Quarterly 52, pp. 467-491

Gajraj, A., Faria, A. and Dickinson, J. (1990). A comparison of the effects of promised and provided lotteries, monetary and gift incentives on mail survey response rate, speed and cost. Journal of the Market Research Society 32, 1, pp. 141-162

Goodstadt, M. S., Chung, L., Kronitz, R. and Cook, G. (1977). Mail survey response rates: their manipulation and impact. Journal of Marketing Research 14, pp. 391-395

Goyder, J. (1994). An experiment with cash incentives on a personal interview survey. Journal of the Market Research Society 36, 4, pp. 360-366

Grice, H.P. (1975). Logic and conversation. In: P. Cole and J.L. Morgan (eds.): Syntax and semantics, 9: speech acts, pp. 41-58. New York: Academic Press

Hansen, R. (1980). A self-perception interpretation of the effect of monetary and nonmonetary incentives on mail survey respondent behavior. Journal of Marketing Research 17, pp. 77-83

Harkness, J. (1996). The (Re)Presentation of Self in Everyday Questionnaires. Paper presented at the International Sociological Association Conference, Colchester, England

Harkness, J. ISSP 1997. Work Orientations. ZUMA report on the German study. ZUMA. Mannheim (in preparation)

Heberlein, T. and Baumgartner, R. (1981). The effectiveness of the Heberlein-Baumgartner models for predicting response rates to mailed questionnaires: European and U.S. examples (Reply to Eichner and Habermehl, ASR). American Sociological Review 46, pp. 363-367

Hubbard, R. and Little, E. L. (1988). Promised contributions to charity and mail survey response rates: replication with extension. Public Opinion Quarterly 52, pp. 223-230

Linsky, A. S. (1975). Stimulating responses to mailed questionnaires: a review. Public Opinion Quarterly 39, pp. 82-101

McDaniel, S. and Jackson, R. W. (1984). Exploring the probabilistic incentive on mail survey research. Proceedings of the American Marketing Association Summer Educators' Conference. Bernhardt, K. et al. (ed.). Chicago, pp. 372-375

Schwarz, N. (1996). Cognition and Communication. Judgmental Biases, Research Methods, and the Logic of Conversation. Mahwah: Erlbaum

Wasmer, M., Koch, A., Harkness, J. and Gabler, S. (1996). Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 1996. Mannheim. ZUMA-Arbeitsbericht 96/08

Yammarino, F., Skinner, St. and Childers, T. (1991). Understanding mail survey response behavior. A meta-analysis. Public Opinion Quarterly 55, pp. 613-639

Yu, J. and Cooper, H. (1983). A qualitative review of research design effects on response rates to questionnaires. Journal of Marketing Research 20, pp. 36-44

Promised Incentives on a Random Digit Dial Survey

DAVID CANTOR, BRUCE ALLEN, PATRICIA CUNNINGHAM, J. MICHAEL BRICK, RENEE SLOBASKY, PAMELA GIAMBO AND GENEVIEVE KENNY

Abstract: *This paper presents the results of two experiments carried out to test the effectiveness of promised incentives for a random digit dial survey (RDD) at the initial household contact (introduction) and at the stage of refusal conversion. The results of the experiments find no effect of at either stage of the survey process. These results are contrary to those found by at least one other recent experiment, as well as the use of promised incentives used routinely to convert reluctant respondents by several survey organizations. The possible reasons for not finding an effect are discussed.*

Keywords: *prenotification, interviewer effects, telephone survey*

There has been increasing pressure to maintain response rates on random digit dial (RDD) telephone surveys (Massey et al. 1997). With these pressures, it has become more common to use incentives on RDD surveys. Incentives have been found to be effective in increasing response rates for many types of surveys (Church 1993, Singer et al. 1996). However, their utility in an RDD context is unclear. Research outside of an RDD context has shown that promising money (or a gift) is not nearly as effective as providing the incentive at the time of soliciting cooperation. Since there is not generally an opportunity to pre-pay all RDD sample persons, the utility of promised incentives on RDD surveys is unclear.

There has not been a great deal of research on the effectiveness of a promised incentive in the context of an RDD survey. Of the evidence that does exist, the results provide a mixed picture (CMOR 1996, Strouse and Hall 1997). The purpose of this paper is to report results from two experiments testing promised incentives on an RDD survey. One experiment tests promised incentives at the initial call, while the second experiment tests incentives at the refusal conversion stage.

1 Design of the National Survey of America's Families

The experiments were conducted as part of the National Survey for America's Families (NSAF), a RDD survey funded by a consortium of private foundations in the United States. It is being conducted by Westat for the Urban Institute. The purpose of the survey is to assess the impact of recent changes in the administration of a number of assistance programs for children and the poor.

The NSAF consists of both a screening and an extended interview. The screening interview consists of a 3-5 minute battery of questions that is designed to select the person that should be administered the extended interview. This involves determining if there are any persons under 65 years old in the household and whether or not the family is above or below 200% of poverty. If there is someone in the right age-range and the household is sampled (based on poverty status) a respondent for the extended interview is selected. The extended interview is approximately 45 minutes in length and covers a wide range of topics, including health, education, child care, income and receipt of social services.

The experiments discussed below were done at the screening stage of this process. Two important features of this design should be noted. First, because of the sampling algorithms, a majority of those that are called to do the screener are not eligible for the extended interview. They are eliminated either because they do not have any person in the household in the right age range or because they are not below 200% of poverty (approximately half of the households that are above 200% of poverty are sampled). Second, the person that answers the screening instrument is not necessarily selected to do the extended interview. For the sample of families, the extended interview respondent is that person that knows the most about the health and well-being of a randomly selected child (based on rosters compiled during the screener). For the sample of adults, the extended interview is done with a randomly selected adult.

2 Promised incentives at the initial contact

The first experiment was conducted as part of a series of small tests that compared alternative formats and designs of the introduction to the screener. This experiment tested the effectiveness of offering a $5 incentive.

Design: There were a total of 6 experienced interviewers who participated in this experiment. Telephone numbers from a telephone directory were selected for calling. The areas called were concentrated in low income areas of Michigan, West Virginia, Virginia, Massachusetts, and New Jersey.

The incentive introduction was:

> Hello, this is (NAME) and I'm calling for the Urban Institute, a private nonprofit research center that is interested in the well being of adults and families. We are preparing for a study on how changes in health care, education and human services in (STATE) are affecting people like yourself. We are offering $5 to those persons that are eligible and agree to participate in a telephone interview.
>
> To find out if someone in your household is eligible, I need to ask a few questions about who lives there. These questions will take about 3 minutes.

The offer was targeted to the person who was eligible for the study and who agreed to complete an extended interview.

The "no-incentive" introduction was:

> Hello, this is (NAME) and I'm calling for the Urban Institute, a private nonprofit research center that is interested in the well being of adults and families. We are preparing for a study on how changes in health care, education and human services in (STATE) are affecting people like yourself.

The primary outcome measure for the experiments was a "cooperation rate," which consisted of :

CR = (Completes + Ineligibles)/(Completes + Ineligibles + Refusals).

The numerator consists of persons who completed the entire screener because there was an eligible subject in the household and those households where no eligible subjects were found (i.e., household had no persons less than 65 years old). The denominator adds in those persons that refused to provide enough information to determine eligibility. No attempts were made to convert refusals.

Results: As can be seen from the results in Table 1, inclusion of an offer of money actually reduced, rather than increased, the cooperation rate (53.3 vs. 60.6). The difference is marginally significant ($p<.15$ using a two tailed test).

Table 1: Results of promised incentives by stage of offer

	Incentive	No Incentive
At Initial Contact (Incentive of $5)	53.3 (210)	60.6 (231)
At Refusal Conversion (Incentive of $25)	27.4 (1,019)	28.5 (1,034)
Advance letter	27.7 (495)	31.1 (486)
No advance letter	27.1 (524)	26.3 (548)
Left message	35.0 (329)	30.3 (277)
Never left message	23.8 (690)	27.9 (757)
Left message and interviewer was:		
Left on conversion	39.0 (197)	30.1 (143)
Not left on conversion	29.1 (134)	30.6 (134)

One issue that became clear during this testing was that many respondents did not stay on the line long enough to hear the incentive offer. Of the 98 persons that refused the incentive offer, 77 individuals hung up either during or shortly after the introduction was read. Slightly over half of these (48) stayed on the line long enough to hear the offer of $5. The other 29 hung up before the introduction was finished. An additional 11 people hung up either during or right after the first question was read ("Is the person at least 18 years old and a member of the household?"). The speed of refusal was similar to the "no-incentive" introduction - 78 of the 91 persons that refused hung up either during the introduction or at the first questionnaire item.

The interviewers generally did not think the offer of money was an effective way to gain the confidence and cooperation of the respondent. It was felt that the offer of money mixed them in with telemarketing firms that try to trick individuals into staying on the

line and eventually sell them something or take their money. It should be noted, however, that these interviewers did not have any experience offering money in this context, they were highly experienced with doing government surveys that do not offer money.

3 Promised incentives at refusal conversion

A second experiment was conducted that tested the effectiveness of offering $25 to convert refusals at the screener level.

Design: This experiment was conducted with a total of 25 interviewers during the initial stages of interviewing for the NSAF. The interviewers were initially selected from among the pool of individuals who were asked to work refusal conversion for the study. The sample consisted of a random subsample of those households where someone had refused to complete the screening interview within 6 of the 11 states that were included in the NSAF supplemented state sample (Texas, New Jersey, Massachusetts, California, New York, Florida). The 6 states were selected because they exhibited the lowest response rates in the sample. Half of the 2,323 cases were assigned to an incentive condition and half were assigned to a "no-incentive" condition.

Cases were randomly assigned among the 25 interviewers. These initial cases were worked by the assigned interviewer until there was a callback assigned that crossed the interviewer's shift (e.g., call during the day, but the interviewer was scheduled for the evening). The average number of cases finalized by an interviewer was 79 (median 83), although there was quite a bit of variation around this average (standard deviation of 30). Several of the 25 interviewers dropped out of the study. Their cases were reassigned to other interviewers.

The interviews were conducted in a room that was isolated from the rest of the interviewing staff. There was some concern that if the experiment were being done in the proximity of normal production work, the offer of incentives would be overheard by other interviewers.

The 25 interviewers were assigned to one of two groups. During the first half of the experiment, one group offered respondents $25 if they agreed to complete the screening interview. The introduction used for this was:

> Hello, my name is (NAME) with the National Survey of America's Families. We are conducting this study for private foundations interested in education, health care, and other services in (STATE). The study has been endorsed by state governments concerned with how recent changes in policies affect

> people's lives. I am not asking for money. To show our appreciation for your help, we will send your household $25 for participating in the survey.

Interviewers were allowed to deviate from this script in ways that they thought appropriate. During the training, interviewers were instructed to use the monetary incentive at any point in the interview they felt it would have the most impact.

The other group of interviewers began with the "no-incentive" cases, using the introduction:

> Hello, my name is (NAME) with the National Survey of America's Families a study to see how recent changes in federal laws affect people's lives in your community. I am not asking for money - this is a study for private foundations on education, health care, and other services in the state of (STATE).

Once half of the cases had been worked, the two groups of interviewers switched conditions. Group 1 worked "no-incentive" cases, while Group 2 worked the "incentive" cases. Switching between conditions was done to minimize any confusion related to erroneously offering money for a no-incentive case.

Before the experiment began, the phone numbers were matched to a list of telephone directories. Those that matched and that had an address were sent a letter prior to the first refusal conversion call. For the cases with an incentive, the letter mentioned that $25 would be provided if the screener was completed.

Results: The basic results of the experiment are displayed in Table 1. These provide the Refusal Conversion Rates (completes/(refusals + completes)) by the incentive and no-incentive conditions. As can be seen, the offer of $25 had no effect on the overall rate of conversions, with 27.4% for the incentive condition and 28.5% for the no-incentive condition.

The prevailing hypothesis of why the incentive did not work is that respondents did not believe the offer of money. One way legitimacy of the offer might have been established, was through attempts to prenotify the respondent about the purpose of the call. Prenotification might provide respondents some opportunity to think about the substance of the study and, hopefully, decide that the call was part of a legitimate research effort (Dillman 1978).

There were two forms of prenotification that were used on the study. As mentioned above, letters were sent to those respondents for whom addresses could be found in public telephone directories. This did not increase the effectiveness of the incentive. For those that received an advance letter, the conversion rate for the incentive group was not

statistically different from those in the "no-incentive" group (data not shown).

A second form of prenotification was to leave a message on an answering machine. A message was left which contained information about the incentive the first time an interviewer had the opportunity to do so (no messages were left in subsequent calls). Table 1 displays data that partially supports the hypothesis that this increased the effectiveness of the incentive. In those instances where a message was left, the conversion rate was about 5% higher in the incentive condition ($t=1.28$; $p<.10$, one tailed test). When messages were not left, the opposite was the case ($t=1.78$; $p<.05$; one tailed test).

There are a number of differences between prenotification by the letter and the message machine. One is the difference in the mode of delivery of the message. The letter was mailed using addresses from published telephone lists. These addresses are not up to date. Consequently, some proportion of the letters are not actually delivered to the address to which the telephone number correspond (Traugott et al. 1997). Even if the letter is to the right address, it may never be opened or carefully read by the person who answers the phone. It may be thrown out as "junk" mail before opening or the person to whom it is addressed (and listed in the phone directory) may not be the person that the interviewer talks to. Finally, another set of letters were sent prior to the initial call made to the household when the study began. The letter for refusal conversion may have just been seen as providing more of the same information that had been rejected the first time. Consequently it may not have been taken very seriously or even read at all.

Alternatively, a message left on a machine could be heard by many people in the household. The person that plays the message is forced to listen to it in order to clear the machine. There is also evidence that leaving messages distinguishes calls from those of telemarketers, who generally do not leave messages (Xu et al. 1993, Tuckel and Shukers 1997). Unlike the letter, it provides an alternative method from the original prenotification letter to inform individuals about the study. It may not be seen, therefore, as providing redundant information and may be taken a bit more seriously.

One final difference are the types of respondents each method reaches. The letter only reaches those that have an address listed in the phone book, while the message machine only reaches those persons that have a machine and did not answer the phone when the interviewer first called.

When debriefed, interviewers were evenly split on the utility of the $25. Some interviewers felt that it provided an effective tool for converting respondents, while others provided the same feedback as in the first experiment - they felt that it made them sound like telemarketers trying to sell the respondent something. There were mixed feelings about when the incentive should be presented. Many interviewers felt that putting the

incentive at the very beginning of the introduction exacerbated problems with sounding like a telemarketer. They felt that it was first important to establish some credibility about the purpose of the call before mentioning the money.

This qualitative information seemed to indicate that effective use of a $25 incentive may be subject to the skills of the interviewer. Subtle differences in the delivery of the introduction may make a significant difference in whether or not the incentive is effectively used. One way this may have manifested itself would be a condition of the use of incentives and interviewer experience. As noted above, all of the interviewers were trained to do refusal conversion as part of the experiment. If interviewers became more confident and skilled at conversion as the field period progressed (as is normal), then the use of the incentive may have also improved as the field period progressed.

This would have shown up as a difference between interviewers who initially started with the incentive (Group 1) and those that offered it during the second half of the experiment (Group 2). This did not seem to be the case however. The differences in the incentive and no-incentive conditions were essentially the same, regardless of when the interviewers were using the incentive during the experiment (Group 1 vs. Group 2).

Use of the incentives may also be subject to more intrinsic qualities related to skills interviewers bring with them to the job. For example, there is quite a bit of variability across interviewers in their abilities to obtain high cooperation rates (Oksenberg and Cannell 1988, Collins et al. 1988). If these abilities are important for utilizing the $25 to convert refusers, then one might expect that the effect would vary by interviewer quality. To test this idea, we used two measures of quality, both of which utilized data obtained during NSAF field period. The first measure was the cooperation rate the interviewer achieved during the January to July NSAF field period.[1] These data are not consistent with the above hypothesis. The differences between the incentive and no-incentive conditions are not statistically different across the three groupings of interviewers (top third, middle third and bottom third).

A second way used to measure interviewer skill was to subset interviewers by whether or not they were kept on refusal conversion for the length of the NSAF field period. When looking at only those that were permanently kept on refusal conversion also does not support the hypothesis that better interviewers could more effectively use the incentive.

---

[1] The experiment was conducted during a 4 week period in March-April of 1997. The NSAF was administering initial screeners for much of 1997.

There is some evidence, however, that skill does interact with prenotification on the effects of incentives. Table 1 provides conversion rates for calls where a message was left by whether or not an interviewer was left on refusal conversion for the entire study. As can be seen, the significant effect of message machines is reduced to zero for those interviewers that were eventually not kept on refusal conversion, while the effect for the best converters is still significant ($t=1.78$; $p<.05$; one tailed test). One should note, of course, that these conclusions are not based on a large number of cases or a large number of interviewers (approximately 25). These conclusions, then, must only be considered tentative and need further confirmation.

4 Summary and conclusions

The experiments reported above did not find an effect of promised incentives for either the initial screening call or during refusal conversion. These results illustrate the relatively complicated role promised incentives may play in the context of an RDD survey. Overall, these results suggest that effective use of incentives varies by: (1) level of payment, (2) prenotification, (3) interviewer skill, and (4) method of presentation.

In conclusion, these data provide no evidence that promised incentives provide a magic bullet in the survey researcher's search for maintaining respectable response rates on RDD surveys. None of the data presented above find a consistent and substantial effect of promised incentives. If there is an effect, these data seem to indicate that it likely varies by a number of factors that are currently not well understood. Before implementing promised incentives on any particular RDD survey, therefore, it would probably be in the interest of both the sponsor and the survey organization to experiment with its use before implementing it on a large scale.

References

Church, A.H. (1993). Incentives in Mail Surveys: A Meta Analysis. Public Opinion Quarterly, 57, pp. 62-79

CMOR (1996). Respondent Cooperation and Industry Image Survey. Conducted by the Council for Marketing and Opinion Research. Technical Report

Collins, M., Sykes, W., Wilson, P. and N. Blackshaw (1988). Non-Response: The UK Experience. In: Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls, W.L. III and J. Waksberg (eds): Telephone Survey Methodology. New York: Wiley, pp. 213-232

Dillman, D. (1978). Mail and Telephone Surveys: The Total Design Method. New York: Wiley

Massey, J.T., O'Connor, D., Krotki, K. and K. Chandler (1997). An Investigation of Response Rates in Random Digit Dialing (RDD) Telephone Surveys. Paper presented at the 1997 Annual Meeting of the American Statistical Association, Anaheim, CA

Oksenberg, L. and C. Cannell (1988). Effects of Interviewer Vocal Characteristics on Nonresponse. In: Groves, R.M., Biemer, P.P., Lyberg, L.E., Massey, J.T., Nicholls, W.L. III and J. Waksberg (eds): Telephone Survey Methodology. New York: Wiley, pp. 257-272

Singer, E., Gebler, N., T. Raghunathan (1996). The Effect of Incentives on Response Rates in Face-to-Face, Telephone, and Mixed Mode Surveys. Paper presented at the Annual meeting of the American Association for Public Opinion Research, Salt Lake City, Utah

Strouse, R. and J. Hall (1977). Incentives in Population Based Health Surveys. Handouts from paper presented at the Annual meeting of the American Association for Public Opinion Research, Norfolk, VA, May 15-17, 1997

Traugott, M., Lepkowski, J. and P. Weiss (1997). An Investigation of Methods for Matching RDD Respondents with Contact Information for Validation Studies. Paper presented at the Annual meeting of the American Association for Public Opinion Research, Norfolk, VA, May 15-17, 1997

Tuckel, P. and T. Shukers (1997). The Effect of Different Introductions and Answering Machine Messages on Response Rates. Paper presented at the Annual meeting of the American Association for Public Opinion Research, Norfolk, VA, May 15-17, 1997

Xu, M., Bates, B. and Schweitzer, J.C. (1993). The Impact of Messages on Survey Participation in Answering Machine Households. Public Opinion Quarterly, 57, pp. 232-237

Does the Payment of Incentives Create Expectation Effects?[1]

ELEANOR SINGER, JOHN VAN HOEWYK AND MARY P. MAHER

Abstract: *Increasing use of incentive payments to survey respondents raises the threat of several unintended consequences, among them the creation of expectations for future payments and the possibility of a deterioration in the quality of response.*
The findings from the present study are somewhat reassuring with respect to both of these unintended outcomes. Although people who have received a monetary incentive in the past are significantly more likely to agree that "people should be paid for doing surveys like this", they are also more likely to participate in a subsequent survey, in spite of receiving no further payments. And respondents who received an incentive six months earlier are no more likely than those who received no incentive to refuse to answer (or to answer Don't Know to) a series of eighteen key questions on the survey. Furthermore, they are more likely than other respondents to express favorable attitudes toward the usefulness of "surveys like this". The generality of these findings, however, needs much further testing.

Keywords: *incentives, nonresponse, survey participation, quality of response, attitudes toward surveys*

1 Introduction

There is some evidence that the difficulty of obtaining cooperation with sample households in the United States and other developed countries is growing over time (de Heer and Israels 1992). In an effort to counter the increasing problem of noncooperation, survey organizations are offering incentives to respondents with increasing frequency, some at the outset of the survey, as has traditionally been done in mail surveys, and some only after the person has refused, in an attempt to convert the refusal. A meta analysis by Church (1993) identified those characteristics of incentives in mail surveys that are associated with greater effects on response rates: prepayment, cash, and larger (vs. smaller) payments. A subsequent examination of the use of incentives in telephone and face-to-face surveys (Singer et al. 1999) demonstrated the utility of incentives in those

---

[1] An earlier version of this paper was presented at the Ninth Nonresponse Workshop, Mannheim, Germany, September 24, 1997. We would like to thank our colleague, Mick P. Couper, for his helpful comments, and the Survey Research Center for financial support of this research. A revised version of the paper will appear in the Summer 1998 issue of *Public Opinion Quarterly*.

surveys, as well. There appear to be no deleterious effects of incentives on the quality of survey responses, though further research is needed in this area.

Despite these encouraging findings, concerns persist about possible unintended consequences of the use of incentives. Three can be mentioned here.

One is a concern that the use of differential incentives to convert refusals will be perceived as unfair by respondents, and will adversely affect their attitudes toward surveys and their willingness to cooperate (Groves et al. 1997, Singer, Groves and Corning 1998).

A second issue that has aroused some concern among survey researchers is whether the offer of an incentive is likely to replace intrinsic motivation to participate with extrinsic motivation, with a resulting decline in the quality of response. In a study of how framing an incentive affects response, Singer, Gebler, Van Hoewyk and Brown (1997) found suggestive evidence that students who respond to a survey request following receipt of a small gift perceive themselves as having responded primarily because of interest, whereas those who responded following receipt of a check for $10 perceive themselves as having responded primarily because of the incentive. An analogous finding is reported by Lengacher et al. (1995), who found that the usual measure of enjoyment of the interview is less predictive of Wave 2 participation among respondents who had received a substantial refusal conversion payment in Wave 1 than among those who had received no such incentive.

Still a third concern is that payment of incentives, especially at the outset of a study, may lead to expectations for such incentives in future surveys.

In an investigation of the effects of differential incentives in one wave on participation in a later wave of a longitudinal survey, Lengacher and her colleagues (1995) found no effect of a large refusal conversion payment on subsequent participation, compared to other Wave 1 reluctant respondents. In this paper, we provide further evidence bearing on two hypothesized unintended consequences of the payment of incentives - namely, increased expectations of being rewarded in the future as a result of having been rewarded in the past, and declines in the quality of response.

2 Methods

Because of concerns about declining cooperation with survey requests, the Survey Research Center at the University of Michigan decided in the fall of 1995 to begin monitoring the changing climate for survey research in the United States by adding five evaluative questions at the end of the Survey of Consumer Attitudes, a national telephone

survey administered monthly to a sample of roughly 500 respondents. Of these 500, 300 are newly selected RDD households and the remaining 200 are reinterviews of respondents first interviewed six months earlier. Because of concerns about their possible biasing effects, the five evaluative questions were asked of only the reinterviewed portion of the sample. The questions were added to the survey in January and February of 1996 and repeated in February and March of 1997 in order to measure changes in the climate for survey research over the approximately 12-month period. In principle, one could use changes in the responses to these questions as leading indicators of changes in the climate for survey research, and take proactive steps to counteract such changes.

The five monitoring questions, which were systematically rotated during their administration, are as follows:

1. If you had it to do over again, would you have agreed to do the interview or would you have refused?
   For each of the following, please tell me whether you agree strongly, agree somewhat, disagree somewhat, or disagree strongly:
2. Surveys like this one provide useful information for decision makers?
3. Surveys like this one are a waste of people's time.
4. People should get paid for doing surveys like this.

4a. How much should they get paid?

5. Everyone has a responsibility for answering surveys like this. (Do you agree strongly, agree somewhat, disagree somewhat, or disagree strongly?)

The initial response rate of those reinterviewed in January and February of 1996 was 67.0% six months earlier; their reinterview rate averaged 77.6%. Thus, the effective response rate of the 1996 sample is 52.0%. For the sample reinterviewed in February and March of 1997, the initial response rate averaged 65.3% and the reinterview rate, 76.8%; thus, the effective response rate for the 1997 sample was a slightly lower 50.2%. (The response rate excludes only nonsample cases from the denominator, and is thus a fairly conservative estimate. Noninterviews for reasons of illness or language, for example, are retained in the denominator.) The monitoring questions are, thus, asked primarily of cooperative respondents and of those who are easier to reach at home, and the comparisons discussed in this paper might be somewhat different if it had been possible to include nonrespondents.

Of particular importance for the present study, approximately half the respondents to the March 1997 survey had been promised a $5 incentive in return for their participation six months earlier, as part of a randomized experiment. In addition, a much smaller number of respondents in three of the four months had received refusal conversion payments of $20-$25 six months earlier. Thus, we are able to evaluate the effect of incentives on

subsequent attitudes and behavior, and to do so in the context of what was essentially a randomized experiment for the large majority of respondents.

3 Results

3.1 Changes in expectations about payment for survey participation

Responses to the five questions (and one subquestion) in 1996 and 1997 are shown in Table 1 (see page 5); they represent two cross-sectional measurements of attitudes toward surveys rather than answers by the same respondents at two different times.

Table 1 indicates that on three of the questions, no significant changes took place from one year to the next. Three others, however, show a significant change: Significantly more respondents (45.7% in 1997, compared to 29.7% in 1996) said that respondents should be paid for doing a survey like this, and the amount they stipulated showed a significant increase as well. In addition, a significantly higher proportion of respondents said, in 1997, that everyone has a responsibility for doing surveys like this.

We had anticipated changes in answers to the question about payment for two reasons. First, the practice of paying incentives to respondents in telephone and face-to-face surveys appears to be increasing. To the extent that awareness of this is diffusing throughout the population, a generalized expectation for payment may be developing. Second, as already noted, a large number of respondents to the March 1997 survey had themselves received a $5 initial incentive payment six months earlier, and a smaller number of respondents in three of the four months had received a refusal conversion payment of $20-$25. These respondents might have developed an expectation for payment based on their personal experience.

In order to separate the effect of these two reasons - generalized expectations vs. personal experience - we looked at the responses to the "People should get paid" question among those who had and those who had not been offered an initial incentive. The results are shown in Table 2 (see page 6). In both years, those who had received an incentive were much more likely to say that people should be paid than those who had not; the differences are significant in both years. Differences among people who did not receive any kind of incentive in either year are not significant. Thus, Table 2 demonstrates that the changed expectations apparent in Table 1 are due almost entirely to the responses of those who had themselves received an incentive - in other words, to personal experience rather than diffuse social norms.

Table 1: Responses to five evaluative questions, 1996 and 1997

	1996 (%)	1997 (%)
1. Do over?		
Yes	76.9	76.1
No	23.1	23.9
(N)	(407)	(406)
	n.s.	
2. Useful information?		
Agree strongly	34.6	39.7
All other	65.4	60.3
(N)	(405)	(401)
	n.s.	
3. Waste of time?		
Disagree strongly	28.6	30.4
Other	71.4	69.6
(N)	(402)	(395)
	n.s.	
4. Get paid?		
Agree	29.7	45.7
Disagree	70.3	54.3
(N)	(411)	(396)
	$\chi^2 = 22.08$, df =1, $p<.01$	
4a. How much?		
0-5	19.6	22.2
6-10	32.4	19.4
11-20	19.6	35.4
Over 20	28.4	22.9
(N)	(102)	(144)
	$\chi^2 = 10.01$, df =3, $p<.05$	
5. Responsibility?		
Yes	44.5	51.4
No	55.5	48.6
	$\chi^2 = 3.84$, df =1, $p<.05$	

Table 2: Response to "Everyone Should Get Paid...," by year and prior receipt of incentive

	Did Not Receive Incentive		Received Incentive	
	1996 (%)	1997 (%)	1996 (%)	1997 (%)
Yes, Should	26.0	31.0	91.3	65.9
No, Should not	74.0	69.0	8.7	34.1
(N)	(388)	(229)	(23)	(167)
	$\chi^2 = 1.77$, df =1, p=.18		$\chi^2 = 6.1$, df =1, p<.05	

The question of interest for this paper, however, is what interpretation should be placed on these responses. Should they, that is, be understood as reflecting changed expectations for the future, or rather as normative statements justifying past behavior? We tested these alternative interpretations by examining the cooperation rate of people to the March 1997 survey. Among people recontacted in March 1997, 139 had received an initial incentive six months earlier and 98 had not; 28 received a refusal conversion payment in March. If the earlier payment of an incentive led to (unmet) expectations for payment in the future, we would expect cooperation rates (without an incentive) in March to be lower among those who had received an initial incentive the preceding September than among those who had not. However, among those who had received an initial incentive in September and who were contacted by interviewers, 81.0% were reinterviewed without an additional incentive in March; among those who had received no incentive in September, the cooperation rate without an additional incentive in March was 66.3%. The difference between those receiving *no* incentive in September and those receiving $5 is significant; $X^2 = 5.43$, df=1, p<.05; those who had received a five dollar initial incentive six months earlier were significantly more likely to cooperate in March than those who had received no incentive. Thus, these data provide no evidence that responses to the question about whether or not respondents should get paid reflect expectations about future behavior, at least in the context of a request for a second interview by the same survey organization.

Because the increase in the percentage saying people should be paid seems to conflict with the increased tendency, also documented in Table 1, to say that everyone has a responsibility to participate in "surveys like this", we cross-tabulated the responses to these two questions in both years. The association is significant in neither year. In both years, people who agree that respondents should be paid for doing a survey like this are neither more nor less likely than those who disagree to say that everyone has a

responsibility for participating in a "survey like this". Nor were there any significant associations between the judgment that respondents should be paid and responses to any of the other monitoring questions.

3.2 Payment of incentives and data quality

We also examined the effect of the payment of incentives, whether offered at the outset of the study or as a refusal conversion payment, on the quality of responses, as measured by an index of nonresponse.[2] We found no effect of either refusal conversion payments or initial incentives on this index. For 1996, when only refusal conversion payments were offered, B=.90, S.E.=1.29, p=.49; for 1997 refusal conversion payments, B= -.03, S.E.=.62, p=.96; for initial incentives, B= -.24, S.E.=.66, p=.71.

3.3 Payment of incentives and attitudes toward surveys

In both years, the payment of incentives affected responses to two of the five evaluative questions in addition to whether or not respondents should get paid. In 1996, respondents who had six months earlier received refusal conversion payments were significantly more likely to say surveys are useful and to disagree that they are a waste of time. In 1997, respondents who had received any type of incentive six months earlier were significantly more likely to agree that surveys are useful and to say that everyone has a responsibility to take part in surveys like this. Thus, payment of incentives seems to lead to more favorable attitudes toward surveys, at least "surveys like this" - i.e., the one for which the respondent has received payment.

4 Summary and conclusions

Increasing use of incentive payments to survey respondents raises the threat of several unintended consequences, among them the creation of expectations for future payments and the possibility of a deterioration in the quality of response. Such deterioration may

---

[2] The index of nonresponse is the percentage of don't knows and no answers to 18 key questions in the Survey of Consumer Attitudes. The questions, whose tabulated responses appear in each SCA monthly report, indicate, among other things, respondents' assessment of their current and future family finances and income, the nation's business and employment conditions, and the government's role in affecting the country's economy.

come about either as a direct result of substituting external for internal motivation, or as a consequence of expectations for rewards that go unmet by the survey organization.

The findings from the present study are somewhat reassuring with respect to both of these unintended outcomes. Although people who have received a monetary incentive in the past are significantly more likely than those who have not to endorse the statement that "people should be paid for doing surveys like this", they are actually more likely to participate in a subsequent wave of the survey, even when they receive no further payments. Thus, it may be that respondents interpret the earlier payment as covering their current participation, as well. Respondents who received an incentive six months earlier are no more likely than those who received no incentive to refuse to answer (or to answer Don't Know to) a series of eighteen key questions on the survey. Furthermore, they are more likely than other respondents to express favorable attitudes toward the usefulness of "surveys like this".

The results of the present study are not, however, grounds for complacency. Payment of incentives is still a rather novel experience for respondents to telephone or personal interviews. Although few organizations would undertake a mail survey without enclosing some monetary or nonmonetary incentive with the advance letter or the questionnaire, interviewer-mediated surveys most commonly reserve incentives for refusal conversion efforts. Whether the absence of negative results observed in the present study will survive the more widespread use of incentives in such surveys remains very much an open question, one deserving continued research.

References

Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta Analysis. Public Opinion Quarterly 57, pp. 62-79

De Heer, W.F. and Israëls, A.Z. (1992). Nonresponse Trends in Europe. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Boston

Groves, R.M., Singer, E., Corning, A. and Bowers, A. (1997). The Effects on Survey Participation of Interview Length, Incentives, Differential Incentives, and Refusal Conversion: A Laboratory Approach. Paper presented at the meeting of the American Association for Public Opinion Research, Norfolk, VA

Lengacher, J.E., Sullivan, C.M., Couper, M.P. and Groves, R.M. (1995). Once Reluctant, Always Reluctant? Effects of Differential Incentives on Later Survey Participation in a Longitudinal Study. Paper presented at the May 1995 annual conference of the American Association for Public Opinion Research, Fort Lauderdale, FL

Singer, E., Gebler, N., Raghunathan, T., Van Hoewyk, J. and McGonagle, K. (1999). The Effects of Incentives on Response Rates in Face-to-Face and Telephone Surveys: Results from a Meta Analysis. Journal of Official Statistics

Singer, E., Gebler, N., Van Hoewyk, J. and Brown, J. (1997). Does $10 Equal $10? The Effect of Framing on the Impact of Incentives. Paper presented at the meeting of the American Association of Public Opinion Research, Norfolk, Va

Singer, E., Groves, R.M. and Corning, A. (1998). Incentives: The Effect of Perceived Equity on Willingness to Participate in a Future Survey. University of Michigan: Survey Research Center

Interviewer Opinions, Attitudes and Strategies Regarding Survey Participation and Their Effect on Response[1]

EDITH DE LEEUW, JOOP HOX, GER SNIJKERS AND WIM DE HEER[2]

Abstract: *Nonresponse is a threat to the validity of conclusions based on survey data. In general, two strategies are used to counteract this threat. The first strategy is to reduce the proportion of nonresponse as far as possible, the second is to statistically adjust for the remaining nonresponse. Interview surveys are still the norm for official statistics, social studies and market research in the Netherlands, and interviewers are an important factor in the battle against nonresponse. We focus on interviewers' opinions on nonresponse and their attitudes regarding the role of the interviewer in persuading potential respondents. In a special project at Statistics Netherlands the continuous survey on living conditions (POLS) was redesigned. During that study interviewer data were collected. It is shown that interviewer attitude and response rate are correlated. Interviewers with a positive attitude towards persuasion strategies attain a higher response rate. No differences between interviewers are found regarding self-reported 'door step' behaviour.*

Keywords: *interviewers, interviewers' attitudes, nonresponse, response rate, survey participation*

1 Introduction

Survey nonresponse is a growing problem in Western Europe and the US, and has been a source of concern for more than a decade (e.g., Steeh 1981; Goyder 1987; Groves 1989; Smith 1994; Schnell 1997; De Heer 1997). Nonresponse, and particularly the possibility of selective nonresponse, poses a serious threat to the validity of conclusions based on survey data. In general, two strategies are used to counteract this threat. The first strategy is to reduce the proportion of nonresponse as far as possible, the second strategy is to statistically adjust for the remaining nonresponse. We focus on the first strategy: reducing nonresponse.

---

[1] The views expressed are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

[2] The authors gratefully acknowledge the assistance of the POLS-interviewers and the field department of Statistics Netherlands. They thank the members of the 1995 Helsinki-nonresponse workshop for sharing their stimulating ideas and enthusiasm.

In interview surveys, be it by telephone or face-to-face, the interviewer is one of the most important means to improve response (Campanelli et al. 1997; Groves et al. 1992; Morton-Williams 1993). In mail surveys the researcher needs other approaches to reach an adequate response (e.g., Dillman 1978; De Leeuw and Hox 1988; Heberlein and Baumgartner 1978; Hippler and Seidel 1985). There is empirical evidence for considerable variation in response rates between interviewers (Lyberg and Lyberg 1991; Lyberg and Dean 1992). As a consequence several studies have addressed the role of the interviewer in nonresponse. There is little evidence that interviewer attributes, such as age and sex as such, influence response rates; however there is some evidence that interviewer experience positively influences response (cf. Groves and Fultz 1985; Couper and Groves 1992; De Leeuw and Hox 1996). What makes these experienced interviewers achieve higher response rates?

Interviewer behaviour at the moment that the first contact is made, has been the focus of recent projects of the International Workshop on Household Survey Nonresponse. Morton-Williams (1993) analyzes tape-recordings of survey introductions, and identifies successful strategies for obtaining respondent cooperation. Important factors were: appear trustworthy (e.g., always identify yourself immediately), appear friendly (e.g., smile, make a compliment), adapt to the situation at the doorstep, and react to the respondent. Interviewer-respondent interaction is also a central concept in the theoretical work of Groves, Cialdini and Couper (1992). Groves and Couper (1992, 1996) introduce the concepts of 'tailoring' and 'maintaining interaction' to emphasize the importance of flexible interviewer behaviour for a successful doorstep interaction.

A different perspective was introduced at the 1995 Helsinki workshop by Lehtonen (1996), who concentrated on interviewers' attitude towards persuasion strategies and the role of the interviewer. Those interviewers who have a strong belief in the importance of voluntariness of participation and feel negative towards strong persuasion strategies also had a higher probability of nonresponse.

Building on these two perspectives, we investigated the influence of the interviewer on survey response in a face-to-face interview. Main variables of interest were self-reported interviewer behaviour and interviewer attitude towards persuasion and the role of the interviewer.

2 Method

2.1 Data

During the months March to May 1996 a field experiment was carried out at Statistics Netherlands using mixed-mode computer assisted data collection. This experiment was part of a larger implementation study for the redesign of the continuous survey on living condition (POLS). During this field experiment interviewers had to perform special tasks such as registering of behaviour codes and using of special probes on the understanding of the survey questions asked. Twenty-two very experienced CAPI-interviewers were selected for this task. Selection criteria were among others, good social skills, research minded, a good response rate and good interviewer performance as evaluated by their supervisors (cf. De Leeuw et al. 1997). The interviewers worked in three separate geographical areas: Utrecht (highly urban), Eindhoven area (urban/rural) and North and Middle Limburg (highly rural). The interviewers were specially trained for this project. However, no special training in obtaining cooperation and doorstep interaction was given.

At the beginning of the training period all interviewers completed a questionnaire on nonresponse and cooperation in surveys[3]. This questionnaire contained general questions on experiences in gaining cooperation (e.g., the profile of refusers), questions on factual 'door step' behaviour during the introduction of the survey, and opinions and attitudes on persuasion strategies, and interviewer experience. This questionnaire was partly based on interviewer questionnaires of Campanelli et al. (1997) and Couper and Groves (1993). Included in the questionnaire was a general interviewer attitude scale developed by Lehtonen (1996). This scale contained five questions on interviewer attitude towards the role of the interviewer in persuading respondents to cooperate in surveys (for the precise wording of the attitude questions see the Appendix).

Based on the interviewer questionnaire four indices were constructed that described interviewer attitude and behaviour. The first index, the 'General Attitude Index' (GAI) based on the Lehtonen-items, indicates a positive attitude towards persuasion of respondents to cooperate. The second index 'Step Back' indicates the interviewers' opinion that it is more important to gain interest and leave a good impression at the first contact, than to push for a quick decision. If necessary one can step back and retry at a better moment. The third index 'Provide Interest Getting Information' indicates that the interviewers in their basic introduction usually give some information about the survey and the proceedings with the

[3] Two questionnaires were developed: one for face-to-face interviewers and one for telephone interviewers. An annotated English translation of these questionnaires is available from the first author (e-mail edithL@educ.uva.nl).

emphasis on the positive aspects. The fourth index indicates that interviewers use the 'Social Validation argument' by pointing out that most people participate and enjoy the survey.

Although we also asked questions on how they start the introduction (e.g., Introduce yourself, show ID, mention survey) we were not able to use these data, because of lack of variance between the interviewers. Interviewers reported that they *always* introduce themselves properly[4].

Furthermore, for each interviewer the following data were available through Statistics Netherlands: interviewer age, time worked at Statistics Netherlands, evaluation by two supervisors, number of instruction sessions attended since 1989. All interviewers were female. Also available were the total number of addresses handed to each interviewer and the total number of completed interviews of each interviewer for the Labour Force Survey 1995 and 1996. Based on these figures the proportion completed interviews was calculated as a conservative indicator of response.

2.2 Analysis

Our main research question was: Does interviewer attitude and behaviour influence response rate? An appropriate technique for binary and proportional response variables would be a logistic regression with response rate as dependent variable and interviewer attitude and behaviour as predictor variables. However, we did have interviewer level response data for the Labour Force Survey on two successive years. Therefore we used a multilevel logistic regression, which is an elegant approach to accommodate time-series data. The separate occasions define the lowest level, the interviewers the highest level. The model uses a logit link function to model the proportions, and second order Taylor expansion with penalized quasi likelihood estimation for the parameter estimates (for details see Goldstein 1995).

As predictor variables we used the interviewer background variables age, experience, number of instructions, supervisor's evaluation, and the four indices on interviewer attitude and behaviour (General Attitude, Step-back, Provide information, and Social-validation argument).

We also used 'district' as a covariate in the analysis. Interviewers worked in three districts ranging from highly urban to highly rural: Utrecht, Eindhoven, and North and Middle Limburg. These districts differed significantly in response. For practical reasons we could not use an interpenetrating design, and as a result the interviewers are nested within districts. We decided to correct for this statistically by using district as covariate in the analyses.

[4] More details about the frequencies of the answers and the index construction are available from the first author.

3 Results

3.1 Interviewer attitudes towards persuasion

We first looked into the answers on Lehtonen's attitude scale. Lehtonen (1996) investigated two groups of interviewers for the Finnish Health Survey, 120 professional interviewers of Statistics Finland and 93 public health nurses. The professional interviewers achieved a higher total response rate than the public health nurses: 88% for the first group versus 74% for the second. The two groups also differed markedly in their attitude towards the role of the interviewer in persuading potential respondents.

When we compare Lehtonen's data with our data (see Table 1), we see that the attitudes of professional interviewers of Statistics Netherlands are closer to those of the Finnish nurses, than to those of the interviewers of Statistics Finland.

Table 1: Proportion of interviewers agreeing or strongly agreeing with questions A-E Netherlands (N=22) and Finland (120+93; Lehtonen 1995)

Question	Stat Netherlands		Stat Finland	Finnish nurses
A. 'should always persuade'	36%	(8)	60%	25%
B. 'can be persuaded'	5%	(1)	29%	15%
C. 'respect privacy'	100%	(22)	96%	99%
D. 'refusal accepted'	32%	(7)	27%	82%
E. 'emphasize voluntary'	9%	(2)	35%	87%

The right to privacy (C) was generally acknowledged in all groups. However the other question associated with 'research ethics' (E) showed marked differences. A large majority of Finnish nurses (87%) agreed that voluntariness of participation should always be emphasized, while only 35% of the interviewers of Statistics Finland agree with that statement. The interviewers of Statistics Netherlands do not think that voluntariness should always be emphasized; only 9% agreed with the statement. In the words of one interviewer: if I emphasize voluntariness, I never get cooperation. This illustrates the difficult 'survey climate' in the Netherlands.

There are some marked differences on the questions measuring *attitudes towards persuading* reluctant respondents. The responses of the Dutch professional interviewers resemble more those of the Finnish nurses than those of the professional Finnish interviewers. The majority of the Dutch interviewers does NOT think that reluctant respondents should always be

persuaded (A), and almost no-one agreed that with enough effort even the most reluctant respondent can be persuaded (B). One therefore should expect that most Dutch interviewers would agree with question D: 'If a respondent is reluctant, refusal should be accepted.' However, only a small number (32%) agreed; on this question the Dutch interviewers were closer to the Finnish professionals. A possible explanation of this contradictory result can be found in the context of the study. The interviewer data were collected at the beginning of a field experiment where among other things 'refusal conversion' of soft refusals was attempted, and interviewers were extensively instructed on the usefulness and practicability of refusal conversion. This clearly introduces the possibility that they have been sensitized towards this attitude question, and in another context they might have given less 'acceptable' answers.

Lehtonen (1996) showed that the professional Finnish interviewers as group reached a decidedly higher response rate than the Finnish nurses. He also showed that interviewers who have a strong belief in the importance of voluntariness of participation and feel negatively towards strong persuasion strategies had a higher probability of nonresponse. In the next section we investigate the influence of interviewers' attitude and behaviour on the response in two successive Labour Force Surveys.

3.2 Predicting interviewer-level response

Using interviewers' background variables and indicators of interviewer attitude and behaviour as predictors we modelled interviewer-level response using a multilevel logistic regression. In this multivariate analysis the interviewer background variables age, experience (time worked at Statistics Netherlands), number of instructions followed, and supervisors' evaluation of the interviewer, were not good predictors of interviewer-level response rates. Also, two indicators of self-reported interviewer behaviour did not predict response rate significantly. These were the indicators concerned with the introductory stage (e.g., giving respondents interest getting information), and with the social validation tactic to persuade reluctant respondents (e.g., others do it too). The general attitude index, however, did significantly predict response rate. It should also be noted that the indicator 'step-back', which indicates the importance interviewers give to maintaining the interaction instead of pushing for a quick decision, showed a tendency in the predicted direction but did not reach statistical significance. The results are summarized in Table 2.

Table 2: Results multilevel logistic regression analysis
Dependent variable: response rate (proportion) for LFS 1995 and 1996

	MODEL				
	I	II	III	IV	V
PREDICTOR:					
intercept	.60	.65	.32	-.10	-.20
year = '96		-.11	-.11	-.11	-.11
district			.39	.34	n.a
(2 dummy vars)			.62	.53	n.a
GAI				.03	.06
STEPBACK				(.05)	(-.01)
variance between itr's	.11	.11	.04	.03	.08
lowest level variance	1	1	1	1	1

Note: Model V is the final model without district as covariate. Therefor n.a. (not applicable is mentioned for the control variables district.

Let us discuss the results in Table 2 in more detail. First of all model I, the intercept-only model, tells us that the mean response over 1995 and 1996 is .6 on a logit-scale. This translates to an average response of 65%. In model II the effect of year is modelled and we see that the response in 1996 was slightly lower than in 1995. We already knew this, of course. What is interesting, is the variance between interviewers. We see that 10% (0.10 = 0.11/(1+0.11)) of all variance in response is interviewer-variance. In the next steps we attempt to model this interviewer variance. Because interviewers are nested within districts, we add district as a covariate in model III, using dummy coding. We see that district explains 64% of all interviewer variance. In the following steps we added the other predictors; the only statistically significant predictor was the general attitude index (GAI). The indicator 'step-back' showed a tendency in the predicted direction but did not reach statistical significance. In model IV the effect of the GAI is added, we also show between parenthesis the nonsignificant effect of 'step-back'. From model IV we see that GAI explains an additional 9% of the interviewer variance. Together region of interviewing and interviewer-attitude explain a total of 73% of all variance.

We used region as a co-variate, but there is a slight danger of 'overcorrection'. Some real interviewer variance could be hidden in the region effect. If we look at differences between the interviewers in the three regions, we see that there is no statistical difference regarding age, attitudes, experience, and evaluation by supervisors. However, there is a statistically significant difference regarding the reported door step behaviour (more information given in

introduction and more social validation arguments in the rural area Limburg), and regarding the number of instructions (more instruction sessions attended by interviewers in the urban region Utrecht, probably reflecting that more different studies are done in that region.)

As a control we repeated the analyses, but without region as a covariate (model V). This did not change the pattern; only interviewer attitude did significantly predict interviewer-level response.

4 Conclusions and discussion

In a pilot study we showed that interviewer-level response rates can be predicted by the interviewers' general attitude towards the interviewer's role. Those interviewers who were more inclined to favor persuading the respondent received a higher response rate, while those interviewers who were more inclined to favor acceptance of refusals and not persuading the respondent, received a lower response rate. Furthermore, our sample of Dutch professional interviewers scored rather low on attitude towards persuasion, when compared with Finnish professional interviewers. It would be worthwhile to try to stimulate a more favorable attitude towards persuasion and persuasion strategies among Dutch professional interviewers.

Changing attitudes is not simple, but a useful start would be a short intensive course to motivate interviewers and teach them successful persuasion strategies (cf. Morton-Williams 1993; Campanelli et al. 1997; Snijkers, Hox and De Leeuw 1996). At the same time, supervisors and trainers should be motivated and through a special 'remedial teaching' course be taught the latest theoretical and empirical findings on response inducement and the role of the interviewer.

However, before starting such an intensive program, we will undertake to replicate this study using more interviewers and more regions in the Netherlands. The main reason for this is the small sample size in the study reported here. Twenty-two interviewers is not a large sample, and although the small sample size is partly compensated by having response rates based on large numbers of respondents, the power for detecting specific interviewer effects is still not very large.

Also it would be extremely fruitful to compare results internationally. Two research questions should be central in an international comparison: 'do interviewers in different countries have different attitudes towards the interviewer role?' and 'does interviewer attitude predict interviewer-level response rate internationally?'. At present a large replication study is being done in Belgium. There are also contacts with Couper (USA) and Campanelli and Sturgis (UK), who used comparable attitude and behavioral questions, to pool the data.

Researchers, who are interested in participating in this international project, are requested to contact the first author.

References

Campanelli, P., Sturgis, P. and Purdon, S. (1997). Can you hear me knocking: An investigation into the impact of interviewers on survey response rates. London: SCPR

Couper, M. and Groves, R. (1992). The role of the interviewer in survey participation. Survey Methodology, 18, pp. 263-277

Couper, M. and Groves, R. (1993). Interviewer questionnaire on issues in survey participation. August 1 Memorandum, Washington DC: US Bureau of the Census

De Heer, W. (1997). Results from the international survey on nonresponse. Paper presented at the 8th International Workshop on Household Survey Nonresponse, ZUMA, Mannheim, Germany

De Leeuw, E.D., Snijkers, G. and Hoezen, D. (1997). Kwalitatieve Veldproef Pols II: Werving, selectie en training van meta-interviewers [Procedures for recruiting, selecting, and training of special interviewers for pilot testing]. Heerlen: Statistics Netherlands. BPA#: H2286-97-GWM. In Dutch

De Leeuw, E.D. and Hox, J.J. (1988). The effect of response stimulating factors on response rates and data quality in mail surveys: A test of Dillman's TDM. Journal of Official Statistics, 4, pp. 241-249

De Leeuw, E.D. and Hox, J.J. (1996). The effect of the interviewer on the decision to cooperate in a survey of the elderly. In: S. Laaksonen (ed). International perspectives on nonresponse; proceedings of the sixth international workshop on household survey nonresponse. Helsinki: Statistics Finland

Dillman, D.A. (1978). Mail and telephone surveys: The Total Design Method. New York: Wiley

Goldstein, H. (1995). Multilevel statistical models. London: Arnold/New York: Halsted

Goyder, J. (1987). The silent minority; Nonrespondents on sample surveys. Cambridge: Policy Press

Groves, R.M. (1989). Survey errors and survey costs (chap 4). New York: Wiley

Groves, R.M. and Fultz, N.H. (1985). Gender effects among telephone interviewers in a survey of economic attitudes. Sociological Methods and Research, 14, pp. 31-52

Groves, R.M., Cialdini, R.B. and Couper, M.P. (1992) Understanding the decision to participate in a survey. Public Opinion Quarterly, 56, pp. 475-495

Groves, R.M. and Couper, M.P. (1992). Respondent-interviewer interactions in survey introductions. Paper presented at the 3th International Workshop on Household Survey Nonresponse, CBS, Voorburg, Holland

Groves, R.M. and Couper, M.P. (1996). Contact-level influences on cooperation in face-to-face surveys. Journal of Official Statistics, 12, pp. 63-83

Heberlein, T.A. and Baumgartner, R.M. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. American Sociological Review, 43, pp. 447-462

Hippler, H.-J. and Seidel, K. (1985). Schriftliche Befragung bei allgemeinen Bevölkerungsstichproben. ZUMA-Nachrichten, 1985

Lehtonen, R. (1996). Interviewer attitudes and unit nonresponse in two different interviewing schemes. In: S. Laaksonen (ed): International perspectives on nonresponse; proceedings of the sixth international workshop on household survey nonresponse. Helsinki: Statistics Finland

Lyberg, L. and Dean, P. (1992). Methods for reducing nonresponse rates: A review. Paper presented at the annual meeting of the American Association for Public Opinion Research. St. Petersburg, Fl.

Lyberg, I. and Lyberg, L. (1991). Nonresponse research at Statistics Sweden. Invited paper Annual Meeting of the American Statistical Association, Atlanta

Morton-Williams, J. (1993). Interviewer approaches. Aldershot: Darthmouth Pub.

Smith, T.W. (1994). Changes in nonresponse on the US General Social Survey, 1975-1994. Paper presented at the 5th International Workshop on Household Survey Nonresponse, Ottawa, Canada

Snijkers, G., Hox, J.J. and De Leeuw, E.D. (1996). Interviewers' tactics for fighting survey nonresponse. Paper presented at the 1996 International Conference on Social Science Methodology, University of Essex, Colchester, UK

Steeh, C.G. (1981). Trends in nonresponse rates, 1952-1979. Public Opinion Quarterly, 45. As reprinted in Singer, E. and Presser, S. (1989): Survey research methods, A reader. Chicago: University of Chicago Press

Schnell, R. (1997). Nonresponse in Bevölkerungsumfragen. Opladen: Leske+Budrich

Appendix

Full text of the General Interviewer Attitude-scale as used in this study

Responses were measured on a five-point Likert scale (1= strongly agree; 5= strongly disagree)

A Reluctant respondents should always be persuaded to participate

B With enough efforts, even the most reluctant respondent can be persuaded to participate

C An interviewer should respect the privacy of the respondent

D If a respondent is reluctant, a refusal should be accepted

E One should always emphasize the voluntary nature of participation

The Effect of Interviewer and Respondent Characteristics on Refusals in a Panel Survey

GEERT LOOSVELDT, ANN CARTON AND JAN PICKERY

Abstract: *In this paper data from an election panel survey are used. The results make clear that respondents who are more interested in politics are more likely to take part in the second interview of an election panel survey and that the initial contact for the second interview is extremely important for the group of poorly educated women. To evaluate the effect of the interviewer a multi level analysis was done. The results of this analysis show that the effect of the interviewers used in '91 on the refusals realized in '95 is more significant then the effect of the interviewers used in '95. This remarkable result stresses the importance of the experience of the first interview. Several interviewer characteristics were used to model the differences between the interviewers. Only the number of interviews done by an interviewer has a significant effect: more interviews result in more refusals.*

Keywords: *interviewer characteristics, respondent characteristics, election panel survey*

1 Introduction

Among others, refusals are an important component of nonresponse. A refusal is an active act of the respondent and can be considered as a crucial aspect of respondent behavior. There are several models to explain that kind of respondent behavior (Goyder 1987, Brehm 1993, Groves et al. 1992, Couper and Groves 1992, Morton-Williams 1993, Hox et al. 1996, Campanelli and Sturgis 1997). In these models respondent characteristics such as the "classic" respondent's background characteristics (age, gender, and education) and the respondent's attitude towards an interview are important components. Also interviewer characteristics (socio-demographics, experience, ...) are part of these models. The models make clear that, although the decision to participate or not is made by the respondent, also the interviewers have an important impact on that decision.

It is generally accepted that some interviewers are better at obtaining cooperation than others. Recently the role of the interviewer during the initial contact and the importance of the initial interaction between respondents and interviewers as an intervening variable

preceding the response decision is stressed (Campanelli and Sturgis 1997, Couper 1997, Maynard and Schaeffer 1997).

In the context of a panel research there is more than the interaction during the initial contact for the current interview. There is also the interviewer-respondent interaction during past interviews of the panel. An important general research question is: What is the effect of the respondent's experience with the interviews during previous waves of a panel on his or her decision to cooperate again. In accordance with the general models for nonresponse some respondent characteristics as well as some interviewer characteristics are used to answer that question.

2 Data

Data from the Belgian general election study are used. Only the interviews conducted in the Flemish region of Belgium are analyzed. Immediately after the national elections of November 1991, the Inter-university Centre for Political Opinion Research (ISPO, K.U.Leuven - Belgium) conducted the first wave of the election study in Flanders. The sample is representative of the Flemish population aged 18-75 years old. The second wave took place after the national elections of May 1995. During the second wave, 2580 respondents were used on the panel. The (non)response rate for the second wave is presented in Table 1.

Table 1: (Non)response rate for the second wave

	%
Completed Interview	68.3
Refusal	22.0
Non contacted	2.9
Ineligible	4.6
Other non-interview	2.2
n	2580

More than one-fifth (22%) of the panel respondents refused to participate a second time. It is clear that "refusal" is the most important reason (69.5% of the nonresponse) for the nonresponse in the second wave of this panel.

During the second wave 165 interviewers were used. 55 of them cooperated also during the first wave of the election panel. Only a small part of the panelrespondents were interviewed twice by the same interviewer.

3 The effect of respondent characteristics on refusals

3.1 The respondent's ability

An important respondent characteristic in a panel research is the **respondent's experience with the interview during the first wave**. His or her experience can be negative or positive and the general idea is that a negative experience with survey research evokes respondent resistance and increases nonresponse (DeMaio 1980, Nederhof 1987, Brehm 1993). We assume that an interview situation is an unpleasant or a negative experience when respondents have not enough cognitive and communicative abilities to execute the respondent's role adequately. For these respondents the tasks and the questions during the interview are too difficult. They experience a lot of difficulties in performing their role and on the basis of this kind of negative experience they are not motivated to participate in a second wave. We expect that refusers in a second wave of a panel are overrepresented in the group of respondents with insufficient abilities.

To measure **the respondent's ability** to perform their task we use three indicators. Two of these indicators are related to the respondent's behavior during the interview of the first wave namely the use of the "don't know" response category (18 questions) and the number of inconsistent answers (three pairs of statements). The third indicator is the interviewer's evaluation of the respondent's ability to answer the questions (a 6-point scale: very high to totally inadequate). The percentage of refusals in the second wave increases with the number of DK answers used during the interview of the first wave. The percentage of refusals is significantly higher if there is at least one inconsistent answer. We find the highest percentage of refusals for the category of respondents with - according to the interviewers - the lowest ability.

Table 2: Percentages of refusals by number of DK answers, inconsistent answers and ability to answer (bivariate associations)

Indicator	% Refusals	n
Number of DK answers		
0 - 2	21.4	1313
3 - 11	26.5	698
12 - 18	32.5	319
$\chi^2 = 18.0$, df= 2, p= 0.001		
Inconsistent answers		
none	23.1	1976
at least one	31.4	354
$\chi^2 = 11.0$, df= 1, p= 0.001		
Ability to answer		
(very) high	18.9	972
adequate	27.8	916
(very) poor tot. inadeq.	29.9	425
$\chi^2 = 28.5$, df= 2, p= 0.001		

In order to combine these three indicators into a typology of respondent's ability we performed a latent class analysis. It was possible to fit a latent class model with one latent variable with three classes (Table 3). In fact, we can consider these classes as a typology of the respondent's ability to answer the questions.

Table 3: Latent class analysis: latent class and conditional latent class probabilities

Type		Interviewer's evaluation of the respondent's ability			Use of DK			Inconsistent answers	
		high	adeq.	inadeq.	0 - 2	3 - 11	12 - 18	0	1+
1	.41	.81	.19	.00	.76	.24	.00	.90	.10
2	.48	.15	.65	.20	.45	.37	.18	.82	.18
3	.11	.00	.05	.95	.18	.25	.58	.84	.16
Chi-square= 2.36, df =3, p=0.5									

The first type contains respondents (41%) with enough cognitive skills to perform their task. The ability of everyone of this type is adequate or more than adequate. Most respondents of this type do not give inconsistent answers and the use of the DK answers is limited. Respondents of the second type (48%) have more problems. Their ability is lower, they give both more "DK answers" and inconsistent answers. Respondents of the third type (11%) experience a lot of problems during the interview. The ability of nearly all the respondents of this type is "poor" or less than "poor". They also have a high frequency of both "DK answers" and inconsistent answers. For these respondents the interview must be a rather unpleasant and negative experience. It certainly does not create a desire to participate a second time.

Given the description of the types, one can order the types from type 1 to type 3 according to their ability to perform their role as a respondent. As expected, there is an increase in the percentage of refusals over the three types.

Table 4: Percentage of refusals by respondent's ability to answer the questions

	Typology		
	Type 1	**Type 2**	**Type 3**
% refusals	19.0	27.7	31.8
n	942	1192	179
$\chi^2 = 27.168$, df= 2, p= 0.001			

The results support the basic idea that if respondents experience difficulties in fulfilling their role during the first interview because they do not have sufficient skills and abilities to perform their task, this will result in a non-participation for the second interview.

3.2 Political interest

The respondent's **level of interest in and knowledge of the topic of the questionnaire** is related to the respondent's abiltity to perform his or her task. If the respondents are interested in the topic of the questionnaire and they know a lot about it, then they do not experience much difficulty in performing their task. In the context of an election study knowledge and interest mean political knowledge and interest in political affairs. We expect the highest refusal rate for the respondents with the lowest political interest.

We used several indicators to measure political interest: a question about reading political news in the newspapers, one about discussing social and political issues among friends and a question about the respondent's evaluation of the complexity of politics (5-point scale were used): (almost) always to never). Table 5 shows that the less respondents read and discuss about the political news, the more they refuse the second interview of the panel. The relationship with reading political news in the newspaper is rather weak. We also see that the percentage of refusals is higher for respondents who (completely) agree with the statement that "politics are too complicated for people like me".

Table 5: Percentage refusals by reading political news in newspapers, discussing with friends and evaluation of the complexity of politics (bivariate associations)

Indicator	% Refusals	n
Reading newspapers		
(almost) always,often	21.1	579
now and then	24.1	523
seldom, never	26.1	1226
$\chi^2 = 5.2$, df= 2, p= 0.07		
Discussing		
(almost) always, often	21.5	452
now and then	21.4	916
seldom, never	28.5	960
$\chi^2 = 15.5$, df= 2, p= 0.001		
Complexity of politics		
(completely) agree	27.0	1236
neither agree nor disagree	22.9	468
(completely) disagree	20.1	617
$\chi^2 = 11.4$, df= 2, p= 0.003		

With the three indicators in Table 5 it was possible to construct a (rather) reliable scale (Cronbach coefficient Alpha = 0.66). On the basis of the first and the third quartile of this scale we distinguish three groups: high (below first quartile), moderate (between first and third quartile) and low (above third quartile) political interest. This more general measure of interest is strongly related to the percentage of refusals. Nearly one-fifth of the

respondents with high political interest and knowledge refuse a second interview; for respondents with low knowledge and interest we find that 29% refuse.

Table 6: Percentage refusals by respondent's political interest

	Political interest		
	low	**moderate**	**high**
% refusals	28.7	25.2	18.7
n	571	1165	594
$\chi^2 =$ 16.66, df= 2, p= 0.001			

The results are consistent with the expectation that respondents who are more interested in politics are more likely to take part in the second interview of an election panel survey.

3.3 Respondents' background characteristics

Respondents' background characteristics such as age, gender, education, occupational status, and place of residence are part of most models for nonresponse. A lot of these characteristics are indeed related to panel nonresponse (Kalton, Lepkowski, Montanari and Maligalig 1990, Rizzo et al. 1996). To make a selection, we performed a logistic regression analysis with only a small number of the most usual and relevant background characteristics as independent variables: education, age, gender, and occupational status. Table 7 presents the Wald (χ^2) statistics for each characteristic.

Table 7: Logistic regression with age, education, gender, and occupational status as independent variables

variable	df	chi-square	prob
Constant	1	308.10	0.000
Education	2	5.96	0.051
Age	2	2.17	0.338
Gender	1	7.54	0.006
Occupational status	2	5.60	0.061
Likelihood ratio= 43.03, df=44, p= 0.51			

There is a significant effect of gender. The effect of age is not significant. The effects of education and occupational status are comparable. However, education fits better with the emphasis on the respondent's ability to perform his or her task. For that reason, we decided to drop occupational status and to select education and gender as background characteristics. To describe the effect of these varaibles we use table 8. The refusal rate is higher for women than for men, and low educated respondents refuse more than high educated respondents. It is clear that the initial contact for the interview of the second wave is extremely important for the group of low educated women.

Table 8: Percentages refusals by gender and education

gender	men			women		
education	**low**	**medium**	**high**	**low**	**medium**	**high**
% refusals	24.4	23.3	13.3	30.4	23.1	22.7
n	631	266	269	634	277	233

3.4 Simultaneous analysis of the effect of the respondent's ability, political interest, and background characteristics

We have seen that the respondent's ability, political interest, gender, and education have an effect on the respondent's decision to participate in the second interview of a panel survey. However, these characteristics are interdependent. Gender and education for example are also strongly related to the respondent's ability (Gender: χ^2 =82.3, df=2, p= 0.001; Education: χ^2 = 616.9, df=4, p=0.001) and to political knowledge and interest (Gender: χ^2 = 152.3, df= 2, p= 0.001; Education: χ^2 = 343.6, df= 4, p= 0.001). Therefore it is necessary to evaluate the partial effects of these variables. The results of a logistic regression analysis show that controlling for education, gender, and type the effect of political interest is not significant.

Table 9: Logistic regression with education, gender, political interest, and respondent's ability

variable	df	chi-square	prob
Constant	1	234.27	0.000
Education	2	5.20	0.074
Gender	1	4.19	0.041
interest	2	2.00	0.367
ability	2	6.82	0.033
Likelihood ratio= 47.59 , df=41, p= 0.22			

4 The effect of interviewer characteristics on refusals

Until now, we have related respondent characteristics measured during the interview of the first wave of the panel to the decision to participate or not with the interview of the second wave. The analysis becomes more complex when interviewer characteristics are introduced in the analysis. There is the interviewer of the first wave (first-interviewer) and the interviewer of the second wave (second-interviewer). A panel respondent is not necessarily interviewed twice by the same interviewer. Most of the time the respondent is confronted with two different interviewers. Only 22,5 % of the panel respondents were interviewed by the same interviewer. Whether the respondent was interviewed by the same interviewer or not is a respondent characteristic. We expect that using the same interviewer will increase the refusal rate when the first interview was a negative experience (respondent type 3) and will decrease the refusal rate when the interview was a positive experience (respondent type 1). Table 10 shows indeed the highest percentage of refusals in the combination of respondent type 3 and the same interviewer. However, the same interviewer with respondent type 1 does not result in the lowest refusal rate. Controling for type of respondent, using the same interviewer results in a higher refusal rate. As a consequence, the advice in a panel survey about politics is not to ascribe the same respondent to the same interviewer. This advice is contrary to the idea that it is necessary to have the same interviewers return to the same respondent in order to maintain good response rates in longitudinal surveys (Campanelli and Sturgis 1997, pp. 2-9).

Table 10: Percentages refusals by the same interviewer or not and type of respondent

interviewer	the same			not the same		
respondent	type 1	type 2	type 3	type 1	type 2	type 3
% refusals	24.4	28.6	41.5	17.5	26.6	28.2
n	230	238	41	692	924	135

The interviewer of the first wave (first-interviewer) plays an important role in how the respondent experiences the first interview. The interviewer of the second wave (second-interviewer) is the one who must convince the respondent to cooperate again. We expect a first-interviewer effect as well as a second-interviewer effect on the decision to cooperate or not a second time. To analyse these interviewereffects we did a multilevel analysis with the result of the respondent's decision as dependent variable (0= refusal, 1= response). In a first analysis the interviewers of '91 were used to specify the second level in the analysis. Futhermore, no independent variables were used (model A). In this model the variance of the dependent variable (refusal or not) is divided into a respondent part (level 1) and an interviewer part (level 2). In a second analysis the same was done with the interviewers of '95 (model B). Both analysis show a significant interviewereffect. At the random part we see that the variance of the constant differs significantly from zero. This means that the differences in refusal rate between the interviewers are significant. The effect of the interviewers used in '91 on the refusals realized in '95 is more significant than the effect of the interviewers used in '95. This remarkable result stresses the importance of the experience of the first interview. However, to evaluate these interviewereffects, it is important to control for relevant respondent characteristics. In Model C and Model D these characteristics are incorporated in the analysis: two dummy variables for type of respondent (type1: 1= a respondent of type 1, 0 = not a respondent of type 1; type 3: 1 = a respondent of type 3, 0 = not a respondent of type 3), gender of the respondent (gender: 1= man, 0= women), same interviewer or not (same: 1= not the same, 0= the same) and two dummies for education (educ1: 1= low, 0= not low; educ3: 1= high, 0= not high). In both models the interviewereffects remain significant after controlling for these respondent characteristics. To evaluate the effect of these characteristics the parameters in the fixed part are used. These parameters can be interpreted as logistic regression coefficients. Although in the model C only the dummy variable type1 is significant at level .05, the same pattern is found in model C and model D. The response increases in type 1, for male respondents and when the second interview is done by another interviewer. There is no significant effect of education.

Table 11A: Results of some multilevel analysis (standard error between brackets)

Parameter	model A	model B	model C	model D
Fixed				
constant	1.211 (0.063)**	1.195 (0.059)**	0.802 (0.169)**	0.792 (0.164)**
type1			0.383 (0.126)**	0.323 (0.124)**
type3			-0.160 (0.185)	-0.137 (0.183)
same			0.255 (0.137)*	0.273 (0.132)**
gender			0.202 (0.105)*	0.231 (0.104)**
educ1			-0.101 (0.132)	-0.132 (0.131)
educ3			0.223 (0.169)	0.232 (0.165)
Random				
Interviewer level				
$\sigma^2_{constant}$	0.195 (0.067)**	0.141 (0.059)**	0.191 (0.067)**	0.126 (0.058)**

* $p < 0.1$
** $p < 0.05$

Another important question is whether some interviewer characteristics can be used to model the differences between the interviewers. Relevant interviewer characteristics are : gender of the interviewer (genderint: 1= men, 0= women); experience (experint: 0= low, 1= high); political knowledge (polknowint: 0= low, 1= high); political interest (polinint: 0= low, 1= high); number of interviews done by the interviewer (numint) and new interviewer in the election survey (newint 0= no, 1= yes). In model E the interviewers of '95 are used as second level and all these interviewer characteristics are added to the significant respondent characteristics of model D. Of these interviewer characteristics, only the number of interviews done by an interviewer has a significant effect: more interviews result in more refusals. Even the general interviewer effect is no longer significant. This can be explained by the fact that part of the interviewer variance is taken into account by the not significant effects of the interviewer characteristics. In this model the effect of the respondent characteristic 'Same' is also not significant. This is due to the high correlation with some of the interviewer characteristics.

In model F, all the characteristics are included with a significant effect in the previous analysis.

Table 11B: Results of some multilevel analysis (standard error between brackets)

Parameter	model E	model F
Fixed		
constant	0.669 (0.232)**	0.98 (0.194)**
type1	0.483 (0.109)**	0.476 (0.109)**
same	0.219 (0.185)	0.295 (0.130)**
gender	0.219 (0.103)**	0.224 (0.103)**
genderint	0.056 (0.121)	
experint	0.186 (0.117)	
polknowint	0.206 (0.168)	
polintint	0.083 (0.133)	
numint	-0. 020 (0.009)**	-0.019 (0.009)**
newint	0.049 (0.172)	
Random		
Interviewer level		
$\sigma^2_{constant}$	0.083 (0.052)	0.111 (0.056)**

5 Discussion

The results presented in this paper support the idea that, after controlling for some relevant respondent characteristics, some interviewers realize more refusals than others. With the interviewer characteristics used in this analysis it was only partially possible to explain these differences.

An important respondent characteristic related to the decision to participate or not, is the respondent's ability to perform his or her task during the first interview. For the respondent the interview should not be a confrontation with his or her inability to answer a lot of questions. For some respondents, long questionnaires with difficult questions create an unpleasant interview situation. Respondents with this kind of a negative

experience are less willing to cooperate again. Consequently, the researcher is responsible for the nonresponse.

It is clear that the respondent makes the decision to participte or not and that the interviewer can influence that decision. Respondent and interviewer characteristics related to the respondent's task (e.g. respondent's ability) and the task of the interviewer (e.g. number of interviews) are important to explain the outcome of the decision.

References

Brehm, J. (1993). The Phantom Respondents: Opinion Surveys and Political Representation. Ann Arbor: The University of Michigan Press

Burchell, B. and March, C. (1992). The effect of questionnaire length on survey response. Quality and Quantity, 26, pp. 233-244

Campanelli, P. and Sturgis, P. (1997a). Separating interviewer and area effects. In: Campanelli, P., Sturgis, P. and Purdon, S.: Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates. London: S.C.P.R.

Campanelli, P. and Sturgis, P. (1997b). Exploring the impact of survey doorstep interactions. In: Campanelli, P., Sturgis, P. and Purdon, S.: Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates. London: S.C.P.R.

Campanelli, P., Sturgis, P. and Purdon S. (1997). Can You Hear Me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates. London: S.C.P.R

Carton, A., Swyngedouw, M., Billiet, J. and Beerten, R. (1993). Source Book of the Voter's Study in Connection with the 1991 General Election. Leuven: Sociologisch Onderzoeksinstituut/ISPO

Clogg, C.C. and Goodman, L.A. (1984). Latent structure analysis of a set of multidimentional contingency tables. Journal of the American Statistical Association, 79, pp. 762-771

Couper, M. and Groves, R. (1992). The role of the interviewer in survey participation. Survey Methodology, 18, pp. 263-277

Couper, M. (1997). Survey Introductions and data Quality. Public Opinion Quarterly, 61, pp. 317-338

DeMaio, T. (1980). Refusals: who, where and why? Public Opinion Quarterly, 44, pp. 223-233

Goodman, L.A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. Biometrika, 60, pp. 179-192

Goyder, J. (1987). The Silent Minority: Nonrespondents on Sample Surveys. Boulder, CO: Westview

Groves, R. (1989). Survey Errors and Survey Cost. New York: John Wiley and Sons

Groves, R., Cialdini, R. and Couper, M. (1992). Understanding the decision to participate in a survey. Public Opinion Quarterly, 56, pp. 475-495

Groves, R. and Couper, M. (1992). Respondent-interviewer interactions in survey introductions. Paper presented at the Third International Workshop on Household Survey Nonresponse, Voorburg, the Netherlands

Hagenaars, J. (1990). Categorical Longitudinal Data: Log-Linear Panel, Trend and Cohort Analysis. Newbury Park, CA: Sage

Hagenaars, J. (1993). Loglinear Models with Latent Variables. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-094. Newbury Park, CA: Sage

Hox, J., de Leeuw, E. and Vorst, H. (1995). Survey participation as reasoned action: A behavioral paradigm for survey nonresponse? Bulletin de Méthodologie Sociologique, 48 , pp. 52-67

Kalton, G., Lepkowski, J., Montanari, G. and Maligalig, D. (1990). Characteristics of second wave nonrespondents in a panel survey. Proceedings of the Section on Survey Research Methods. Alexandria, VA: American Statistical Association

Loosveldt, G. (1995). The profile of the difficult-to-interview respondent. Bulletin de Méthodologie Sociologique, 48, pp. 68-81

Maynard, D. and Schaeffer, C. (1997). Keeping the Gate. Declinations of the Request to Participate in a Telephone Survey Interview. Sociological Methods and Research, 26, pp. 34-79

McCutcheon, A. (1987). Latent Class Analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-064. Newbury Park, CA: Sage

Morton-Williams, J. (1993). Interviewer Approaches. Aldershot: Darthmouth Publishing Company Limited

Nederhof, A. (1987). When neutrality is negative, pleasantness of most recent survey experience and nonresponse. Quality and Quantity, 21, pp. 425-432

Rizzo, L., Kalton, G. and Brick, M. (1996) A comparison of some weighting adjustment methods for panel nonresponse. Survey Methodology, 22, pp. 43-53

Taylor, B., Heath, A. and Lynn, P. (1996). The nature of attrition in the British election Panel study. Paper presented at the Fourth International Social Science Methodology Conference, Essex 1996

Longitudinal Nonresponse in the Current Population Survey (CPS)[1]

BRIAN A. HARRIS-KOJETIN AND CLYDE TUCKER

Abstract: *A longitudinal database was created consisting of seven complete "cohorts" who were in the CPS sample all eight months over nearly a 2 year period. Comparisons were made among complete respondents and partial respondents. Households that were complete respondents were relatively more likely than households that were partial respondents to be located in rural areas and in any region of the country except the West, and to have more persons living there, and to be occupied by owners. Members of complete respondent households were relatively more likely than members of partial respondent households to be a married couple, children, white, non-Hispanic, over 65 years of age, and were relatively more likely to answer a question on total family income. Persons in partial respondent households had a higher level of employment in the first four months-in-sample (MIS) and a higher unemployment rate in MIS 1 than persons in complete respondent households.*

Keywords: *panel nonresponse.*

1 Introduction

It is a fact of modern life that organizations that survey households, individuals, or business establishments will not obtain complete cooperation from everyone they solicit. Survey nonresponse occurs at several different levels, with individuals or households not responding at all (unit nonresponse), not responding during one wave or panel of a longitudinal survey (wave/panel nonresponse), or simply omitting certain survey items (item nonresponse). Ideally, one would hope that nonrespondents are a random cross-section of the sample, reflecting the same demographic, geographic, and economic groups as respondents. If this is the case, then one need not be concerned about obtaining biased results from a survey in which there was some degree of nonresponse. However, it is typically the case that nonrespondents differ from respondents on these characteristics (for reviews see Goyder 1987, Groves 1989). Therefore, even surveys with high response rates may have some degree of bias in their results to the extent that nonrespondents differ from

[1] The views expressed in the paper are those of the authors and do not necessarily represent those of the U.S. Bureau of Labor Statistics.

the respondents.

Of course, it is the nature of nonresponse that we almost never know exactly what we wanted to know about survey nonrespondents. Longitudinal surveys or those with a panel component can provide insights into some of the characteristics of some nonrespondents who provide information on a previous or subsequent panel or wave. Longitudinal surveys spanning several years are particularly concerned with bias entering due to attrition (e.g., Zabel 1994), whereas shorter panel surveys may be more likely to have people missing a wave and then returning. With panel nonresponse, if one can assume that the characteristic of interest was stable during the entire time then little is lost by missing one observation in time, because the data can be drawn from another point in time. However, most panel surveys are primarily concerned with characteristics that are (at least potentially) changing over time and, therefore, observations from other points in time provide a clue as to what the actual response would have been, but may not be sufficient to infer accurately the desired characteristics.

1.1 The present study

The focus of the present study is on panel nonresponse in the Current Population Survey (CPS). We created a longitudinal database file for the CPS consisting of a number of "cohorts" who were in the CPS sample for the full 8 months over a 16 month period. In this paper, we examine patterns of longitudinal nonresponse and the characteristics of nonrespondents using information obtained in months in which there was an interview. Previous research on panel nonresponse to the CPS has been severely hampered by the nature of the control card data and the difficulty of tracking individuals and households accurately across time. For example, one prior study was able to match only 50% of the cases across all 8 panels of the CPS (Dippo et al. 1992). Also, it was not possible to distinguish movers from other unmatched households. Beginning in January, 1994 the redesigned questionnaire of the CPS was implemented with new computer-assisted data collection methods which included among several improvements in data quality, the creation of a longitudinal identification number that would allow better longitudinal record-level matching of CPS data.

2 Design

The CPS is the monthly household labor force survey for the United States conducted by the U.S. Census Bureau for the U.S. Bureau of Labor Statistics. Approximately 60,000 eligible households are sampled each month in a two-stage clustered design. Households selected for sample are interviewed for 4 consecutive months, are not interviewed 8

months, and then are interviewed again for 4 consecutive months. Furthermore, in any given month, one eighth of the sample is composed of households participating for the first time (month-in-sample 1 (MIS 1)), one eighth the second time (MIS 2), etc.

Data for the present investigation were drawn from the first seven cohorts that completed all of the eight months-in-sample (MIS 1-8) since their initial selection during or after January 1994. The first cohort began in January 1994 and finished in April, 1995, and the last cohort began in July, 1994 and completed in October, 1995. This data set includes a total of 75,854 households.

2.1 Matching procedure

Household level matching. The longitudinal matching involved several levels of analysis because nonresponse is at the household level whereas survey response is at the person level. First, a longitudinal household level data file was created. A total of 60% of the households matched across all eight MIS. The majority of the non-matching cases were out of scope at least one month (e.g., demolished, nonresidential use, vacant, etc.) or were movers. Of the 45,395 households that matched across all eight MIS, 35,018 had some combination of interviews or Type A nonresponses (i.e., refusals, noncontacts, other noninterviews) for all eight MIS and are the focus of analyses in this paper. Of those, 28,724 households had interviews only for all eight MIS, 699 households had nonresponses only for all eight MIS, and 5595 households had at least one interview and at least one nonresponse during their eight MIS.

Person level matching. For the present paper, only those households with interviews all eight months or at least one interview and at least one nonresponse were selected for examination at the person level. A total of 99,639 persons were in those households, with 84,665 persons in households that were interviews all eight MIS, and 14,974 persons in households that were in all eight MIS and had at least one interview and at least one nonresponse.

2.2 Measures

The measures included in the present investigation covered several different levels of analysis. Geographic area information available included region of the country, urbanicity, and whether the area had a high percentage of people living below the poverty line. These data were available for all households regardless of their participation in the CPS. Additional data at the household and person levels was available only for those households with at least one interview. Household-level information included whether the

housing unit was occupied by the owner or a renter, the total number of persons in the household, the type of family (e.g., couple with children, single parent, single adult), and total family income. Person-level information included demographic characteristics such as age, sex, race, ethnicity, education level, marital status, and household members' relationships with each other. For the present study, the only substantive survey information examined was labor force classification (employed, unemployed, and not in the labor force), which is the primary focus of the CPS.

2.3 Analysis plan

Analyses comparing respondent and nonrespondent cases were carried out at both the household and person level. Analyses were conducted to compare characteristics of households that were interviewed all eight months (complete respondents) with those interviewed at least once with one or more Type A nonresponses (partial respondents) on geographic, household, and person characteristics that were available only for cases which had been interviewed at least once. The geographic analyses also included households that were nonrespondents all eight MIS (complete nonrespondents). All available geographic, household, and person-level information was entered into a logistic regression to compare, as completely as possible, persons who were complete respondents to partial respondents. Initial comparisons of the labor force status of complete and partial respondents were also made. All analyses were conducted using unweighted data.[2]

3 Results

3.1 Longitudinal response patterns

Approximately 82% of the households were interviewed all eight MIS. The most common pattern involving Type A nonresponse was for the first month to be a nonresponse and all subsequent months to be interviews (2.8% of all the households) and the second most common pattern was nonresponse during all eight MIS (2.0% of all households). The other most common patterns most frequently involved one nonresponse and seven interviews. Patterns that showed attrition from the sample made up a total of 3.1% of the households.

[2] All significance tests reported in this paper use unweighted data and do not take into account the complex sampling design of the CPS.

3.2 Household level analyses

3.2.1 Geographic area characteristics

To examine whether there were significant differences in geographic area characteristics among households that were complete respondents, partial respondents, and complete nonrespondents, distributions of characteristics for these three groups were compared using chi-square tests. Significant differences were revealed between these three groups for region of the country, urbanicity, and poverty of the area (all p's $< .01$). The largest differences were between the complete respondents and the complete nonrespondents. There were a relatively greater proportion of households that were complete nonrespondents in the Northeast, West, and South, in urban areas (especially central cities), and in low poverty areas, while a relatively greater proportion of households were complete respondents in the Midwest, in rural areas, and in high poverty areas.

3.2.2 Household characteristics

Comparisons were also made at the household level for characteristics of the households that were obtained in the interview, but this information was not obtained for households that were complete nonrespondents. Partial respondent households were relatively more likely than complete respondent households to be occupied by renters, have fewer persons living there (1 or 2), and to be headed by a single parent or single person (all $p's < .01$). In addition, members of partial respondent households were relatively less likely to answer a question on total family income, while members of complete respondent households were relatively more likely to provide family income data or have a higher family income than partial respondent households.

3.3 Person level analyses

3.3.1 Demographic characteristics

There were also significant differences between complete and partial respondents on a variety of demographic characteristics. Households that were partial respondents were relatively more likely to contain persons who were nonwhite, Hispanic, who had graduated college, and who were 25-34 years of age; while households that were complete respondents were relatively more likely to contain persons who were white, non-Hispanic, under 19 years old, and who have never been married.

3.3.2 Multivariate analyses of demographic, household, and geographic variables

To take into account all of the geographic, household, and demographic characteristics in comparing complete respondents to partial respondents, logistic regression analyses were conducted entering all of the variables into the equation in order to maximally distinguish between these two groups. The overall difference between complete respondent households and partial respondent households was significant, $p < .01$, and the odds ratios are given in Table 1. The results are essentially the same as the univariate results reported above with most of the characteristics independently predicting interview status, except the effects of poverty and marital status are no longer significant. Although the overall equation was significant, even all together these variables were able to account for only a small percentage of the variance between these two groups, $R_L^2 = .040$.[3]

3.3.3 Labor force classification

Unlike most of the demographic characteristics noted above, a person's labor force classification may change from month-to-month during the time that they are in the CPS sample; therefore, one cannot determine with certainty a person's labor force classification during the months the person was a nonrespondent. Accordingly, comparisons of labor force classification were made for each MIS separately between the complete respondents and partial respondents. Only persons who were interviewed in a particular MIS and had a labor force classification are included. Therefore, the persons who were partial respondents and who had a labor force status in MIS 1 did not respond at some later month, while the persons who were partial respondents who had a labor force status in MIS 8 did not respond during some previous month. The distribution of labor force classifications for persons in complete and partial respondent households showed significant differences (p's $< .05$) in MIS 1-4, demonstrated marginally significant differences for MIS 5 & 6 (p's $< .10$), but were not significantly different in MIS 7 & 8. Persons in partial respondent households were relatively more likely to be employed during MIS 1-4 than persons in complete respondent households. The unemployment rate for persons in partial respondent households was also significantly higher ($p < .05$) than the rate for persons in complete respondent households during MIS 1, but the differences in the other MIS were not significant.

---

3 $$R_L^2 = \frac{-2\log L_0 - (-2\log L_1)}{-2\log L_0}$$

Table 1: Results of logistic regression predicting partial respondents (compared to complete respondents) with geographic area, household, and demographic characteristics

Variable	Odds Ratio
Region (West omitted)	
Northeast Region	.93*
Southern Region	.90**
Midwest Region	.66**
Urbanicity (rural omitted)	
Central City	1.25**
Balance of MSA	1.08**
Poverty area	1.04
Own Residence	.90**
Household Size (5+ omitted)	
1 Person Household	1.96**
2 Person Household	1.79**
3 Person Household	1.42**
4 Person Household	1.21**
Family type (singles omitted)	
married couple family	.90*
single parent	1.02
Refused/DK income	4.06**
Sex (1=male, 2=female)	.97
White/nonwhite	.75**
Hispanic/nonhispanic	.90*
Age (65 + omitted)	
16-19 years	2.41**
20-24 years	2.06**
25-34 years	1.93**
35-44 years	1.80**
45-54 years	1.62**
55-64 years	1.43**
Education (College degree + omitted)	
Less than High School degree	.85**
H.S. diploma	.95
some college	.91**
Marital status (never married omitted)	
Married	1.00
Widowed/Divorced/Separated	.97
Relationship (non-relative omitted)	
Reference person	1.07
spouse of reference person	1.06
child	.72**
other relative	.99

** $p < .01$, * $p < .05$

4 Discussion

4.1 Summary of findings

The present study found a variety of differences in geographic, household, and person characteristics between households that were complete respondents and those that were only partial respondents, i.e., households that had interviews and at least one nonresponse. Specifically, households that were complete respondents were relatively more likely than households that were partial respondents to be located in rural areas and in any region of the country except the West, and to have more persons living there, and to be occupied by owners. Members of complete respondent households were relatively more likely than members of partial respondent households to be a married couple, children, white, non-Hispanic, over 65 years of age, and were relatively more likely to answer a question on total family income. Although there were many significant differences between complete and partial respondent households on these characteristics, they actually accounted for a rather small proportion of the variance between the groups.

Differences in labor force classification were also demonstrated between persons in complete and partial respondent households. Persons in partial respondent households had a higher level of employment in the first four MIS and a higher unemployment rate in MIS 1 than persons in complete respondent households. Furthermore, the differences between these two groups were the most extreme at MIS 1 and decayed over the months in sample until there were no differences in MIS 7 & 8.

4.2 Implications

The evidence that households that are nonrespondents in some MIS are more likely to be from distinguishable geographic, demographic, economic groups leads to concern. The present results have possible implications for post-survey nonresponse adjustment procedures as well as field procedures. Current CPS nonresponse adjustment procedures do not utilize prior information obtained from the household. Of the variables examined in the present investigation, response adjustment utilizes only MSA status. Clearly, now that accurate longitudinal data on households and persons is easily accessible, there are many more possibilities for enhanced and refined nonresponse adjustment and imputation. The current results suggest efforts in this area may be quite worthwhile.

Very little information is currently obtained in the field about nonrespondents. Changes in field procedures that put more emphasis on obtaining more characteristics of the household and persons living there is essential in furthering research and understanding of nonresponse. The households that were complete nonrespondents all 8 MIS appeared to be distinctly different from households that were complete respondents as well as

households that were partial respondents. However, there were very few available characteristics on which comparisons of these households could be made. To the extent that complete nonrespondents are distinctly different from partial nonrespondents, our ability to make more accurate nonresponse adjustments will be limited by the information that is obtained in the field.

Further analyses need to be conducted on the labor force status of persons in partial respondent households. There is some indication of differences in labor force status of persons who become nonrespondents in later MIS (see also Tucker and Harris-Kojetin, pp. 45-54, this volume). Whether this leads to a bias in labor force estimates requires further investigation.

4.3 Limitations

Because the data used in this study were from a limited time period, the impacts of seasonal effects and real economic change could not be assessed. These factors also may have influenced our findings. All analyses presented in this paper have been based on unweighted data. Therefore, conclusions pertaining to rates and percentages do not necessarily represent the entire U.S. civilian noninstitutional population.

4.4 Future directions

The present findings should be considered preliminary as this is the first in what we expect will be an ongoing line of research on the longitudinal aspects of the CPS. Further analyses will also examine the demographic characteristics associated with different patterns of nonresponse and reasons for nonresponse (refusal, noncontact, other). The use of unweighted data is a good starting point in this type of research, however, to make any geographic conclusions or any others that would reflect the nation, base weights, which reflect the probability of selection, need to be used. The baseweighted data has not been adjusted for nonresponse and thus would not exaggerate nonresponse results (whereas final weighted data would). Evaluations of the current weighting procedures are planned, and additional analyses of the effect of methodological factors such as mode of interview on response are also planned.

References

Dippo, C. S., Silberstein, A. and Rope, D. (1992). An Examination of Longitudinal Nonresponse in the CE and CPS. Paper presented at the Third International Workshop on Household Survey Nonresponse, Voorburg, The Netherlands

Goyder (1987). The Silent Minority: Nonrespondents in Sample Surveys. Boulder, CO: Westview Press

Groves, R. M. (1989). Survey Errors and Survey Costs. New York: Wiley

Tucker, C. and Harris-Kojetin, B. (1997). The impact of nonresponse on the unemployment rate in the Current Population Survey. In: Koch, A. and Porst, R. (eds.). Nonresponse in Survey Research. Proceedings of the Eighth International Workshop on Household Survey Nonresponse 24-26 September 1997. Mannheim: ZUMA, pp. 45-54

Zabel, J. E. (1994). A Comparison of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation. Unpublished manuscript. U. S. Census Bureau, Washington, D.C.

A Bootstrap Strategy for the Detection of a Panel Attrition Bias in a Household Panel with an Application to the German Socio–Economic Panel (GSOEP)

ULRICH RENDTEL AND FELIX BÜCHEL

Abstract: *A bootstrap strategy for detecting non–ignorable panel attrition is proposed. The strategy is based on the difference of the original estimate and an estimate that is obtained by reducing the original sample by a second attrition experiment. The attrition propensities are estimated from previous wave information and field work information of the current panel wave. The routine may be used to estimate the bias due to panel attrition. The bootstrap routine is applied to two income estimates with data from the first 8 waves of the GSOEP.*

Keywords: *attrition bias, bootstrap, household panel*

1 Introduction

Panel surveys are plagued by the successive attrition of people, who refuse to continue to participate or who are lost because of problems in recontacting them in the next wave of the panel. Such losses not only reduce the sample size, they may also bias estimates based on the remaining sample. The panel attrition is ignorable if conditioning on the participation does not affect the distribution of interest $f(Y \mid X)$; which means we have $f(Y \mid X, S = 1) = f(Y \mid X, S = 0)$, where $S = 1$ indicates participation and $S = 0$ indicates attrition. A selection rule is called non–ignorable, if the these two distributions differ. A recent survey on non–ignorable panel attrition was presented by Verbeek and Nijman (1996).

The main difficulty in the treatment of attrition is the lack of knowledge about $f(Y \mid X, S = 0)$. In fact, without any knowledge about $f(Y \mid X, S = 0)$ the problem is not solvable on the basis of the observed data alone. In a panel, however, there exists a lot of information about attriters. The information arises from the characteristics observed in the previous panel waves. Also the field work of the present panel wave produces relevant information; namely, whether the person or the household has moved since the last interview or whether the

household has split up into two separate households. In the case of the German Socio-Economic Panel (GSOEP), an ongoing household panel started in 1984 (cf. Wagner et al. 1993a + b), field-related characteristics turned out to be the most relevant indicators for explaining drop-out during the panel (cf. Rendtel 1990, 1995). For a panel study which is based on face to face interviews, this is a plausible finding.

Such information may be used to estimate drop-out probabilities for panel members. The attrition probabilities help synthesize our knowledge about the attrition process using information on attriters and non-attriters. In order to answer the question of whether panel attrition affects the estimation of the model of interest, it is crucial to exploit the relationship between the variables of the model of interest and the characteristics from field work.

The attrition bias is defined here as follows: let $\hat{\theta}$ be an estimator of some parameter θ that characterizes the distribution $f(Y \mid X, \theta)$. For each unit we observe covariates Z that predict attrition, which is indicated by S. Let $\hat{\theta} = \hat{\theta}(X, Y)$ be the estimate of θ on the basis of the sample in the absence of attrition. Of course, we cannot observe $\hat{\theta}$. Denote by $\tilde{\theta} = \hat{\theta}(\tilde{X}, \tilde{Y})$ the estimate of θ on the basis of the observed sample $\tilde{X}, \tilde{Y}$ after attrition. We use here the following definition of an attrition bias of $\hat{\theta}$:

$$\text{bias}(\hat{\theta}) = E_S(\hat{\theta}(X, Y) - \hat{\theta}(\tilde{X}, \tilde{Y}) \mid X, Y, Z)$$

Hence, the bias($\hat{\theta}$) is the expected difference of the estimation results with and without attrition. The expectation is with respect to S conditional on the value of X , Y and Z. The effect of an attrition rule depends strongly on the marginal distribution of the model variables and the attrition predictors in the sample before attrition, which is reflected by conditioning on X,Y and Z .

The basic idea of the bootstrap strategy presented here is to resample from the observed sample $\tilde{X}, \tilde{Y}$ B replicates X^*, Y^*. This is done by Poisson sampling, which means that each unit is resampled according to its propensity of non-attrition. The use of the Poisson sampling is different from the standard Bootstrap routines which use sampling with replacement(Efron and Tibshirani 1993). The Poisson sampling results in a second, artificial attrition experiment on those units that survived the first attrition.

The bootstrap strategy presented here is also different from Efron's (1994) "Full-mechanism Bootstrap", since we do not try to reconstruct the distribution of X and Y before attrition.

Under some regularity conditions we may assume that the second attrition experiment produces a similar bias as the first attrition experiment. In this case we can check whether the distribution of the $\hat{\theta}^* = \hat{\theta}(X^*, Y^*)$ is centered around

$\widetilde{\theta} = \hat{\theta}(\tilde{X}, \tilde{Y})$. The average of the differences $\widetilde{\theta} - \hat{\theta}^*$ is taken as an estimate of bias$(\hat{\theta})$.

The bootstrap routine is applied to data from the first 8 waves of the GSOEP. The example deals with males who experienced a period of unemployment of at least 12 months. We want to know whether a joint analysis of incomes before and after unemployment is affected by panel attrition.

2 Selection on observable and unobservable variables

A necessary condition for the bootstrap strategy to work is the selection on observable variables. The distinction between selection on observable and selection on unobservable variables is discussed here within the the framework of the standard econometric model for selection. Here, we have a regression equation:

$$Y_{i,t} = X'_{i,t}\beta + \epsilon_{i,t} \tag{1}$$

which is observed if $S_{i,t} = 1$. Here i indicates units (individuals or households) and t indicates the panel wave. The model for $S_{i,t}$ is the stochastic censoring model for a latent response propensity $S^*_{i,t}$:

$$S^*_{i,t} = Z'_{i,t}\gamma + \delta_{i,t} \tag{2}$$

and

$$S_{i,t} = \begin{cases} 1 & \text{if } S^*_{i,t} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The standard orthogonality assumptions are: $X_{i,t} \perp \epsilon_{i,t}$, $Z_{i,t} \perp \delta_{i,t}$. Selection on unobservables holds, if:

$$E(\epsilon_{i,t} Z_{i,t} \mid X_{i,t}) = 0 \text{ and } E(\epsilon_{i,t}\delta_{i,t} \mid X_{i,t}) \neq 0 \tag{4}$$

This case applies if all Z–variables that are not regression covariates have no impact on $Y_{i,t}$. However, there exist unobserved variables that affect both the regression equation and the selection equation.

In the observable selection case the above relationship is interchanged. Here we have:

$$E(\epsilon_{i,t} Z_{i,t} \mid X_{i,t}) \neq 0 \text{ and } E(\epsilon_{i,t}\delta_{i,t} \mid X_{i,t}) = 0 \tag{5}$$

This case applies if some of the Z–variables that are not regression variables have an impact on $Y_{i,t}$ which is not explained by the covariates of the regression model.

However, the model ignores the existence of unobserved variables that affect both, the regression equation and the selection equation. The observable selection appears to be well suited for the case in which $Z_{i,t}$ contains lagged dependent variables like $Y_{i,t-1}$ or $Y_{i,t-2}$. The occurence of $Y_{i,t-1}$ as a covariate for the attrition propensity scores is quite natural in a panel since $Y_{i,t-1}$ is in general known for all attriters in wave t. However, it is unrealistic to assume $E(\epsilon_{i,t}Z_{i,t} \mid X_{i,t}) = 0$ as long as $Y_{i,t-1}$ is not a covariate in the regression model.

It is immediately clear that the bootstrap strategy will fail to detect an attrition bias in case of the selection on unobservables: the resampling strategy explicitely uses the independence of $\epsilon_{i,t}$ and $\delta^*_{i,t}$, which is the simulated random propensity part in equation 2.

However, also the selection on unobserved variables makes strong asumptions about the joint distribution of $X_{i,t}, Y_{i,t}$ and $Z_{i,t}$. It assumes that the regression coefficient of Z–variables not contained in $X_{i,t}$ are 0 in the unselected population. Such a restriction cannot be tested on the basis of the observed data. It has to be deduced from a–priori knowledge. For example, if the survey organisation introduces some random variation of fieldwork rules which are known to have different impacts on the participation behavior, the fieldwork treatment indicator can be guaranteed to have no impact on Y but surely it has an influence on the participation behavior. This happened, for example, in the British household panel survey (BHPS), where a change of the interviewer was randomly introduced[1]. Such changes are known to be a source of an increased attrition risk (Rendtel 1990,1995).

However, such experimental rules of fieldwork are expensive (also with respect to attrition rates) and therefore seldom. Fitzgerald et al. (1997) conclude in their analysis of sample attrition in the PSID "that there are no suitable candidates for instruments for nonresponse[2] in the PSID and hence that we cannot adjust for selection on unobservables".

3 The implementation of the bootstrap procedure

The implementation of the bootstrap procedure is much facilitated if the data base contains variables with the estimated propensity to attrit in the current wave given participation in the proceeding wave [3]. In order to replicate the original attrition

[1]Such a rule is different from changes of the interviewer that are caused be the move of a household.

[2]i.e. Z–variables where a correlation with Y can be excluded by a-priori knowledge.

[3]The GSOEP data base containes variables which describe the reciprocal value of the risk that a household attrits from the preceeding wave to the current wave. Details that decribe their generation can be found in Rendtel (1995)

process the following rules appear to be appropriate:

1. Since the attrition risk occurs sequentially from one wave to another, the bootstrap attrition should also be performed sequentially. This is especially useful if the estimation procedure bases on an unbalanced sample.
2. In many panels attrition is an absorbing state, i.e. attrited units do not re-enter the panel. Consequently the bootstrap attrition should be absorbing. Therefore,after an attrition has occured all observations of the unit and their household splitt-offs after that wave should be skipped in the bootstrap routine.
3. One should also reflect dependencies in the participation of household members. As a rule, household members react unanimously, i.e. all household members cooperate or refuse their cooperation, see Rendtel (1995) for empirical results from the GSOEP. This strictly votes for an application of the bootstrap attrition at the household level.

4 An application of the bootstrap attrition routine

In this section we use the bootstrap approach for a sample from the German Socio-Economic Panel (GSOEP). The GSOEP is an ongoing household panel, which is similar to the Panel Study of Income Dynamics (PSID). It started in 1984 with a sample of about 6000 households and 12000 interviewed persons. A short description of the data base is given in Wagner et al. (1993a). Detailed information can be found in Wagner et al. (1993b) and Haisken-DeNew and Frick (1997).

All household members who are older than 16 years are interviewed. The regular interviewing method is a personal interview or a self-filled questionaire in the presence of the interviewer. All persons that have given an interview are followed up as long as they stay within Germany.

The GSOEP was at its 13th wave in 1996. Up to that time it had lost more than 40% of its wave 1 members through panel attrition. Because wages and labor force participation are the central topics of the GSOEP they have been choosen to be checked for effects of panel attrition.

The model used in this section is the basic model of human capital theory, which explains the log of the earned monthly gross income by the duration of the education (schooling), the duration of the participation in the labor force (experience) and the firm specific human capital expressed by the length of the job at the present employer (tenure). Such a model was also used by Becketti et al. (1988) and Fitzgerald et al. (1996) to evaluate the PSID[4].

[4]Becketti et al. did not use tenure. Instead they used some race dummies.

The choice of our example was motivated by the following considerations:

- The selected group should have a high potential risk of attrition.
- The aim of the analysis should be panel specific, for example, a before/after event comparison.

We have chosen here the group of male employees that experienced a period of unemployment of at least 12 months. The risk of attrition of these people is supposed to be high because they are less educated and have a higher propensity of residential mobility in order to get a new job. The aim of the analysis is to assess the effects of unemployment on earned income[5] if we control for basic variables such as schooling, experience and tenure.

In this example the human capital model is augmented by interactions with the indicator LTU for observations after the long–term unemployment period. Hence, the coefficient of LTU*schooling measures the depreciation of school–specific human capital, while LTU*experience and LTU*(experience^2) describe the depreciation of occupationally achieved human capital. We did not include an interaction term with tenure since the firm–specific human capital is usually completely devalued after long–term unemployment[6].

In order to reduce the number of parameters we deflated the gross income. Therefore we used only one time dummy measuring real wage increases after 1987, the year that marks the end of the economic recession in the reference period in Germany.

The sample consists of 224 male employees with 761 valid income observations. Figure 1 shows the sample status of these people during the first eight waves of the GSOEP. It appears that there are to be almost no losses due to panel attrition during the first 3 waves of the panel, contradicting general knowledge that the panel attrition is highest at the beginning of the panel. The reason for this discrepancy arises from the fact that in most cases it is necessary to observe a person in two subsequent waves to assess a long–term unemployment period. Those people who are unemployed for less than a year and attrit are not in the sample.

[5]Therefore we excluded people from the analysis, where no before or after income exists, i.e. people who enter the labor force or retire.

[6]We assume that employees do not return to their old firms, which is the usual pattern of re–entering into the labor force after unemplovment in Germany.

Figure 1: Participation behavior during the panel: Male employees with long–term unemployment. Waves 1 to 8 of the GSOEP.

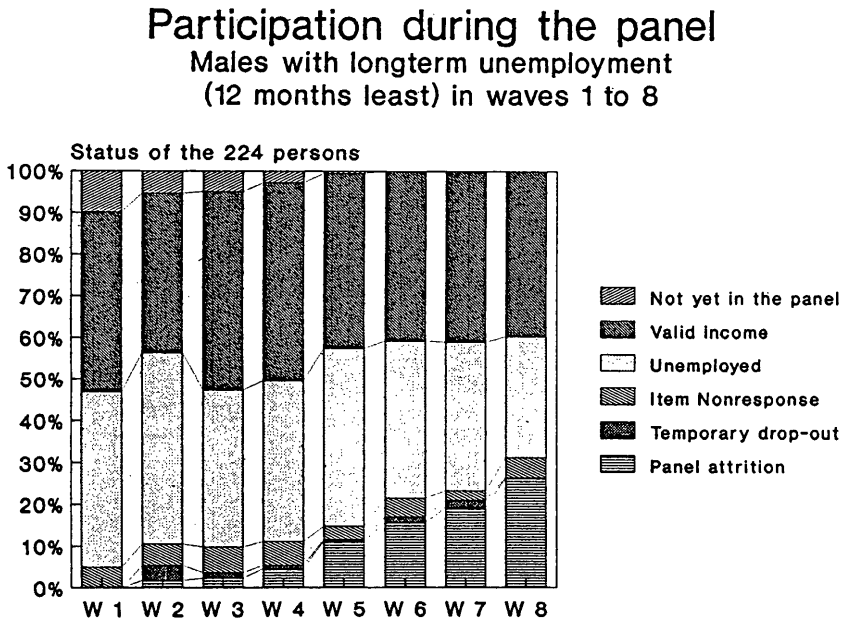


Table 1 displays the results of an unbalanced Feasable Generalized Least Squares (FGLS) analysis of a random effects model, which assumes an individual specific variance component for the error terms (Hsiao 1986, p.34). In this model all interactions of LTU with the other covariates turn out to be insignificant. Thus, there appears no further depreciation of human capital after long–term unemployment. However, there remain permanent effects due to at least one year of missing experience.

Table 1: Estimation of the income of male employees with an observed period of long–term unemployment (LTU) of at least 12 months. Dependent variable: ln(monthly gross income). Source: Waves 1 to 8 of the GSOEP. Number of persons: 224. Number of valid income measurements: 761.

Characteristic	FGLS estimate	t-values
Constant	7.261	50.9
After 1987	0.047	2.2
Schooling	0.031	2.5
Experience	0.020	2.4
Experience^2	-3.6×10^{-4}	-1.6
Tenure	0.006	2.1
After LTU	0.132	0.9
After LTU*Schooling	-0 007	-0.6
After LTU*Experience	-0.004	-0.5
After LTU*(Experience^2)	0.3×10^{-4}	0.1

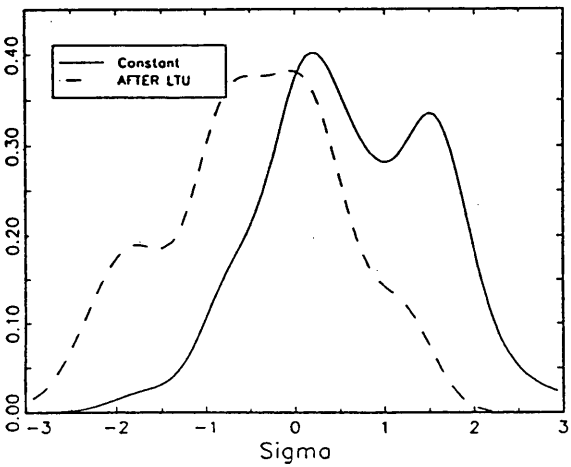
In Table 2 we compare the means and standard deviations of $\widetilde{\beta} - \hat{\beta}^*$. We find no indication of a large systematic bias. The highest bias is indicated for the constant and for the interaction of LTU with the constant. The distribution of $\widetilde{\beta} - \hat{\beta}^*$ for these two coefficients is displayed in Figure 2. There appears a bimodality for the simulated distribution of both coefficients and it is obvious that the minor modal values correspond to each other[7]. Such a plausible correspondence indicates that the estimated trade–off of both coefficients is sensitive to the panel attrition.

[7]For the distribution of the other coefficients, not documented here, there is no such bimodality.

Table 2: Results of B=100 bootstrap attrition experiments. $\widetilde{\beta}$: Estimate of β on the basis of the observed sample. $\hat{\beta}^*$: Estimate of β on the basis of the bootstrap sample.

Characteristic	Mean of $\widetilde{\beta} - \hat{\beta}^*$	Std. Deviation of $\widetilde{\beta} - \hat{\beta}^*$
Constant	0.0437	0.0650
After 1987	-0.0090	0.0126
Schooling	-0.0025	0.0057
Experience	-0.0021	0.0049
Experience^2	0.4×10^{-4}	1.1×10^{-4}
Tenure	-0.0004	0.0018
After LTU	-0.0577	0.1173
After LTU*Schooling	0.0051	0.0113
After LTU*Experience	0.0019	0.0058
After LTU*(Experience^2)	-0.5×10^{-4}	1.4×10^{-4}

Figure 2: The distribution of $\widetilde{\beta} - \hat{\beta}^*$ for the constant and the coefficient LTU. Kernel density estimate on the basis of B=100 bootstrap attrition experiments.



5 Conclusions

In many panel studies there are good reasons to assume that panel attrition is caused by field–related variables. If the model variables are not correlated with the field work variables attrition will turn out to be ignorable.
The bootstrap routine we proposed here efficiently uses the available knowledge about the attrition process in a panel study. The attractive features of the routine are:

- It is not necessary for the reseacher to estimate an attrition model if there are appropriate variables with attrition propensities in the data base.
- The researcher does not have to use a new estimation routine for his model of interest. It suffices to apply the same estimation routine to different data sets.
- The routine works for every analysis, not only for regression analysis.
- The routine gives a reasonable estimate of the size of an attrition bias.

All the researcher has to do is the programming and the execution of the bootstrap attrition experiments, which seems relatively simple.

In order to achieve these merits one has to rely on the assumption that the selection process can be controlled by observable variables and that these observable variables are contained in the model underlying the generation of the attrition propensities in the data base.

There are some alternatives to the bootstrap routine, especially in the case of regression analysis: First, one can augment the regression equation with the observed attrition variables. However, nonzero estimated slope coefficients of these variables are not always an indicator of an attrition bias nor differences in the estimated β–values indicate always an attrition bias.

Second, one can run a weighted regression analysis where the weights are generated by the inverse of the attrition propensities. Under the assumption of selection on observables such a weighted regression analysis yields consistent estimates of β, see Cosslett (1993), DuMouchel and Duncan (1983), Fitzgerald et al. (1996), Little (1991) and Nathan and Holt (1980).

References

Becketti, Sean; Gould, William; Lillard, Lee; Welch, Finis (1988). The Panel Study of Income Dynamics After 14 Years: An Evaluation. Journal of Labor Economics, 6, 472–492.

Cosslet, S. (1993). Estimation from Endogenously Stratified Samples. In: Handbook of Statistics, Vol.11, eds. Maddala, Rao, Vinod, Elsevier.

DuMouchel, William; Duncan, Greg (1983). Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples. Journal of the American Statistical Association, 78, 535–543.

Efron, Bradley (1994). Missing Data,Imputation, and the Bootstrap, Journal of the American Statistical Association, 89, 463–475.

Efron, Bradley; Tibshirani, R. (1993). An Introduction to the Bootstrap, Chapman and Hall,London.

Fitzgerald, John; Gottschalk, Peter; Moffitt, Robert (1996). An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics, Paper presented at the 7th Conference on Panel Data, Paris.

Haisken–Denew, John; Frick, Joachim (1997). Destop Companion to the German Socio–Economic Panel Study (GSOEP), German Institut for Economic Research (DIW), Berlin.

Hsiao, Cheng (1986). Analysis of Panel Data, Cambridge University Press, Cambridge.

Little, Roderick (1991). Inference with Survey Weights. Journal of Official Statistics, 7, 405–424.

Nathan, G. ;Holt, D. (1980). The Effect of Survey Design on Regression Analysis. Journal of the Royal Statistical Society, B, 42, 377–386.

Rendtel, Ulrich (1990). Teilnahmeentscheidung in Panelstudien: Zwischen Beeinflussung, Vertrauen und sozialer Selektion. Über die Entwicklung der Antwortbereitschaft im Sozio–ökonomischen Panel. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 42, 280–299.

Rendtel, Ulrich (1995). Panelausfälle und Panelrepräsentativität, Campus Verlag, Frankfurt/M. New–York.

Verbeek, Marno; Nijman, Theo (1996) Incomplete Panels and Selection Bias: A Survey. In: Matyas, L.; Sevestre, P. (eds): The Econometrics of Panel Data: Theory and Applications, $2^n dEdition$, Kluwer Academic Publishers, Dordrecht London.

Wagner, Gert; Burkhauser, Richard; Behringer, Friederike (1993a). The English Language Public Use File of the German Socio–Economic Panel. Journal of Human Resources, 28, 429–433.

Wagner, Gert; Schupp, Jürgen; Rendtel, Ulrich (1993b). Das Sozio–oekonomische Panel (SOEP) — Methoden der Datenproduktion und –aufbereitung im Längsschnitt. In: Hauser, R.; Ott, N.; Wagner, G. (Hrsg.): Mikroanalytische Grundlagen der Gesellschaftspolitik — Band 2: Erhebungsverfahren, Analysemethoden und Mikrosimulation, Akademie Verlag, Berlin, 70–112.

Regression-Based Nearest Neighbour Hot Decking

SEPPO LAAKSONEN

Abstract: *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 ratio-scale variable which consists of a high number of non-known zero values. The results obtained by means of the method are compared with those obtained by random hot decking. 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. In the first part of the paper, imputation methods and imputation strategies are discussed more generally. The paper also develops a diagnostic test for the quality of imputations; this test checks how many times the same donor is used in imputing missing values.*

Keywords: *diagnostics of imputed values, imputation variance, model-value imputation, nearest neighbour imputation, real-value imputation*

1 Introduction

Imputation is typically used when needed to substitute missing item values with certain fabricated values in surveys or censuses. The method may even be practicable by replacing the missing values of unit non-respondents, being thus a competitor for reweighting methods. Numerous alternative imputation methods are mentioned in the literature (see e.g. Kalton and Kasprzyk 1986, Little and Rubin 1987). I have been confused with most classifications, because some methods are only variations of a more general methods family. For example, mean imputation is a simple application of regression (or modelling) imputation. Therefore, I classify the imputation methods into the four main categories, the first of which is not a real imputation method, but instead a course of action, or a baseline:

0. Use of available/complete cases, when any missing items have not been imputed.
1. Deductive or logical imputation; there is a known function (identity equation) between certain observed values and missing values.
2. Imputed values are derived from a (behavioural) model, that is, imputed values may be non-observable in a real life world. I call this methods family *model-value imputation.*

3. Imputed values are derived from a set of observed values, from a real donor respondent. This is called *real-value imputation*. Note that the methods group 2 may provide a real value as well, but this is not derived directly from a real donor.

This classification thus essentially reduces the number of separate imputation methods, compared with the lists presented for example in Kalton and Kasprzyk (1986), in Särndal et al. (1992) or in West et al. (1996). I see the distinction between methods 2 and methods 3 to be useful for better understanding the nature of imputations, since the latter one gives always natural, possible values, whereas the former may provide impossible values as well. This feature is not always an advantage for a certain method; this is the case, for example, if the observed values do not cover all potential values exhaustively. Real-value imputation is impossible to correctly apply if there are no respondents within some areas. It is as well problematic to use if the share of respondents is low. In such cases a modelling technique may be more helpful supposing that a model is predictable enough. The best method is in such cases to collect additional information from these units.

Standard methods may be broken down in the groups above. Cold decking is rather a method of group 1 than that of group 2. Regression-based and other model-based imputation methods (including mean/median/mode and ratio imputation), being deterministic or stochastic, belong to group 2, whereas hot decking with its ordinary variations such as random hot decking, sequential hot decking and nearest neighbour hot decking belongs to group 3. Instead, it is not fully clear on how to classify nearest neighbour hot deck when used a linear or other interpolation in the ordered list (West et al. 1996). This method is a mixed method, exploiting both model-value imputation (but very simple) and real-value imputation. More generally, mixed imputation methods may be best in many practical situations.

A newer technique, that of neural networks (e.g. Nordbotten 1996), is a model-value method which exploits non-linear 'learning' models. The division into single and multiple imputation methods (Rubin 1987 and 1996) is set under group 2 or group 3 depending on its mother method. Methods 0 and 1 are not of real interest, method 1 is considered part of the editing process. Method 0 is used in particular when comparing the effects of real imputation methods. The second aspect of method 0 concerns the serious practical question whether the files with numerous imputed values for several variables could be without problems utilised in all further analyses, including distribution measures and multivariate methods, or whether it would be better to use only complete cases or partially imputed cases in such analyses.

So, although a certain imputation method may be advantageous to some measures, it might be useless or even disadvantageous to some others. The choice of the best imputation method for a certain situation is a difficult task, correspondingly. This paper

first in Section 2 discusses principles important when selecting an imputation method, and then makes an exercise for a real situation in Sections 3 and 4. The method applied is called regression-based nearest neighbour hot decking, since it exploits both regression model and hot decking. This method belongs to the family of real-value imputation methods, since the imputed values are derived from respondents.

Our method consists of some new elements, although it is familiar with *predictive mean matching* which method is considered earlier in Little (1988), for instance. Recently, Landerman et al. (1997) test this method in the situation where the variable to be imputed is an independent rather than a dependent variable in a substantive model. The term 'predictive mean matching' is derived from the fact that the method was first used for statistical matching, not for real imputations. The specification of Little (1988) is partially different from that used by Landerman et al. (1997) and by the present author, as well. As a conclusion, I want to use this new term '*regression-based nearest neighbour hot decking*' because it is more illustrative than the old term.

From the point of view of variance estimation the methods groups differ essentially as well. When using only completed cases, it is needed to take into account *sampling variance* solely. The same concerns logical imputation in which case we usually presume that the imputation model used fits exhaustively. However, if this does not hold true, we will have a certain component of variance/bias which enhances the inaccuracy of the estimates. As far as methods 2 and 3 are concerned, we should in all cases add *another variance component* because of the fact that some missing values have been imputed. There are various methods for this purpose. We consider a replicate method in Section 4.

Our regression-based nearest neighbour method is applied to the variable 'overtime hours of workers in enterprises' of the Finnish data of the European Union Wages Structure Survey. Imputations are needed since the data provider exempted a high number of enterprises from the reply to some special questions in order to reduce their response burden (and to increase response rate) and also the work burden of statisticians themselves (intentional item nonresponse). The original overtime hours and imputed hours are used in the second step to impute the amount of overtime wages. The method used here is standard regression imputation. This paper does not deal especially with this last application, since this second step is not difficult after the successful first imputation step. This is due to a strong regression model when predicting overtime wages through overtime hours as the best explanatory variable.

2 Needs and strategies for imputations

Imputation is a standard technique for substituting missing values with fabricated ones. Usually, it is used when some item values are missing but it may be used instead of reweighting methods as well. For example, if we impute all the values for one missing unit B drawing those from a real respondent/donor A, this hot deck method corresponds to such a reweighting method in which the original sampling weight of unit A is multiplied by two. Särndal (1996) even proposes a common framework for analysing the estimation error following from the both techniques. He uses in this context the term a 'surrogate estimator.' Nevertheless, the approaches and the techniques of both adjustment methods differ from each other, although they have much similarities. Imputations yield finally one or more completed data sets which can be utilised as ordinary ones for point estimation, but variance estimation may be problematic.

Another common feature of both adjustment methods is naturally the need for auxiliary information. There are some differences in types of this information; for example, it is not useful to exploit aggregated/macro auxiliary information in imputations, that is typical in such reweighting methods as post-stratification and calibrations. On the contrary, micro-level auxiliary data give advantages for both methods, but these data are necessary for imputation methods which can exploit partial auxiliary data easier than reweighting methods. Another important common feature of both missing data adjustment methods concerns adjustment cells, or groups within which the response mechanism is assumed to be ignorable (Rubins term, e.g. 1987). This means for example that these cells are expected to be homogeneous so that the respondents and non-respondents within these cells are similar.

Eltinge and Yansaneh (1997) wish to define these groups so that there are approximately equal response probabilities, or equal values of a specific survey item. They also discuss an important question on how to define these cells optimally. The method used by Ekholm and Laaksonen (1991), Laaksonen (1991) and Heiskanen and Laaksonen (1996), based on logistic regression when adjusting for unit nonresponse, is one appropriate method. The same modelling technique may be useful when forming imputation cells as well. The assumptions required for these cells are demanding, and if not satisfied, harmful biases in estimates may result. Obviously most cells are fairly homogeneous, but there are in practical situations also such cells which consist of only few if any respondents but of a number of non-respondents. These cells are often situated in certain extreme areas, e.g. comprising relatively many poor or rich people in income surveys (Laaksonen 1991), and the imputation method used may have a substantial impact on estimates, correspondingly. This type of problematic areas may be observed and diagnosed by a good imputer, but this does not seem to be a normal case in today survey quality reports, unfortunately. It is more difficult or impossible to correctly diagnose such cases when the cells are

homogeneous enough, but the values of an outcome variable may be varying within these cells and variously from a respondent to a non-respondent.

What criteria should be used when choosing an imputation method itself? This is another substantial question to which any simple answer does not exist. The solution depends at least on the following four aspects:

(i) The importance of a variable being imputed. If this variable is of high importance, it is natural that the selection of an imputation technique should be especially careful.
(ii) The type of a variable being imputed. We have to distinguish here the scale of a variable, that is, whether it is metric (ratio-scale, interval scale) vs. non-metric (categorical, ordinal). This leads to similar situations as in standard model specifications, that is, we should consider such questions as whether to use a linear specification, or a logit/probit/tobit model or some else. We do not discuss the details of these alternatives. In the case of a ratio-scale, that is typical for variables of economic surveys, the minimum of the possible values is thus zero, and the negative values are not acceptable. The situation is easier, if we know which missing values are non-zero, since in such cases we only need to impute these values, and we can use an appropriate model-value imputation without severe problems (see e.g. Heeringa et al. 1997 who have also certain bracketed information on non-zero values). On the contrary, if we have no idea what missing values are zero, what are non-zero, the choice of the imputation method is more limited. The example of the present paper considers just such a case.
(iii) The statistical figures desired to estimate. If the means and the totals only are of interest, a simple method such as mean/median/ratio imputation may be reasonable, although there are problems to estimate the correct variances. The requirements for imputations are more demanding, if the distributions of variables or the associations between variables are to be properly analysed in partially imputed data sets. The exercise of this paper aims at tracking distributions as well as possible, and also at taking into account some associations. It should be noted that we impute only few variables. The problem will be worsened if the number of imputed variables increases. On the other hand, the problem will be slight, if the imputation model is strong.
(iv) The nonresponse rates and the accuracy needs. Imputation is a post-survey adjustment method, which should not be used too much, and less often if the (item) nonresponse rate is high. The problem is not so severe, if a customer receives the correct information about the accuracy of the statistical measures. This is obviously not an usual situation today due among others to the fact that the estimation of accuracy measures such as variances is a fairly difficult task.

3 Regression-based hot decking

The Finnish data set for the 1995 European Union Wages Structure Survey is based on a complicated survey design. It is derived from several data providers (both public and private associations of employers, and Statistics Finland), some parts of these data sets are based on samples, some on censuses. The content of the whole survey changed from the former one, being now wider than in old wages statistics, in particular, concerning items of wages. All these new items were not possible to easily collect and hence, a sub-sample of enterprises was drawn. The required new questions were inquired only from this sub-sample which covered about 40 per cent of all the enterprises. The sampling fraction of the sub-sample was higher for larger enterprises than for smaller ones, since the larger ones were more able to give this additional information.

Our task is to impute the missing values for the workers of those enterprises who do not belong to a sub-sample. This is necessary in order to cover the whole wage/salary paid for each worker. We have two types of variables with item nonresponse: wages paid from overtime work, and bonuses or other additional wages paid occasionally. These behave fairly differently: paid overtime work is not common whereas some additional wages are paid for most workers. Basically, both situations lead to use the essentially similar methodology. We here only consider the former case that is more difficult and also more interesting because of a two-step strategy used. We next describe this strategy:

- In the first step we impute overtime hours for each non-respondent and
- in the second step our final target, paid overtime wages for them.

Next we pay mostly attention to the imputation methodology of the first step, since the second step is fairly easy assuming that the first is successful, due to a high predictability of overtime wages when overtime hours are known or imputed, exploiting such explanatory variables as regular time wages per hour, industry classification, age, gender, firm size and occupation.

Let k (k=1,...,K) be a worker of a certain respondent enterprise, and y_k = an observed value of overtime work hours of this worker, correspondingly. Our target is to provide the best substitute for the rest of the workers, say l=1,...,L. These imputed values are here denoted by y_l*. The methodology used runs as follows:

1. Take the data of respondents.[1]
2. Construct a multivariate regression model[2] so that y_k is the dependent variable and the variables without missing values are potential independent or explanatory variables.
3. Compute the predicted values for both K respondents ($\hat{y}_k$) and L non-respondents (y_l).
4. Order the data set by the predicted values, and calculate the distances for each non-respondent.[3]
5. Search the nearest neighbour from K respondents for each missing unit l using this ordered data set.[4] Let this value be y_k^l .
6. Put $y_l^* = y_k^l$.

This method is here called *regression-based nearest neighbour hot decking*, since it first exploits standard regression, but finally picks up the imputed values from the really observed data set, analogously to nearest neighbour hot decking.

Our method is expected to be better than regression imputation and random hot decking for the reasons discussed below. Pure regression imputation, thus replacing the missing values with predicted values from the estimated regression model, is not competitive, since it would have given a high number of negative imputed values in our situation (the same problem is common in less complicated data sets as well). Secondly, it is well-known that pure regression imputation underestimates the variance. Hence, it is usual to add a noise term to the predicted values in order to avoid this problem. There are the two

---

1 Our data for a real situation consist of 155000 employees, about 60000 of which responded. In our simulation exercises we have this set of the 60000 respondents as a benchmarking data set. The missing values were given for sub-samples of these data. We constructed several sub-samples, with various item nonresponse rates. All these were drawn at enterprise level and quite similarly to the real situation. Stratification by size class was used as well. In each simulation run we made attempts to impute the missing y values.

2 In our empirical exercise we had the following explanatory variables: industry classification (15 dummies), occupation (8), size class (6), gender (1), age and square of age, regular time wage, regular time hours, number of years worked for the enterprise. The estimated model fits fairly well (R-square = 12%) if we take into account that our data are micro-based and very heterogeneous, 87% of the y values being zero, and the others varying a lot. All the other explanatory variables except the last one were significant, regular time wage and occupation group being most significant.

3 We only checked 15 nearest neighbours in our practical situation but 30 neighbours in simulation experiments. If no respondents found, no imputation used.

4 It is possible to do this search separately for sub-groups. In our real situation we used four sub-groups which were obtained by cross-classifying gender and enterprise size (small and medium sized vs. large enterprises). Our users were more satisfied with such results, but we have not been able to find any assertive rules which type of sub-groups or if any should be used. This requires further research work. In simulation exercises we have only two sub-groups, males and females, in order to avoid too small imputation 'fields'.

alternatives to do this: (i) to draw those terms from the normal distribution (0, mean square error); we assume here that the model assumptions hold true; (ii) to draw those from the set of respondent residuals at random. In our situation alternative (i) was not possible because we were not able to build such a model that would have satisfied the reasonable assumptions of regression models. This would not have as well given the correct distributions of our y variable (obs. that there are a high number of zeros). Alternative (ii) would have led to the last problem unless conditioned in a good way.

Our method is expected to succeed well with respect to many criteria, although we cannot build any excellent regression model for the first step of our imputations. It however ranks the respondents and non-respondents so that the probability for a non-zero imputed value will be increased while the predicted values will be increased. This being the case the imputed values look basically similar to observed values. There may be problems due to the poor balance between respondents and non-respondents within certain sub-intervals.

The second alternative for our method, random hot decking, is (too) much used so that the imputed values are drawn from the respondents within the same imputation cell. The crucial question is thus how to construct these cells. In this situation we have no exact idea for this, although many auxiliary variables are available. The best solution could be to form the imputation cells by dividing the interval of the predicted values into the reasonable number of such cells. This method is fairly close to that we used, but however less efficient. In the empirical part we computed some crucial estimates using random hot decking without cells in order to compare the results from both methods.

Variance estimation

The point estimates are not enough in survey sampling. During recent years several methods have been presented to estimate variances for imputed data. Rubins (1987, 1996) multiple imputation is one of these methods. Fay (see the discussion of Rubin 1996) presents a fractionally weighted imputation which is close to multiple imputation as far as point estimation is concerned, but his variance estimation takes benefit of the Jack-knife method of Rao and Shao (1992), done for single hot deck imputation. Shao (1997) has some further developments concerning other imputation methods as well. Särndal (see, e.g. 1996) has provided variance estimates for single imputations.

For our regression-based hot decking, we have the two components of the whole variance: (1) sampling variance and (2) imputation variance. The sampling variance is estimated from the imputed data set assuming that these imputed values are as correct as the observed values. The imputation variance takes into account the uncertainty in imputations themselves. This uncertainty is estimated in this case as follows:

a) Assume that the error term of the regression model is normally distributed with the zero mean and with the variance equal to the mean square error [$\varepsilon \sim N(0, MSE)$].
b) Take the uncertainty of the regression model into account when searching the nearest neighbour for each non-respondent by drawing a random number of N(0,1), say z_1, and add $z_1 * RMSE$ to the predicted value of the regression model (RMSE = root MSE).
c) Perform the operations 4 to 6 of the original scheme above.
d) Repeat the steps (b) and (c) a reasonable number of times.
e) Calculate the point estimates needed after each simulation run.
f) Calculate the variance of the point estimates over all the simulations, that is our *imputation variance.*

The final variance is the weighted sum of the sampling variance and the imputation variance. We here emphasise on imputation variance, the variability of which is derived from various alternatives to rank the respondents and the non-respondents, and to find the nearest respondent for each non-respondent, correspondingly. Note that the roots of this method are derived from a specification of multiple imputation of Rubin (e.g. 1987). He proposes to create several completed data sets by multiple imputation and then to use the variability in the estimates obtained from these simulation runs in variance estimation. He says that 3 or 5 completed data sets are a reasonable number in practice. In the empirical part of this study we produced 64 simulation runs (Table 2). This number gave fairly robust variance estimates, but using only 3 to 5 runs, the results would have been too inaccurate.

4 Empirical findings

We checked the quality of our imputations on the following three aspects: (i) asking the evaluation of the main users, (ii) analysing the bias, (iii) computing imputation variances. The first aspect concerns the real data set. The users had some expectations and they were quite satisfied with the results which covered several tabulations for means and totals, by gender, industry classification, size class and occupation group. In addition, we compared the estimated means derived from the data set of the respondents on one hand, and the completed data set on the other. These differences were not dramatic which was a good thing from the users' point of view.

However, the users' evaluation/opinion is not in general enough to confirm the quality of imputations. Therefore, we extended the evaluation using simulations for points (ii) and (iii). To understand the bias we drew a number of random samples of the data set of the responding enterprises, performed the imputations and estimated various test figures. The average of these estimates was compared with the known population (benchmarking)

value on one hand, and with the results obtained from random hot decking. Our application of the random hot decking method is based on 'sampling without replacement', that is, each respondent can be a substitute only for one non-respondent. Table 1 gives some comparative figures.

Table 1: The biases of some point estimates based on 64 simulations.[5] The item nonresponse rate is in both cases the same, 33 percent

		Difference from the Benchmarking, %			
Sub-Group	Number of workers in the bench-marking data	Regression based nearest neighbour hot decking		Random hot decking	
		Total	Standard deviation	Total	Standard deviation
All	59878	-0.5	+0.4	-0.9	+0.1
Gender					
Male	15402	-1.4	-0.8	-7.1	-4.5
Female	44476	-0.1	+1.3	+2.3	+3.1
Size Class					
-9.9	1514	+14.7	+9.6	+17.2	+6.5
10.0-19.9	2942	-0.3	-0.2	-5.6	-5.8
20.0-49.9	3147	-2.1	-5.0	-6.4	-7.5
50.0-99.9	8110	+1.0	+0.7	+0.9	+0.3
100.0-499.9	6954	+0.2	+1.3	-11.7	-7.7
500.0+	37211	-1.5	-0.8	+3.5	+5.2

The results based on our method are promising for most point estimates. The bias derived from it is in almost all cases lower than obtained by random hot decking; the latter is slightly better only if the bias is very low with both methods. The most fatal biases are observed for small sub-groups, in particular, for the smallest one. It seems that this sub-group is in some sense exceptional, and hence our method as well succeeds badly with it. Random hot deck gives systematically biased figures, sometimes negative, sometimes positive, in few cases also fairly good ones (at random). We made tests with higher nonresponse rates as well, the results were in the same direction as presented in Table 1.

The second series of simulations, based on 64 simulations, was done for the various

[5] This fairly low number of simulations may be criticised but our experience shows that even the lower number gave already the base line of the results.

proportions of missing values. Table 2 presents the most interesting results. We here do not present the estimates of sampling variances since these here are not of high importance. Another reason is that these are dependent on the assumed sampling design. If the design is simple random sampling, the estimates are around 2 percent in the largest size class and more than 8 percent in the smallest one. The two-stage cluster sampling gives much higher estimates of sampling variances.

Table 2: Square roots of relative imputation variances (coefficients of variation), %, for some point estimates based on 64 simulations

	Coefficients of variation in some examples by item nonresponse rate					
Sub-group	10%	20 %	33%	50%	67%	80%
Gender						
Male	1.2	1.6	1.8	1.9	2.7	3.1
Female	0.6	1.1	0.9	1.3	1.7	2.1
Size Class						
-9.9	4.8	6.6	6.9	8.8	6.9	9.5
10.0- 19.9	2.4	4.2	5.1	6.2	5.8	6.0
20.0- 49.9	1.9	3.9	3.7	5.9	5.7	8.3
50.0- 99.9	1.6	2.1	2.6	3.5	3.4	3.8
100.0-499.9	0.9	2.3	2.1	2.5	3.0	3.3
500+	0.9	1.1	1.5	1.5	1.9	1.9

These findings are interesting and believable in many respects. The outcome variable itself is skew due to some high values, and hence the estimates including the variance estimates may vary sensitively. We observe this sensitiveness on the results based on the various sub-samples from the respondents, but the main line is clear: the variance estimates are increasing with the increase of item nonresponse rates. The exceptions from this base line occur in smaller sub-groups. These demonstrate the possibility that this mechanism may generate slightly biased point and variance estimates. We checked some problematic sub-groups in more detail, and observed a congeries of non-respondents in a certain part of the interval. Correspondingly, these non-respondents were too often substituted by one type of respondents; e.g. if this substitute has a non-zero value, the point estimate is more often overestimated, whereas it is underestimated in the case of a zero-value substitute. The effects of such outliers are not as dramatic for larger sub-groups; for example, the variance estimates for males and females are increasing very correctly by nonresponse rate.

It is not trivial to avoid the above mentioned over/underestimation problem. The problem is difficult even to detect in large data sets. Something we should try to do. At least, we could diagnose how many times each respondent is a substitute for any non-respondent. It is natural to compare this number with the average number of required substitutes within each imputation interval (or imputation cell). If the item nonresponse rate in interval/cell c is f_c then the average number of possible respondents is simply $(1- f_c)/ f_c$.

We checked the frequencies of the same donors for a number of simulated data sets. Table 3 gives the thorough results. These illustrate the sensitiveness of this hot deck imputation method by item nonresponse rate. We wanted especially to check the numbers of donors for largest (extreme) observations (group B), which may have a considerable effect on estimates. Note that it is not automatically a problem to use the same donor several times, if this is absolutely similar to a close non-respondent, but in a usual survey situation we cannot know that. The print such as Table 3 should be considered as an example of a diagnosis for an imputer. This should be leading to check some individual observations more carefully. As a consequence, the imputer may also be looking forward to an alternative imputation method.

Table 3: Percentages of the same donors by item nonresponse rate based on simulation data (the results are the averages of the two independent simulations, that is, these are not exact values). Group A covers 99% of the smallest values (varying from 0 to 60 hours), group B the 1% rest (from 60 to 130 hours)

	Item nonresponse rate and groups											
	10%		20 %		33%		50%		67%		80%	
Number of the same donors	A	B	A	B	A	B	A	B	A	B	A	B
1	99.2	100.0	96.0	93.8	88.7	94.7	74.8	74.5	60.1	64.9	38.9	52.8
2	0.8		3.7	6.2	9.4	5.3	17.4	13.4	22.4	15.6	20.5	11.8
3			0.3		1.6		5.8	6.1	12.6	13.0	15.8	8.5
4					0.3		1.5	5.2	3.1	5.2	10.1	7.4
5							0.4	0.8	2.6	1.3	6.5	7.0
6							0.0		0.5		4.0	5.8
7									0.2		2.5	4.1
8+											1.7	2.6

5 Discussion

The use of imputations has become more common during last years. It is due to the increase in item/unit nonresponse rates, and to the advanced methodology for imputations. This trend is not only a good thing, since an imputed value will never be an ideal substitute for the real observed value, imputation always provides fabricated values to a certain extent. A general problem is also that a user cannot recognise easily the quality of survey estimates in the case of imputations. The quality checking requires a careful analysis of the completed data set using various statistical measures and covering also small sub-groups of the whole population.

The sensitiveness of imputed values may be considered using various imputation techniques and various specifications, although one of these techniques should have been chosen, finally. The checking may reveal for example that there are difficulties to correctly impute some target groups because of the poor information about non-respondents through auxiliary variables. This means, in the case of model-value imputations, that a model is not fitting well within such groups. Correspondingly, in the case of real-value imputation, the same respondents are used 'too often' to provide substitute values for missing items. To get some understanding of the last point, it is useful to calculate the frequencies of such duplications, and to know in which groups these occurred. The problems of this sort are in practice worst in margins of distributions where are not many observations.

The present paper pays attention to the above mentioned problems while it presents a somewhat new application to nearest neighbour hot decking. In the empirical analysis we test a fairly difficult metric variable that is very heterogeneous and consists of a high number of non-known zero values. Our results based on simulation experiments are promising. The bias is not bad for most sub-groups, and in most cases much smaller than using an alternative method, that is, random hot decking. We also estimate imputation variances for several sub-groups and by various nonresponse rates. The variance estimation needs further research in the future.

References

Ekholm, A. and Laaksonen, S. (1991). Weighting via Response Modeling in the Finnish Household Budget Survey. Journal of Official Statistics 7, pp. 325-337

Eltinge, J.L. and Yansaneh, I.S. (1997). Diagnostics for Formation of Nonresponse Adjustments Cells, With an Application to Income Nonresponse in the U.S. Consumer Expenditure Survey. Survey Methodology 23, pp. 33-40

Heeringa, S.G., Little, R.J. and Raghunatan, T.E. (1997). Bayesian Estimation and Inference for Multivariate Coarsened Data on U.S. Household Income and Wealth. Invited Paper for the 51st Session of the ISI, Istanbul

Heiskanen, M. and Laaksonen, S. (1996). Nonresponse and Ill-Being in The Finnish Survey on Living Conditions. In: Laaksonen, S. (ed.). International Perspectives on Nonresponse. Research Reports 219, pp. 81-100. Statistics Finland

Kalton, G. and Kasprzyk, D. (1986). The Treatment of Missing Survey Data. Survey Methodology 12, pp. 1-16

Landerman, L.R., Land, K.C. and Pieper, C.F. (1997). An Empirical Evaluation of the Predictive Mean Method for Imputing Missing Values. Sociological Methods and Research 26, pp. 3-33

Laaksonen, S. (1991). Adjustments for Nonresponse in Two-Year Panel Data. The Statistician 40, pp. 153-168

Little, R. (1988). Missing-Data Adjustments in Large Surveys. Journal of Business & Economic Statistics 6, pp. 287-297

Little, R. and Rubin, D. (1987). Statistical Analysis with Missing Data. New York: John Wiley & Sons

Nordbotten, S. (1996). Neural Network Imputation Applied to the Norwegian 1990 Population Census Data. Journal of Official Statistics 12, pp. 385-402

Rao, J.N.K. and Shao, J. (1992). Jack-knife Variance Estimation With Survey Data Under Hot Deck Imputation. Biometrika 79, pp. 811-822

Rubin, D. (1987). Multiple Imputation in Surveys. New York: John Wiley & Sons

Rubin, D. and the papers and the discussion by B. Fay, J. Rao, D. Binder, J. Eltinge and D. Judkins (1996). Multiple Imputation After 18+ Years. Journal of the American Statistical Association 91, 434, pp. 473-520

Särndal, C-E. (1996). For a Better Understanding of Imputation. In: Laaksonen, S. (ed.). International Perspectives on Nonresponse. Research Reports 219, Statistics Finland, pp. 7-22

Särndal, C-E., Swensson, B. and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer

Shao, J. (1997). Variance Estimation for Imputed Survey Data With Non-Negligible Sampling Fractions. Invited Paper for the 51st Session of the ISI, Istanbul

West, S.A., Kratzke, D-T. and Robertson, K.W. (1996). Alternative Imputation Procedures for Item-Nonresponse from New Establishments in the Universe. ASA Proceedings of the Section in Survey Research Methods

Handling of Household and Item Nonresponse in Surveys

RAJENDRA P. SINGH AND RITA J. PETRONI

Abstract: *For the 2000 Census, the U.S. Census Bureau will select a quality check, also known as integrated coverage measurement (ICM), sample to improve Census estimates. The ICM sample is subject to missing data due to household and item nonresponse. This paper discusses alternative methods researched to deal with nonresponse in the ICM sample. These methods include no adjustment for household nonresponse and no item imputation, use of Census short form characteristics to perform household nonresponse adjustment, substitution of Census data for ICM missing items, and alternative hot deck imputation procedures.*

Keywords: *noninterview adjustment, imputation, logistic regression, hot deck*

1 Introduction

As in any other data collection process, ICM sample data is also missing in some cases due to either whole household noninterview or nonresponse to one or more characteristics for an interviewed person. In this paper, we discuss and attempt to integrate research related to 1) noninterview adjustment for whole household noninterviews and 2) imputation methods to handle missing demographic characteristics (item) for persons. Section 2 presents an overview of the 1995 Integrated Coverage Measurement (ICM) sample design, procedures for handling missing data, and the estimation methodology used for Census Plus (C^+) and Dual System Estimation of the total population. Sections 3 and 4 describe methodological research for handling whole household nonresponse and characteristic nonresponse, respectively. In Section 3, we also discuss how the methodology impacts the allowable nonresponse rate in ICM for the 2000 Census. Summary and conclusions are presented in Section 5.

2 Background

2.1 1995 ICM sample design

The Bureau conducted the 1995 Census Test in three sites: Oakland, CA; Paterson, NJ; and Northwest Louisiana. The ICM sample, consisting of block clusters (single blocks or groups of blocks, generally with 30 or more housing units), was selected separately for each site.

The ICM methodology used three separate rosters: the R-Sample, the P-Sample, and the E-Sample. The R-Sample which tries to obtain a „true" roster from the ICM blocks was created for all three sites and was used in Census Plus estimation. Census Plus estimates are calculated based on the assumption that the R-Sample is the „truth" for the ICM blocks. The P- and E-Samples were only created for the Oakland and Paterson sites and were used in Dual System Estimation (DSE). DSE tries to obtain a roster from the ICM blocks independently of the Census. The independent roster called the P-Sample and the Census roster called the E-Sample are matched and the results of the matching are used to estimate the number of persons missed by both rosters. The E-Sample is also used to adjust the Census for erroneous enumerations. Further details on DSE and Census Plus estimation can be found in Schindler (1996).

In 1995, the ICM collected information for both DSE and Census Plus in a single interview. Initially, an independent ICM roster was collected, and then matched during the interview to a preliminary Census roster. Census Plus combined the preliminary Census roster and the independent roster into a „true" household roster. DSE used the independent roster to form the P-Sample. An overview of the 1995 ICM sample design and operations is given in Mulry and Singh (1995).

2.2 Outline of procedures for handling missing data

2.2.1 Noninterview adjustment

Whole-household noninterviews in the R- and P-Samples are accounted for by using a noninterview adjustment. Noninterview adjustment is not applied to the E-Sample since it is used only to make adjustments for erroneously enumerated persons in interviewed households.

The noninterview adjustment is done using block cluster x type of structure. The type of structure categories are: (1) one family detached house, (2) one family attached house, (3) building with two or more apartments, and (4) mobile home or trailer, boat, tent, van, etc., and other. If predefined criteria are not met at the block cluster level, block clusters are

collapsed according to predefined rules to control mean square error. For details, see Ikeda and Petroni (1996).

2.2.2 Characteristic imputation

Some persons in interviewed households are missing demographic characteristics required to assign persons into estimation post-strata. Missing characteristics are filled in using an imputation procedure. The 1995 ICM used a different method from the 1990 Post Enumeration Survey (PES) to impute missing characteristics. We describe both methods below along with the 1995 Test Census imputation method.

1990 PES imputation method

The item nonresponse imputation method used in the 1990 Census is a hot-deck imputation procedure that fills in values for the missing data in the P- and E- Samples. Certain information about other household members is used in the hot-deck procedure when such information is available. When information on other household members is unavailable, the hot-deck procedure imputes values based on either a previous household with reported values or the distribution of reported values in the entire file. Tenure is imputed first, followed by race, Hispanic origin, sex, and age.

When information on other household members is unavailable, race and Hispanic origin imputations use values from a previous reporting household. Tenure is imputed from a previous reporting household.

The sex and age imputations use values based on the distribution of reported values in the entire file. If one spouse reports sex and the other doesn't, the nonreporting spouse's sex is imputed as the opposite sex of the reporting spouse. For any other cases of missing sex, the imputation is based on the reported sex distribution by household size.

Missing age is imputed based on household size. One-person households with missing age are imputed from the age distribution of all reporting one-person households based on marital status. Households of two or more people have missing ages imputed from the age distribution of all other reporting individuals in households of two or more people with similar relationships to head of household, marital status and age of head of household as the individual with missing age. For more information on the 1990 imputation method, see Diffendal and Belin (1991).

1995 Test ICM imputation method

The 1995 ICM Methodology imputes missing data similarly for all three ICM samples. We impute tenure using the previous household with nonmissing tenure and sex of married householder (spouse of householder) as opposite to that reported for spouse of householder (married householder). We use the Flexible Matching procedure to impute race, sex of unmarried persons, age, and Hispanic origin (Ikeda and Petroni 1996). The Flexible Matching procedure is a hot deck imputation which identifies matching variables and uses the variables to match an incomplete record with a complete record (Williams 1995a, 1995b).

1995 Test Census imputation method

We use a two part imputation process for the 1995 Census methodology. For part one, we first edit reported Census responses for race, age, and date of birth within a household. Then, within households we edit household relationships based on age, reported relationships, and sex. When relationship is missing we allocate it based on other available data. Next, we allocate age based on relationships and ages of other persons in the household. We assign missing race (Hispanic origin) based on relationships and race (Hispanic origin) of other persons in the same household if at least one person in the household reports race (Hispanic origin). Otherwise, we assign the race of all persons in the household to be the race of the householder of the closest previous neighbor with identical Hispanic origin. We base the Hispanic origin of all persons in the household on race.

For part two, we use hot deck imputation to substitute the nearest previous responding unit having the same race, Hispanic origin, and household size. The Census Bureau determines household size during data acceptance processing. We take race and Hispanic origin to be the race and Hispanic origin of the first person in the household reporting the items. When no one in the household reports race or Hispanic origin, we use race or origin of the nearest previous responding unit. We set initial cells in each matrix by a cold deck procedure (Spencer 1995).

2.3 ICM estimation

We discuss below two estimation methods namely the Dual System Estimation and the Census Plus.

2.3.1 Dual System Estimation

For the 1995 Test Census, the Bureau used the P- and E-Samples to obtain the DSE estimate as follows:

$$\hat{N}_{..} = \left\{ \frac{[(N_c^* - \hat{II})\ (1 - \hat{EE}/\hat{N}_e)]\hat{N}_p}{\hat{M}\, N_c^*} \right\} N_c^* = AF_d\, N_c^*$$

where

$\hat{N}_{..}$ = population estimate

$\hat{N}_P$ = weighted P-Sample total

N_c^* = Census estimate based on ICM and non-ICM blocks

$\hat{II}$ = estimated number of whole-person Census imputations

$\hat{EE}$ = estimate of E-Sample insufficient information for matching cases and erroneous enumerations

$\hat{N}_e$ = weighted E-Sample total

$\hat{M}$ = estimate of P-Sample matches

AF_d = DSE adjustment factor

See Wolter (1986) for theoretical understanding.

2.3.2 Census plus estimation

The Bureau used the R-Sample to estimate the Census Pluc (C^+) population total as:

$$\hat{N}_{..} = \frac{\hat{R}}{\hat{N}_c} N_c^* = AF_c\, N_c^*$$

where

$\hat{R}$ = weighted R-Sample total

$\hat{N}_c$ = weighted total based on Census enumerations in ICM blocks

AF_C = C^+ adjustment factor

3 Methodological research for handling whole household noninterviews

The Census Bureau pursued three whole household noninterview research projects. We discuss briefly these research efforts and their outcomes.

3.1 Treating noninterview household persons as not captured

The research was motivated from the point of view of reducing data processing time and effort. An attractive alternative is to treat persons in noninterviewed households as not captured in the P-Sample. Under this alternative, we do no household noninterview adjustment for the P-Sample. Petroni, Kearney, Town and Singh (1995) calculated 357 national level estimates using the original and the alternative DSE definition of capture along with percent differences between the alternative and original estimates. In this study, we used the 1990 PES data and the 1990 noninterview household adjustment method. The 1990 PES and 1995 ICM noninterview household adjustment methods are similar.

Results suggested differential affects on particular race and tenure groups. As a result, we recommended to not drop noninterview adjustments.

3.2 Reducing household noninterviews

Petroni, Kearney, and Gbur (1996) studied effects of differential noninterviews on C^+ and DSE estimates for the 1995 ICM test in Oakland. The focus of the study was whether the large difference in noninterview rates of the P- and E-Samples could introduce bias into comparison of C^+ and DSE methodology. They analyzed data from this study to see how changes in noninterview rates affect C^+ and DSE.

For a given household, the Census Bureau collected data for both procedures simultaneously using one computer assisted personal interviewing instrument. The instrument was expected to do equally well in obtaining interviews for both procedures. The instrument contained Census rosters which were to be displayed to the interviewer after collection of an independent household roster. The independence between the roster obtained during the ICM interview and the Census roster is critical for the P-Sample. However, interviewers could sometimes view the Census roster before obtaining the independent household roster. Thus, since the independence of the initial roster was lost due to the design of the instrument and incorrect implementation of collection instructions, we had to treat such households as noninterviews for the P-Sample, but not for the R-Sample. For Oakland, the P-Sample noninterview rate was 15.06%. The R-Sample rate was 8.54%. For details, see Ikeda and Petroni (1996).

Our weighting approach to handle household noninterviews implicitly assumes that the average noninterviewed housing unit is similar to the average interviewed housing unit with respect to the characteristic(s) being estimated at the level we calculated the adjustment. As noninterview rates increase, actual difference in interviewed and noninterviewed households can increase bias. Hence, because R- and P-Sample nonresponse rates differed, comparison of C^+ and DSE estimates could be contaminated. As a result, the Bureau mounted two research efforts: first, increase the interview rates for

both samples through a field followup (Method 1); second, make the sample more comparable by treating P-Sample households that were classified as noninterviews due to the instrument problem as noninterviews in the R-Sample too (Method 2). The P-Sample household noninterview rate for the second method is the same as the original rate. The Oakland noninterview rates for the R- and P-Samples respectively are provided in the C^+ and the DSE headings of Table 1.

For Oakland, we conducted both research endeavors and for Paterson, only the second. We compared the recomputed Census Plus and DSE adjustment factors to the original (production) factors. Selected factors for C^+ and DSE are presented in Table 1. Basically, the comparison shows that lowering versus raising noninterview levels does result in significant differences in C^+ and DSE factors.

Table 1 shows that for C^+, Method 2 factors for total persons and total owners respectively are about one half percentage point and one percentage point higher than original factors. There are three percentage points fewer persons estimated with Method 1 compared to Method 2 and owners is short by seven percentage points. Most of the Method 1 factors for C^+ are lower than the original (production) factors. In general, there is a little difference between Method 2 and the original factors for the C^+. Both the Method 2 and original factors are very different from Method 1. For DSE, Method 1 factors are higher than original factors. The maximum observed difference between Method 1 and the Original factors occurred for Hispanic renters and is about four percentage points.

Differences between the Method 2 factors and the original factors are larger in Oakland than Paterson. This is to be expected, because of the higher noninterview (NI) rates in Oakland. The original Patersons rates were 2.18% for R-Sample and 8.49% for P-Sample. We examined factors for 56 groups based on race/ethnicity, tenure, age and sex. All of the Method 2 Oakland factors are within 5% of the original factors, most (52 of 56 for both Census Plus and DSE) are within 3% of the production factors, and a substantial majority (41 of 56 for Census Plus, 44 of 56 for DSE) are within 2% of production factors.

Table 1 shows that increasing the response rate had different effects on C^+ and DSE as compared to original factors. For C^+, factors decreased while DSE factors increased. C^+ factors are more affected by 10 percentage point reduction in noninterview rate (from Method 2 to Method 1) than DSE factors. This could mean that either DSE estimates are more robust to the noninterview rate as compared to C^+ or the 3% households (5.58%-2.38%) which could not be interviewed for DSE were very different from the interviewed households. Further evaluation of C^+ for the three Methods shows that a 5%

Table 1: C^+ and DSE Adjustment factors for different household noninterview (NI) rates for Oakland

Characteristics	C^+ Estimate By NI Rates			DSE Estimate By NI Rate	
	Method (NI Rate (%))			Method (NI Rate (%))	
	Method 1 (2.38)	Original (8.54)	Method 2 (13.7)	Method 1 (5.58)	Original (15.06)
	(1)	(2)	(3)	(4)	(5)
Total	0.978	1.005	1.010	1.108	1.087
Owner	0.964	1.026	1.034	1.074	1.060
Renter	0.988	0.989	0.992	1.135	1.107
Black	0.934	0.950	0.958	1.121	1.105
Owner	0.968	0.987	0.985	1.091	1.097
Renter	0.915	0.928	0.942	1.138	1.109
Non-Black/Non API Hispanic	1.163	1.132	1.148	1.245	1.203
Owner	1.220	1.212	1.246	1.203	1.176
Renter	1.130	1.086	1.091	1.269	1.219
API	0.891	0.976	0.967	1.083	1.075
Owner	0.877	0.981	0.978	1.046	1.034
Renter	0.900	0.972	0.958	1.109	1.104
All Others	0.998	1.039	1.042	1.035	1.007
Owner	0.915	1.021	1.035	1.027	0.999
Renter	1.114	1.064	1.053	1.046	1.019

Note: Adjustment factors for Method 1 are in Columns (1) and (4), for Method 2 in Column (3) and for the original are in columns (2) and (5).

Source: Petroni, Kearney, and Gbur (1996)

drop from 13.7% to 8.54% in noninterview rates did not make as large difference on estimates as the next 6% drop from 8.54% to 2.4%. This suggests that obtaining response for the tail end of the respondents is very important to reduce bias.

3.3 Classifying households into noninterview cells using Census data

Ikeda (1996) studied the effect of using data from the Hundred Percent Edited Response File (HERF) households to assign R-, and P-Sample noninterviewed households into noninterview adjustment cells. With this approach data from matching HERF households was obtained for 1995 ICM R-Sample and P-Sample housing units. The HERF data was used to help define noninterview adjustment cells and the R-Sample and P-Sample noninterview adjustment weights were used in the calculation of new HERF based C^+ and DSE poststrata estimates.

The production noninterview adjustment system which used adjustment cells defined mostly by block cluster x type of structure was rerun using the pseudo-cluster code (crossed with the ICM sample selection stratum) instead of the block cluster code. Pseudo-cluster codes were defined by household size, tenure, and race categories.

The new and the original C^+ and DSE estimates and adjustment factors are presented in Table 2. The overall Census Plus estimate for Oakland decreases from 334,493 (production) to 331,849 (HERF based). Census Plus estimate for owner poststrata tend to decrease somewhat more than Census Plus estimates for renter poststrata (although Asian/Pacific Islander owners and renters decrease by similar proportions). The overall DSE estimate for Oakland increases from 361,550 to 366,137. DSE estimates for renter poststrata tend to increase somewhat more than DSE estimates for owner poststrata (although Hispanic owners and renters increase by similar proportions).

The C^+ and DSE factors based on field followup operations (Method 1 in Table 1) were compared with factors in Table 2. Differences between factors of race/origin x tenure from the Method 1 and production C^+ were in the same direction as the difference between the HERF based and production C^+ factors. However, the Method 1 differences were considerably larger (about 10% of the production compared to less than 2%).

The differences between the Method 1 and production DSE race/origin x tenure factors (see Tables 1 and 2) tended to be in the same direction as the differences between the HERF based and production DSE factors. However, all three sets of DSE factors were close to each other. Note that none of the differences between the Method 1 and production DSE race/origin x tenure factors were significant at the 0.10 level.

Table 2: Production and HERF based Census plus and DSE adjustment factors for Oakland

Postratum	C^+ Estimates		C^+ Adj Factors		DSE Estimates		DSE Adj Factors	
	Prod	HERF	Prod	HERF	Prod	HERF	Prod	HERF
Black Owner	50542	49989	0.9874	0.9766	56158	56470	1.0972	1.1033
Black Renter	79169	79471	0.9277	0.9312	94665	96618	1.1092	1.1321
Black	129711	129460	0.9501	0.9482	150823	153089	1.1047	1.1213
Hispanic Owner	21175	20693	1.2117	1.1841	20542	20836	1.1755	1.1924
Hispanic Renter	32645	32719	1.0858	1.0882	36663	37394	1.2194	1.2437
Hispanic	53819	53412	1.1321	1.1235	57205	58230	1.2033	1.2248
API Owner	22829	22520	0.9813	0.9680	24051	24149	1.0338	1.0380
API Renter	31248	30636	0.9717	0.9526	35520	36231	1.1045	1.1266
API	54077	53155	0.9757	0.9591	59571	60379	1.0748	1.0894
Other Owner	55512	54666	1.0212	1.0057	54306	54375	0.9990	1.0003
Other Renter	41374	41156	1.0637	1.0581	39645	40064	1.0192	1.0300
Other	96886	95822	1.0389	1.0275	93951	94438	1.0075	1.0127
Owner	150057	147867	1.0258	1.0108	155058	155830	1.0600	1.0653
Renter	184436	183982	0.9891	0.9867	206492	10307	.1074	1.1279
Oakland	334493	331849	1.0052	0.9973	361550	366137	1.0866	1.1003

Source: Ikeda (1996)

Similar analysis was performed for Paterson (NJ) and Northwest Louisiana. The results were similar to those obtained for Oakland (CA). However, the differences between HERF based factors and production factors were smaller due to lower noninterview rates in production in these sites.

These two studies conclude that using HERF data to classify noninterviewed households into noninterview adjustment cells brought estimates closer to the Method 1 estimates. Therefore, if an alternative source of good quality is available to provide data on nonresponding households, it is desirable to use it at least to assign these noninterviewed households to appropriate noninterview cells.

4 Methodological research for handling characteristic nonresponse

Characteristic nonresponse is also called item nonresponse. Research in this area for ICM included 1) evaluation of disagreement of imputed and nonimputed characteristics between R-Sample and Census; 2) excluding persons in interviewed households with missing characteristics (item) from estimation; 3) substituting missing items in P-, R-, and E-Samples with those reported in Census; 4) replacing P-, R- and E-Sample person data with Census data reported for that person. A brief summary is presented below.

4.1 Disagreement of imputed and nonimputed characteristics between R-Sample and Census

Since the Census and R-Sample used two different methods for handling missing data, Petroni (1996a and 1996b) investigated the disagreement of characteristics (imputed or non imputed) between R-Sample cases linked to the Census unedited file (CUF) and 1995 Census cases from the HERF. Petroni examined differences in reported and/or imputed race, tenure, age, origin and sex for Oakland. She revealed that there are large disagreements (3%-12%) for race, tenure, origin, and sex in reported and/or imputed characteristics for matched persons. Reported and/or imputed race for Oakland shows that even if the differences in reported and/or imputed race at the micro level were large, in general, 60%-80% of these differences canceled out at the macro level. Similar results hold for tenure, sex and age.

Results from these studies suggest that for the Census Plus estimation methodology we may need to be concerned about response variation and consistency in data collection between ICM and Census even more than about differences in imputation. We pursued three research efforts to get more insight into the effect of disagreement in characteristics between ICM and Census on the C^+ and DSE estimates.

4.2 Persons with missing characteristics treated as not captured in the P-Sample

For this study we treated persons with missing characteristics in interviewed households as not captured in the P-Sample and, hence, excluded them from DSE. Petroni, Kearney, Town and Singh (1995) calculated 357 national level DSE estimates. They compared the original DSE estimates to estimates for the alternative definition of captured persons for race/ethnicity (nonHispanic White, Black, nonBlack Hispanic, Asian and Pacific Islanders, and American Indians), age (0-17, 18-29, 30-49, 50+), sex (Male and Female), and regions (Northeast, South, Midwest, West). The categories in the parenthesis define subgroups of the corresponding group.

They found that:

- A significant percent of alternative estimates differ by more than two percent from the original estimates.
- For all groups except region, the alternative subgroup estimates differ in closeness to the original estimates.
- For all subgroups, except American Indians on Reservations, most estimates for alternative definition are higher than the original estimates.
- For all groups except region and race/Hispanic origin, the subgroups have roughly the same percent of estimates that are higher than the original estimates.

Also, comparison of total person estimates by region, tenure, race/Hispanic origin, sex, age and age by sex showed that estimates from the alternative definition are higher than the original estimates but all are within two percent of the original estimates. Because of the differential effects on total person estimates by region, tenure, race/Hispanic origin, sex and age, it was recommended to treat persons in interviewed households with missing characteristics as captured in the P-Sample.

4.3 Replacement of P-, R-, and E-Sample person imputed data with Census data

To examine the impact of different ICM and Census imputation methods, we linked R-, P-, and E-Sample persons to Census persons and replaced R-, P-, and E-Sample imputed data by Census data (which may or may not have been imputed). Where there was no link, we kept the ICM imputed data. We recomputed DSE adjustment factors when only E-Sample imputed data were replaced with Census linked data, and when both P- and E-Sample imputed data were replaced with Census linked data. We also recomputed C^+ adjustment factors. Table 3 summarizes these factors for Oakland (see columns 3-5).

Table 3: 1995 Dual system and Census plus estimation adjustment factors* for Oakland, California

Characteristics	Original Factors		Factors After Replacing ICM Imputed Data With Census Data			Factors After Replacing ICM Data With Census Data		
	DSE	C^+	DSE Only E Replaced	DSE P and E Replaced	C^+	DSE Only E Replaced	DSE P and E Replaced	C^+
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Oakland	1.0600	1.0052	1.0868	1.0871	1.0052	1.0872	1.0834	1.0069
Owner	1.0600	1.0258	1.0594	1.0587	1.0290	1.0587	1.0599	1.0201
Renter	1.1074	0.9891	1.1081	1.1095	0.9865	1.1096	1.1019	0.9965
Black	1.1047	0.9501	1.1053	1.1062	0.9488	1.1054	1.0991	0.9600
Owner	1.0972	0.9874	1.0981	1.0967	0.9924	1.0952	1.0951	0.9929
Renter	1.1092	0.9277	1.1096	1.1118	0.9226	1.1115	1.1015	0.9402
Hispanic	1.2033	1.1321	1.1977	1.1967	1.1408	1.1978	1.2012	1.1287
Owner	1.1755	1.2117	1.1664	1.1648	1.2253	1.1672	1.1796	1.1954
Renter	1.2194	1.0858	1.2159	1.2152	1.0917	1.2155	1.2138	1.0900
API	1.0748	0.9757	1.0770	1.0777	0.9724	1.0802	1.0602	1.0189
Owner	1.0338	0.9813	1.0301	1.0301	0.9741	1.0306	1.0284	0.9987
Renter	1.1045	0.9717	1.1110	1.1121	0.9711	1.1160	1.0832	1.0335
Other	1.0075	1.0389	1.0089	1.0091	1.0382	1.0084	1.0142	1.0063
Owner	0.9990	1.0212	1.0012	1.0009	1.0240	1.0014	1.0018	0.9987
Renter	1.0192	1.0637	1.0196	1.0204	1.0582	1.0181	1.0315	1.0170

* We use the term adjustment factor loosely. Technically the term implies a factor that would be applied to a Census count to produce an estimate. In this sense, the only adjustments are at the poststratum level. That is, technically the above are not adjustment factors, but these factors minus one represent coverage rates.

Source: Petroni, Kearney, and Ikeda (1996)

For both methods, most of the DSE factors were numerically close to the original method factors. All new alternatives result in substantial differences in factors for 50+ male and female renters. For C^+, Table 3 shows that the alternatives provided approximately the same adjustment factors as the original C^+ factors.

4.4 Replacement of P-, R-, and E-Sample characteristics with Census data

To examine whether differences in data reported for the ICM and Census could be responsible for the differences in DSE and C^+ adjustment factors, we replace R-, P-, and E-Sample data by Census data when ICM and Census data disagreed regardless of whether differences were due to imputation. Again, we recomputed both C^+ and DSE adjustment factors when only E-Sample data are replaced with Census linked data and DSE factors when both P- and E-Sample data are replaced with Census linked data. These factors are presented in Table 3 (see column 6-8).

We compared the original and resulting adjustment factors to analyze the impact of different ICM and Census imputation procedures and the impact of disagreement of Census and ICM reported data. The results for these three methods were similar to those in section 4.3. This is because the large differences in reported/imputed characteristics at the micro level reduce significantly at the macro level and, hence, the adjustment factors at the post-strata level are not affected significantly.

4.5 Comparison of 1990 PES and 1995 ICM imputation methods

One simulation study (Dorinski et al. 1996) compared and evaluated two imputation methods - one used in 1990 and another used in 1995 - for DSE estimation. The items imputed in both the 1990 Census and the 1995 Census Test are tenure, race, Hispanic origin, sex and age. Thus we focus on these items when evaluating the methods. However, the 1990 Census imputation method allows hot-decking based on missing values for relationship and marital status, so we included those items when simulating missing data. The 1995 Census Test did not include group quarters, so we excluded group quarters records from the 1990 E-Sample data file for this study.

To evaluate the two imputation methods, Dorinski et al. compared two measures of success: the number of correct imputations each method produces, or the „closeness" of the marginal distributions produced by each imputation method to the reported marginal distributions. The number of correct imputations is a micro-level measure, while the „closeness" of the marginal distribution to the reported distribution is a macro-level

measure. The focus of the ICM samples is to produce accurate results at aggregate levels so that differential undercounts may be examined. Thus, Dorinski et al. first compared the „closeness" of the marginal distributions to evaluate the methods. Secondarily, they looked at the number of correct imputations each method produces.

Overall, the 1990 method of imputation performs better for the characteristics studied. However, for Hispanic origin, the 1995 method produces slightly better marginals, but fewer correct imputations. Sex imputation also had some problems in both 1990 and 1995 methods.

Further analysis (Dorinski 1996) showed that most of the errors in the 1990 sex imputation occurred when imputing the sex of single householders. Most households with two or more persons are married-couple households, with the husband being listed as the householder almost all the time. This causes the sex distribution of reporting householders to be predominantly male. However, householders who are single tend to be female, so using the overall sex distribution of all reporting householders causes most single householders to be imputed as male. Dorinski suggests that the 1990 sex imputation method for householders based on the presence of a spouse in the household will provide better overall imputation than the 1990 PES imputation method. Sex imputation method for householder should be divided into married householder, householder with no spouse in a single person household, and householder with no spouse in a two or more person household. Householders in each group should be imputed based on the distribution of all reporting householders in that group.

5 Summary and conclusions

From the noninterview adjustment studies, we conclude that we should use the noninterview adjustment to handle whole household noninterviews. Even though the use of Census reported data on the HERF have potential to improve ICM estimates, the improvements were not significant. However, if the nonresponse rate is higher, use of the HERF should be researched. C^+ results show clearly that reducing the last few percent of the household noninterview affected estimates significantly. It is not so obvious for DSE estimates since the noninterview rate dropped to only 5.4%. Reducing nonresponse rates affected C^+ and DSE differently. In general, estimates for C^+ decreased while they increased for DSE when nonresponse was reduced. We believe that, generally, the best approach for handling noninterviews is to reduce the noninterview rate to as low as practical. With a low noninterview rate, a simple adjustment will work as well as a more complex one.

The discrepancy in unimputed and imputed characteristic in the ICM sample and the

Census is large at the micro level. However, the discrepancy at the aggregate level is significantly reduced due to canceling out of some micro-level discrepancies. These discrepancies may have a larger affect on C^+ than on DSE estimates. Results from these studies suggest that for the Census Plus estimation methodology we may need to be concerned about response variation and consistency in data collection between ICM and Census even more than about differences in imputation. The comparison of 1990 and 1995 imputation methodologies showed that, in general, 1990 and 1995 methods provide approximately the same marginal distribution for imputed characteristics except for sex. However, the 1990 method produces more correct imputations. Therefore, we should use the 1990 type imputation for the 2000 Census. For the sex imputation, the 1990 sex imputation method for householders should be divided further by 1) householder with spouse present, 2) householder in a single person household, and 3) householder without spouse in two or more person households.

References

Diffendal, G. and Belin, T. (1991). Results of Procedures for Handling Noninterviews, Missing Characteristics Data, and Unresolved Enumeration Status in 1990 Census/Post-Enumeration Survey. Internal Census Bureau Memorandum, STSD Decennial Census memorandum Series V-112, July 1, 1991

Dorinski, S. M. (1996). Evaluation of Sex Imputation Method Used in 1990 PES. Internal Census Bureau Memorandum for Documentation, September 23, 1996

Dorinski, S., Petroni, R., Ikeda, M. and Singh, R. (1996). Comparison and Evaluation of Alternative ICM Imputation Methods. Proceedings of the Survey Research Methods Section, American Statistical Association

Ikeda, M. (1996). Effect of Using Data From Matching HERF Households to Help Define Noninterview Adjustment Cells for the 1995 ICM. Internal Census Bureau Document dated August 29, 1996

Ikeda, M. and Petroni, R. (1996). Handling of Missing Data in the 1995 Integrated Coverage Measurement Sample. Proceedings of the Survey Research Methods Section, American Statistical Association

Mulry, M. and Singh, R. (1995). Development and Evaluation of Census Methodology for 2000. Proceedings of the International Conference on Survey Measurement and Process Control, 1995

Petroni, R. (1996a). Disagreement of Characteristics Between R-Sample and Census Linked Cases for Oakland - More Findings. Internal Census Bureau Memorandum from Petroni to the Distribution List, May 7, 1996

Petroni, R. (1996b). Disagreement of Characteristics Between R-Sample and Census Linked Cases for Oakland - Preliminary Findings. Internal Census Bureau

Memorandum from Petroni to the Distribution List, March 11, 1996

Petroni, R., Kearney, A. and Gbur, P. (1996). Handling Noninterviews to Provide Equitable Comparisons of ICM Estimates. Proceedings of the Survey Research Methods Section, American Statistical Association

Petroni, R., Kearney, A. and Ikeda, M. (1996). Item Imputation's Effect on 1995 Census Test Estimates. Presented at the 7th International Workshop on Household Survey Nonresponse, Rome, Italy, October, 1996

Petroni, R., Kearney, A., Town, M. and Singh, R. (1995). Should We Account for Missing Data in Dual System Estimation? Presented at the 6th International Workshop on Household Survey Nonresponse, Helsinki, Finland, October, 1995

Schindler, E. (1996). 1995 Census Test Memorandum Series - IS #8, Computer Specifications for ICM Site Level Estimation for the 1995 Census Test - Revision 3. Internal Census Bureau Memorandum from Singh to Thompson, June 18, 1996

Spencer, G. (1995). Draft of 1995 Census Imputation Specifications. Internal Census Bureau Memorandum from Spencer to Philipp, June 19, 1995

Williams, T. (1995a). Methodology Used for the Modeling of Missing Variables in the Flexible Matching Imputation Software. Internal Census Bureau Draft Memorandum for Documentation, July 14, 1995

Williams, T. (1995b). Using the Flexible Matching Imputation Software. Internal Census Bureau Draft Memorandum for Documentation, July 17, 1995

Wolter, K. (1986). Some Coverage Error Models for Census Data. Journal of the American Statistical Association, 81, pp. 338-346

Aspects Concerning Data Fusion Techniques

SUSANNE RAESSLER AND KARLHEINZ FLEISCHER

Abstract: *Data fusion techniques merge data sets of different survey samples by means of statistical matching on the basis of common variables. As a result a virtual sample of complete, but artificial nature is generated. Because being completely unobserved, the missing information of an individual in one sample is imputed using the observed data values of some individual which is found to be most similar in the other sample. The power of data fusion techniques is analysed and the parameters of the distributions of all variables in the artificial sample are formulated. The correlation between variables not jointly observed, which can only be estimated by means of the matched file, is of main interest herein. Furthermore the influences of nearest neighbour matches, several so–called marriage processes, and small sample sizes are the focus of simulation studies.*

Key words: *missing information, imputation, merging data sets, statistical matching.*

1 Introduction

Empirical studies concerning the association between individual television viewing and purchasing behaviour, for instance, occure to be difficult in the majority of cases. The ideal medium would be a very large consumer panel, where each individual's purchasing and television viewing behaviour would both be measured. However, the costs of running a large single source panel of this kind are prohibitively high. Furthermore, a high percentage of nonresponses or poor quality of data are to be expected. A powerfully attractive alternative is to make use of a data fusion technique to link together, for example the viewing information available from a television measurement panel with the purchasing data available from an existing large market tracking panel.

Especially in the area of media analyses, data fusions have been performed in France and the UK with a reasonable degree of accuracy as published by Antoine

(1987), Baker (1990) and Roberts (1994). Further descriptions of data fusions done in practice could be found by Okner (1972a and b), Okner (1974), Ruggles and Ruggles (1974) or Scheler and Wiegand (1987). On the other hand there is scepticism among theoretical and practical statisticians about the power of fusion techniques (see Sims (1972a and b), Bennike (1987) or Gabler (1997)). Only a few publications, for instance, Woodbury (1983), Sims (1972a and b), Kovacevic and Liu (1994) or Wiedenbeck (1995), are known to mathematically study fusion algorithms and investigate their efficiency under certain circumstances.

This paper analyses the power of some data fusion techniques. Their influence on the estimated joint distribution of the variables not jointly observed will be shown herein with the help of mathematical methods as well as simulation studies.

2 Fusion algorithm

Data fusion is initiated by two samples, one usually of larger size than the other, with the number of individuals appearing in both samples (i.e. the overlap) clearly negligible. Only certain variables, say Z, of the interesting individual's characteristics can be observed in both samples; they are called common variables. Other variables, Y, appear only in the larger sample while others, X, are observed exclusively in the smaller sample. (For generalization purpose X, Y, Z can be treated as vectors of variables.) Since no single sample exists with information on X, Y and Z together, an artificial sample has to be generated by matching the observations of both samples according to Z. The matching is performed at an individual level by means of statistical matching; this is often called the marriage process.

Without a loss of generality, let the smaller (X, Z) sample be the so–called recipient sample and the larger (Y, Z) sample the donor sample. For every unit i of the recipient sample with the observations (x_i, z_i) a value y from the observations of the donor sample is determined and a data set $(x_1, y_1, z_1), \ldots, (x_{n_E}, y_{n_E}, z_{n_E})$ is constructed with n_E elements of the recipient sample. The main idea is to search for a statistical match, i.e. a donor unit j whose observed data values of the common variables are identical to those of the recipient unit i.

As long as the overlap is poor, there is little chance to find a perfect match for each individual, especially if (some) common variables are continuous. Described by Baker (1990), Roberts (1994) or Antoine (1987), the marriage process is carried out using an algorithm based on nearest neighbour techniques calculated by means of a distance measure $d(.,.)$. The marriage algorithm may use all or some of the

common variables, weighted or not, to find for each recipient unit i one (or more) donor unit(s) j whose distance $d(z_i, z_j)$ is minimal. By restricting the multiple choice of a donor for different recipients, further variations on the algorithm can be created. To limit the number of times a donor is taken, a penalty weight may be placed on donors already used, as the multiple choice will otherwise reduce the effective sample size and lead to underestimation of the true variance. Another modification is to take the next three (or any other number) donors and impute their (weighted) mean. If the multiple use of donors is restricted or combined with a penalty function, the resulting artificial sample, i.e. the fusion sample, may vary depending on the order the donor units are taken. Other algorithms are known to limit this problem by, for instance, cross-checking all matches after fusion and sometimes abandoning certain matches in order to find a better donator–recipient combination afterwards. Antoine (1987) gives a short description of such algorithms.

2.1 Distributions computed by fusion

In the following all density functions (joint, marginal or conditional) and their parameters produced by the fusion algorithm will be marked by the symbol $\tilde{}$.

Let X, Y, Z be multivariate random variables with joint discrete or continuous density function $f_{X,Y,Z}$. Thus, for discrete variables, $f_{X,Y,Z}(x_i, y_i, z_i)$ describes the probability to draw a certain unit i with observation (x_i, y_i, z_i) and for continuous variables it is the value of the joint density function at the point (x_i, y_i, z_i). To keep things simple, only the expression "probability" will be used hereafter. In case of continuous variables, f as the density function instead of the probability function may be taken.

If the units of the two samples are drawn independently from each other, the distribution of the donor sample of size n_S is $\prod_{i=1}^{n_S} f_{Y,Z}(y_i, z_i)$ and likewise the recipient sample is distributed with probability function $\prod_{i=1}^{n_E} f_{X,Z}(x_i, z_i)$.

Furthermore, let the fusion algorithm be one of multiple choice of the donor units without any penalty function. Thus the units of the artificial sample can be treated as being drawn independently with probability $\tilde{f}_{X,Y,Z}(x, y, z)$ each. The fusion algorithm therefore induces the probability distribution $\prod_{i=1}^{n_E} \tilde{f}_{X,Y,Z}(x_i, y_i, z_i)$ on the set of all possible fusion samples. They can be handled as simple random samples drawn from an artificial population with distribution $\tilde{f}_{X,Y,Z}(x, y, z)$, which may

be called the "fusion distribution".

Often the fusion sample is used to estimate parameters (such as means, variances, covariances or higher moments) of the "initial" population following $f_{X,Y,Z}(x,y,z)$ with traditional methods. To judge the quality of such estimates, which in fact means the power of the fusion, the relation between $\tilde{f}_{X,Y,Z}(x,y,z)$ and $f_{X,Y,Z}(x,y,z)$ has to be examined.

2.1.1 Distribution of the artificial sample

As already specified, let the probability to get a particular unit i after the fusion with observation (x_i, y_i, z_i) be $\tilde{f}_{X,Y,Z}(x_i, y_i, z_i)$. This is equivalent to the probability of drawing a particular unit i of the recipient sample with observation (x_i, z_i) and merging this unit with a unit j from the donor sample with observed values (y_j, z_j), where $z_i = z_j$. The probability for a donor unit j with observed value $z_j = z_i$ from Z to have the observation y_j from Y is obviously $f_{Y|Z}(y_j|z_j)$.

Hence the probability to observe (x, y, z) for any unit of the fusion sample is

$$\tilde{f}_{X,Y,Z}(x,y,z) = f_{X,Z}(x,z) f_{Y|Z}(y|z) \tag{1}$$

provided that donor and recipient sample have been drawn independently from the same population. Thus,

$$\begin{aligned} \tilde{f}_{X,Y,Z}(x,y,z) &= f_{X,Z}(x,z) f_{Y|Z}(y|z) = f_{X|Z}(x|z) f_Z(z) f_{Y|Z}(y|z) \\ &= f_{X|Z}(x|z) f_{Y,Z}(y,z) \end{aligned} \tag{2}$$

and the conditional distribution is given by

$$\tilde{f}_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z)\, f_{Y|Z}(y|z)\,. \tag{3}$$

It should be noted that this derivation is only admissible if, for every recipient unit, there is a donor unit with the same observed value z for Z. Especially in case of continuous distributions, this will not happen often, and a nearest neighbour unit in z has to be merged. The influence of such a nearest neighbour match on the computed distribution after the fusion will be discussed on the basis of simulation studies as well hereinafter.

2.1.2 Marginal distributions after the fusion

The marginal distributions of $\widetilde{f}_{X,Y,Z}$ are now easily obtained with the help of (2):

$$\begin{aligned}\widetilde{f}_X(x) &= \int\int \widetilde{f}_{X,Y,Z}(x,y,z)dydz = \int f_{X,Z}(x,z)\int f_{Y|Z}(y|z)dydz = \\ &= f_X(x) \qquad (4)\end{aligned}$$

as is $\widetilde{f}_Y(y) = f_Y(y)$ and $\widetilde{f}_Z(z) = f_Z(z)$. Furthermore

$$\begin{aligned}\widetilde{f}_{X,Z}(x,z) &= \int \widetilde{f}_{X,Y,Z}(x,y,z)dy = f_{X,Z}(x,z)\int f_{Y|Z}(y|z)dy = \\ &= f_{X,Z}(x,z) \qquad (5)\end{aligned}$$

and also $\widetilde{f}_{Y,Z}(y,z) = f_{Y,Z}(y,z)$. Different from their initial distributions are

$$\widetilde{f}_{X,Y}(x,y) = \int \widetilde{f}_{X,Y,Z}(x,y,z)dz = \int f_{X|Z}(x|z)f_Z(z)f_{Y|Z}(y|z)dz, \quad (6)$$

$$\begin{aligned}\widetilde{f}_{X,Y,Z}(x,y,z) &= f_{X,Z}(x,z)f_{Y|Z}(y|z) = \\ &= f_{X,Z}(x,z)f_{Y|X,Z}(y|x,z)\,\frac{f_{Y|Z}(y|z)}{f_{Y|X,Z}(y|x,z)} \\ &= f_{X,Y,Z}(x,y,z)\,\frac{f_{Y|Z}(y|z)}{f_{Y|X,Z}(y|x,z)} = \\ &= f_{X,Y,Z}(x,y,z)\,\frac{f_{X|Z}(x|z)}{f_{X|Y,Z}(x|y,z)} \qquad (7)\end{aligned}$$

Thus, the distribution of X, Y, Z after the fusion is equal to the initial distribution if X and Y are independent, conditional on every possible value z of Z, i.e.

$$f_{Y|X,Z}(y|x,z) = f_{Y|Z}(y|z) \quad \text{or equivalent} \quad f_{X|Y,Z}(x|y,z) = f_{X|Z}(x|z).$$

Especially Sims (1972a and b) called for conditional independence as a main assumption for a reliable fusion.

Moreover, all marginal distributions, which could have been estimated already by the two separate samples, are identical before and after the fusion. Only the joint distributions of variables not jointly observed are different. Note that all derivations above are valid for vectors of random variables X, Y, Z as well. Accordingly all moments concerning variables of one or the other sample are identical for the fusion distribution and the initial distribution. See for instance $\widetilde{\mu}_X = \mu_X$,

$\widetilde{\mathrm{E}}(X^i) = \mathrm{E}(X^i)$, $\widetilde{\sigma}_X^2 = \sigma_X^2$ and so on. Thus testing the accuracy of the fusion by properties of the marginal distributions of variables observed in one of the two samples is by no means useful for validation of fusion results.

Consider now the moments of the joint distribution of variables from different samples. The correlation between X and Y generated by the fusion, measured as covariance $\widetilde{\mathrm{Cov}}(X,Y)$, is

$$\widetilde{\mathrm{Cov}}(X,Y) = \mathrm{Cov}(X,Y) - \mathrm{E}(\mathrm{Cov}(X,Y|Z)). \tag{8}$$

This result, however, can easily be obtained by first calculating $\mathrm{E}(\mathrm{Cov}(X,Y|Z))$:

$$\begin{aligned}
\mathrm{E}(\mathrm{Cov}(X,Y|Z)) &= \int [\mathrm{E}(X\cdot Y|Z=z) - \mathrm{E}(X|Z=z)\mathrm{E}(Y|Z=z)]\, f_Z(z)dz \\
&= \int \left[\int\int xyf_{X,Y|Z}(x,y|z)dxdy \right. \\
&\quad \left. - \int xf_{X|Z}(x|z)dx \int yf_{Y|Z}(y|z)dy\right] f_Z(z)dz \\
&= \int\int\int xyf_{X,Y|Z}(x,y|z)f_Z(z)dxdydz \\
&\quad -\int\int\int xyf_{X|Z}(x|z)f_{Y|Z}(y|z)f_Z(z)dxdydz \\
&= \int\int\int xyf_{X,Y,Z}(x,y,z)dxdydz \\
&\quad -\int\int\int xy\widetilde{f}_{X,Y,Z}(x,y,z)dxdydz \\
&= \mathrm{E}(XY) - \widetilde{\mathrm{E}}(XY). \qquad (9)
\end{aligned}$$

$$\begin{aligned}
\text{Thus}\quad \widetilde{\mathrm{E}}(XY) &= \mathrm{E}(XY) - \mathrm{E}(\mathrm{Cov}(X,Y|Z)) \quad \text{and} \\
\widetilde{\mathrm{E}}(XY) - \mu_X\mu_Y &= \mathrm{E}(XY) - \mu_X\mu_Y - \mathrm{E}(\mathrm{Cov}(X,Y|Z)) \text{ and because of } \mu = \widetilde{\mu}. \\
\widetilde{\mathrm{Cov}}(X,Y) &= \mathrm{Cov}(X,Y) - \mathrm{E}(\mathrm{Cov}(X,Y|Z)) \quad q.e.d.
\end{aligned}$$

This leads to

$$\mathrm{E}(\mathrm{Cov}(X,Y|Z)) = \mathrm{Cov}(X,Y) - \widetilde{\mathrm{Cov}}(X,Y), \tag{10}$$

i.e. the average covariance of X and Y is just the difference of the covariances from the initial, the real distribution and the fusion distribution. It may be used

as a quality measurement of the fusion. The closer this value gets to zero, the better the true correlation is reproduced by the fusion.

In analogy to (9) it is quite simple to show, that

$$\begin{aligned} \mathrm{E}(\mathrm{Cov}(X^i, Y^j|Z)) &= \mathrm{E}(X^iY^j) - \widetilde{\mathrm{E}}(X^iY^j) \quad \text{and} && (11)\\ \widetilde{\mathrm{Cov}}(X^i, Y^j) &= \mathrm{Cov}(X^i, Y^j) - \mathrm{E}(\mathrm{Cov}(X^i, Y^j|Z)) \quad (i, j \in \mathbb{N}). && (12) \end{aligned}$$

Accordingly the fusion can produce "good results" concerning the true correlation between the variables X and Y never jointly observed only if they are **on the average conditionally uncorrelated**, i.e. $\mathrm{E}(\mathrm{Cov}(X, Y|Z)) = 0$. The same applies to higher moments. Therefore the independence of X und Y conditional on Z, as postulated by Sims (1972a and 1972b), is sufficient but not necessary.

2.2 Application on certain distributions

Under the assumption of a multivariate normal distribution for the joint distribution of X, Y, Z Wiedenbeck (1995), has shown the following results after the fusion, independent of the real correlation $\mathrm{Cov}(X, Y)$:

$$\widetilde{\mathrm{Cov}}(X, Y) = \widetilde{\sigma}_{X,Y} = \frac{\sigma_{X,Z}\sigma_{Y,Z}}{\sigma_Z^2}. \tag{13}$$

Using the expression (10) this leads to

$$\mathrm{E}(\mathrm{Cov}(X, Y|Z)) = \mathrm{Cov}(X, Y) - \widetilde{\mathrm{Cov}}(X, Y) = \sigma_{X,Y} - \frac{\sigma_{X,Z}\ \sigma_{Y,Z}}{\sigma_Z^2}.$$

Thus, after the fusion process the variables X and Y are computed uncorrelated without respect to their initial correlation, if X, Z or Y, Z are uncorrelated. Otherwise, if X, Z and Y, Z are correlated, then X, Y are computed correlated as well, although they may be uncorrelated initially.

Consider now X, Y, Z as being transformed via (e^X, e^Y, e^Z) to lognormally distributed random variables whose means, variances and covariance are specified by

$$\begin{aligned} \mu_X^\star &= e^{\mu_X + 0.5\sigma_X^2}\\ \sigma_X^{2\star} &= e^{2\mu_X + \sigma_X^2}(e^{\sigma_X^2} - 1) = \mu_X^{\star\,2}(e^{\sigma_X^2} - 1)\\ \sigma_{X,Y}^\star &= e^{\mu_X + 0.5\sigma_X^2 + \mu_Y + 0.5\sigma_Y^2}(e^{\rho_{X,Y}\sigma_X\sigma_Y} - 1) = \mu_X^\star\mu_Y^\star(e^{\sigma_{X,Y}} - 1) \end{aligned}$$

and so on. Let the parameters of the distribution of transformed variables based on a normal distribution be marked by the symbol *. Finally the reproduced correlation of X and Y is given by

$$\begin{aligned}\widetilde{\mathrm{Cov}}(e^X, e^Y) &= \widetilde{\sigma}^\star_{X,Y} = \mu^\star_X \mu^\star_Y \left(e^{\sigma_X \sigma_Y \frac{\sigma_{X,Z}\sigma_{Y,Z}}{\sigma_Z^2}} - 1 \right) = \\ &= \mu^\star_X \mu^\star_Y \left(e^{\sigma_X^2 \sigma_Y^2 \rho_{X,Z} \rho_{Y,Z}} - 1 \right) \qquad (14)\end{aligned}$$

In general, it could be difficult to calculate the exact formulas of the covariance reproduced by the fusion algorithm. Therefore, the investigation hereinafter will be if and to what degree of accuracy the presented results can be computed by simulation.

3 An experimental design

As mentioned before the fusion distribution was derived assuming the existence of a donor unit with identical z–values for every recipient unit. Since it is common practice to use several sociodemographical variables often combined with other continuous variables as common Z variables, the above assumption is most unlikely.

Hence the simulation study in the following is performed to consider the accuracy of the fusion distribution and estimators such as means, variances and covariances derived from it and affected by different continuous variables. Likewise, the influence of nearest neighbour matches, different marriage processes, and varying sample sizes will be discussed.

To keep it simple, the simulation study is limited to trivariate normal and lognormal distributions.

3.1 Random number generation

To generate random numbers considered as realizations of a standard normal distribution, a simple random number generator *randn()* of the MATLAB 4.0 program is used. All the programs needed for the simulation study have been created in the matrix programming language MATLAB 4.0 which is a product and trademark of The Math Works, Inc.

Based on standard normally distributed variables U, V, W produced by *randn()*, the multivariate normally distributed variables X, Y, Z are given by

$$\begin{aligned}(U,V,W) &\sim N(0,I)\\ Z &= U\cdot\sigma_Z+\mu_Z\\ X &= V\cdot\sigma_{X|Z}+\mu_{X|Z}\\ Y &= W\cdot\sigma_{Y|Z,X}+\mu_{Y|Z,X}\end{aligned}$$

assisted by their conditional distributions, for further notes see Johnson (1987), p. 50. Realizations of lognormally distributed random variables are easy to obtain by calculating (e^X, e^Y, e^Z).

3.2 Marriage processes

The statistical match or marriage process is carried out by using a simple nearest neighbour algorithm first. For each recipient unit i, the missing information Y is imputed taken from the donor unit j whose distance $|z_i - z_j|$ is minimal. The donor sample size is twice the recipient sample size with $n_S = 10000$ and $n_E = 5000$. A donor unit can be used many times without restrictions for other marriages; this process may be called "polygamy".

Moreover, the sample sizes are reduced considerably. Further variations of the algorithm are dealt with by restricting the multiple choice of the donor units. Thus the following experimental design results:

(1) "Polygamy", i.e. any multiple use of donor units is allowed with

- $n_S = 1000, n_E = 500$ and
- $n_S = 500, n_E = 500$.

(2) "Bigamy", i.e. any donor unit can be used twice only with

- $n_S = 1000, n_E = 500$ and
- $n_S = 500, n_E = 500$.

(3) "Monogamy", i.e. any donor unit can be used once only with

- $n_S = 1000, n_E = 500$ and
- $n_S = 500, n_E = 500$.

If the data sets are sorted, the order of sampling donor units influences the resulting fusion sample. Since the sample units are ordered at random, no special effect is to be expected by sampling one unit after the other.

(4) "Free triple", i.e. imputing the observations' mean of the next three donor units allowing multiple choice with

- $n_S = 1000, n_E = 500$ and
- $n_S = 500, n_E = 500$.

Since the variance of the mean of n observations is no longer identical with the variance of the population, it is not possible to reproduce even the true variance of Y by the fusion. In case of normal distributed Y and Z variables, the reproduced variance of Y is now given by

$$\begin{aligned}
\widetilde{\mathrm{Var}}(Y) &= \mathrm{E}(\widetilde{\mathrm{Var}}(Y|Z)) + \mathrm{Var}(\widetilde{\mathrm{E}}(Y|Z)) = \mathrm{E}(\mathrm{Var}(\overline{Y}|Z)) + \mathrm{Var}(\mathrm{E}(\overline{Y}|Z)) \\
&= \mathrm{E}\left(\frac{\sigma_Y^2}{n}\,(1-\rho_{Y,Z}^2)\right) + \mathrm{Var}\left(\mu_Y + \rho_{Y,Z}\,\frac{Z-\mu_Z}{\sigma_Z}\,\sigma_Y\right) \\
&= \frac{\sigma_Y^2}{n}\,(1-\rho_{Y,Z}^2) + \rho_{Y,Z}^2\,\frac{\sigma_Z^2}{\sigma_Z^2}\,\sigma_Y^2 = \sigma_Y^2\left(\frac{1}{n}\,(1-\rho_{Y,Z}^2) + \rho_{Y,Z}^2\right) \\
&= \sigma_Y^2\left(\frac{1}{n} + \rho_{Y,Z}^2\left(1-\frac{1}{n}\right)\right) = \frac{\sigma_Y^2}{n}\left(1+(n-1)\rho_{Y,Z}^2\right)
\end{aligned}$$

Thus, with the assumption of normal distributions for the free triple the following is true:

$$\widetilde{\mathrm{Var}}(Y) = \frac{\sigma_Y^2}{3}(1+2\rho_{Y,Z}^2), \tag{15}$$

which does not match the variance σ_Y^2 of the initial population.

3.3 Simulation of the reproduced covariance

Now n_E random variables (X_i, Z_i) and n_S random variables (Y_j, Z_j) are generated independently due to a given trivariate normal distribution or its transformations with mean vector μ and covariance structure Σ. This could be done either in accordance with their marginal distributions or by generating a sample of the trivariate distribution and splitting it at random. Then the two samples are merged in accordance with the specified algorithms. The empirical covariance, i.e. the

estimate of the reproduced covariance, is calculated by

$$\widehat{\widetilde{\sigma}}_{X,Y} = \frac{1}{n_E - 1} \sum_{i=1}^{n_E} (x_i - \overline{x})(y_i - \overline{y})$$

from the fusion sample. This procedure is repeated k times. In this manner the estimated mean and variance of the empirical covariance are obtained from the simulated distribution. In particular, for every considered distribution

$$\widehat{\mathrm{E}}(\widehat{\widetilde{\sigma}}_{X,Y}) = \frac{1}{k} \sum_{i=1}^{k} \widehat{\widetilde{\sigma}}_{X,Y_i}, \quad s^2(\widehat{\widetilde{\sigma}}_{X,Y}) = \frac{1}{k-1} \sum_{i=1}^{k} \left(\widehat{\widetilde{\sigma}}_{X,Y_i} - \widehat{\mathrm{E}}(\widehat{\widetilde{\sigma}}_{X,Y}) \right)^2 \tag{16}$$

and $\widetilde{\sigma}_{X,Y} - \widehat{\mathrm{E}}(\widehat{\widetilde{\sigma}}_{X,Y})$ are calculated. To assure the accuracy of the simulation, to what extend $\widehat{\mathrm{E}}(\widehat{\widetilde{\sigma}}_{X,Y})$ and $\widetilde{\sigma}_{X,Y}$ agree is checked since the fusion only reproduces this covariance and not $\sigma_{X,Y}$. Furthermore, a t–statistic like value is computed with $t = \frac{\widehat{\mathrm{E}}(\widehat{\widetilde{\sigma}}_{X,Y}) - \widetilde{\sigma}_{X,Y}}{s(\widehat{\widetilde{\sigma}}_{X,Y})} \sqrt{k}$ to ease interpretation.

As the true variance of the imputed variable Y is changed by use of the free triple, the value of $\widehat{\mathrm{E}}(\widehat{\widetilde{\sigma}}^2_Y)$ is tabulated as well when using this algorithm.

To get results in reasonable time, the simulation has to be restricted to $k = 100$. On a 100 MHz pentium computer it takes about 4.2 hours to generate one empirical distribution with $n_S = 10000$.

4 Results of the simulation

4.1 Reproduced covariances

The simulation study is done with the following parameter set for the normal distribution

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{und} \quad \Sigma = \begin{pmatrix} 1 & \sigma_{X,Y} & \sigma_{X,Z} \\ \sigma_{Y,X} & 1 & \sigma_{Y,Z} \\ \sigma_{Z,X} & \sigma_{Z,Y} & 1 \end{pmatrix} \tag{17}$$

That means $\sigma_{.,.} = \rho_{.,.}$; the parameters $\rho_{X,Z} = \rho_{Y,Z} = 0$ and 0.5 together with $\rho_{X,Y} = 0.1, 0.5$ and 0.9 each are assumed.

Using the multivariate lognormal distribution, it is possible to specify mean vectors

and covariance structures such as shown by Johnson (1987), p. 83. As mentioned before, the parameters of the lognormal distribution are marked by the symbol *.

As shown in detail below, the results are quite stable despite the rather small simulation size of $k = 100$. The sample sizes are $n_S = 2n_E = 10000$ allowing multiple choice of donor units, i.e. polygamy. The different distributions used and the real correlations between X and Y have no influence on the reproduced covariances. Likewise the reproduced covariances are uninfluenced by the need to merge nearest neighbours instead of donor units identical in Z.

Table 1: Normal distribution with $n_S = 2n_E = 10000$ using polygamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\rho_{X,Y}$	$\tilde{\sigma}_{X,Y}$	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\widehat{\tilde{\sigma}}_{X,Y})}$	t
0	0.1	0	0.0012	0.0147	0.836
0	0.5	0	−0.0031	0.0149	−2.104
0	0.9	0	0.0026	0.0143	1.818
0.5	0.1	0.25	0.2469	0.0144	−2.161
0.5	0.5	0.25	0.2524	0.0174	1.365
0.5	0.9	0.25	0.2526	0.0148	1.739

Table 2: Lognormal distribution with $n_S = 2n_E = 10000$ using polygamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\rho_{X,Y}$	$\tilde{\sigma}^\star_{X,Y}$	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}^\star_{X,Y})$	$\sqrt{s^2(\widehat{\tilde{\sigma}}^\star_{X,Y})}$	t
0	0.1	0	0.0090	0.0747	1.199
0	0.5	0	−0.0231	0.0594	−3.880
0	0.9	0	0.0074	0.0629	1.181
0.5	0.1	0.7721	0.7553	0.1204	−1.388
0.5	0.5	0.7721	0.7903	0.1331	1.369
0.5	0.9	0.7721	0.7760	0.1352	0.294

As mentioned before, the true correlation $\rho_{X,Y}$ has no influence on the correlation generated by the fusion.

4.2 Influences of the marriage processes and the sample sizes

Even reducing the rather large sample sizes of recipient and donor samples to only $n_S = 1000$ and $n_E = 500$ does not affect the results. The same holds when several

marriage processes are considered as is reported in the following tables. Since the true correlation $\rho_{X,Y}$ is of no influence, the simulation is done via k recipient and k donor samples generated for different $\rho_{X,Z}$ and $\rho_{Y,Z}$ values.

Table 3: Normal distribution with $n_S = 2n_E = 1000$ using polygamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\hat{\mathrm{E}}(\hat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\hat{\tilde{\sigma}}_{X,Y})}$	t
0	0	−0.0094	0.0409	−2.293
0.2	0.04	0.0315	0.0417	−2.042
0.4	0.16	0.1573	0.0509	−0.528
0.6	0.36	0.3669	0.0497	1.379
0.8	0.64	0.6430	0.0601	0.501

Table 4: Normal distribution with $n_S = n_E = 500$ using polygamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\hat{\mathrm{E}}(\hat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\hat{\tilde{\sigma}}_{X,Y})}$	t
0	0	0.0041	0.0487	0.839
0.2	0.04	0.0468	0.0434	1.574
0.4	0.16	0.1600	0.0533	0.001
0.6	0.36	0.3624	0.0535	0.443
0.8	0.64	0.6397	0.0587	−0.044

Table 5: Normal distribution with $n_S = 2n_E = 1000$ using bigamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\hat{\mathrm{E}}(\hat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\hat{\tilde{\sigma}}_{X,Y})}$	t
0	0	−0.0029	0.0414	−0.694
0.2	0.04	0.0325	0.0458	−1.628
0.4	0.16	0.1587	0.0457	−0.283
0.6	0.36	0.3553	0.0540	−0.876
0.8	0.64	0.6387	0.0584	−0.220

Table 6: Normal distribution with $n_S = n_E = 500$ using bigamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\widehat{\tilde{\sigma}}_{X,Y})}$	t
0	0	−0.0140	0.0505	−2.764
0.2	0.04	0.0372	0.0429	−0.647
0.4	0.16	0.1568	0.0498	−0.641
0.6	0.36	0.3604	0.0501	0.085
0.8	0.64	0.6354	0.0433	−1.058

Table 7: Normal distribution with $n_S = 2n_E = 1000$ using monogamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\widehat{\tilde{\sigma}}_{X,Y})}$	t
0	0	−0.0001	0.0416	−0.028
0.2	0.04	0.0354	0.0454	−1.013
0.4	0.16	0.1573	0.0435	−0.613
0.6	0.36	0.3579	0.0500	−0.419
0.8	0.64	0.6338	0.0495	−1.263

Table 8: Normal distribution with $n_S = n_E = 500$ using monogamy

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\widehat{\tilde{\sigma}}_{X,Y})}$	t
0	0	0.0024	0.0510	0.470
0.2	0.04	0.0439	0.0504	0.770
0.4	0.16	0.1510	0.0486	−1.856
0.6	0.36	0.3414	0.0482	−3.858
0.8	0.64	0.6007	0.0467	−8.422

Table 10: Normal distribution with $n_S = 2n_E = 1000$ using free triple

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\widehat{\tilde{\sigma}}_{X,Y})}$	t	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}^2_Y)$	$\tilde{\sigma}^2_Y$
0	0	−0.0011	0.0238	−0.465	0.3342	0.3333
0.2	0.04	0.0363	0.0261	−1.409	0.3542	0.3599
0.4	0.16	0.1589	0.0343	−0.334	0.4360	0.4400
0.6	0.36	0.3662	0.0394	1.566	0.5729	0.5733
0.8	0.64	0.6376	0.0546	−0.446	0.7549	0.7600

Table 11: Normal distribution with $n_S = n_E = 500$ using free triple

$\rho_{X,Z} = \rho_{Y,Z}$	$\tilde{\sigma}_{X,Y}$	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}_{X,Y})$	$\sqrt{s^2(\widehat{\tilde{\sigma}}_{X,Y})}$	t	$\widehat{\mathrm{E}}(\widehat{\tilde{\sigma}}^2_Y)$	$\tilde{\sigma}^2_Y$
0	0	0.0047	0.0285	1.657	0.3321	0.3333
0.2	0.04	0.0442	0.0307	1.381	0.3615	0.3599
0.4	0.16	0.1594	0.0380	−0.163	0.4373	0.4400
0.6	0.36	0.3545	0.0432	−1.269	0.5743	0.5733
0.8	0.64	0.6297	0.0523	−1.976	0.7461	0.7600

Again the simulation results turn out very stable and match the theoretical results quite well. Therefore, it seems not necessary to examine more combinations such as $\rho_{X,Z} \neq \rho_{Y,Z}$ or positive/negative ρ.

Few absolute t–values are greater than 2. Neglecting the case of monogamy, this seems not to be systematic but a matter of small k. The different marriage processes show no further influence on the reproduced covariances as long as the donor sample is twice the recipient sample (or just larger). Even the very small sample sizes do not much affect the results. Only when using monogamy and identical sample sizes, meaning that every donor unit is used once, are bigger differences reported. If recipient and donor sample sizes are similar, this algorithm is of no practical use, of course. Furthermore, using the free triple the reproduced covariance is simulated just as well as the reproduced variance; for evidence see (15). Even this algorithm is not able to reproduce the true correlation between X and Y nor the true variance of Y.

5 Conclusions

The results are obvious. Fusion of data sets using such rather simple algorithms can reproduce the true correlation between variables X and Y not jointly observed if and only if they are uncorrelated on the average conditional on the common variable Z, i.e. if $\mathrm{E}(\mathrm{Cov}(X, Y|Z)) = 0$.

The stronger demand for conditional independence is not necessary if the interest is focused on the correlation (or higher moments) between X and Y only.

In general, the parameters reproduced by the fusion are not affected by merging nearest neighbour units instead of statistical twins. The influence of several marriage processes on the reproduced parameters is likewise low. The free triple (or any mean of n observations) should not be used if inference is done without correcting the variance of Y reproduced by the fusion. Finally, the sample sizes are not important, but the donor sample should be of larger size than the recipient sample if multiple use of donor units is restricted anyway.

References

Antoine, J. (1987). A Case Study Illustrating the Objectives and Perspectives of Fusion Techniques. In: Readership Research: Theory and Practice, ed. Henry, H., Amsterdam: Elsevier Science Publishers, 336–351.

Baker, K. (1990). The BARB/TGI Fusion. Technical Report on the Fusion Conducted in February/March 1990, Ickenham, Middlesex: Ken Baker Associates.

Bennike, S. (1987). Fusion — An Overview by an Outside Observer. In: Readership Research: Theory and Practice, ed. Henry, H., Amsterdam: Elsevier Science Publishers, 334–335.

Gabler, S. (1997). Datenfusion. Mannheim: ZUMA-Nachrichten, 40, 81–92.

Johnson, M.E. (1987). Multivariate Statistical Simulation. New York: John Wiley and Sons.

Kovacevic, M.S. and Liu, T. (1994). Statistical Matching of Survey Datafiles: A Simulation Study. Proceedings of the Section on Survey Research Methods, American Statistical Association, 497–484.

Okner, B.A. (1972a). Constructing a New Data Base from Merging Microdata Sets: The 1966 Merge File. Annals of Economic and Social Measurement, 1, 325–341.

Okner, B.A. (1972b). Reply and Comments. Annals of Economic and Social Measurement, 1, 359–362.

Okner, B.A. (1974). Data Matching and Merging: An Overview. Annals of Economic and Social Measurement, 3, 347–352.

Roberts, A. (1994). Media Exposure and Consumer Purchasing: An Improved Data Fusion Technique. Marketing and Research Today, 150–172.

Ruggles, N. and Ruggles, R. (1974). A Strategy for Merging and Matching Microdata Sets. Annals of Economic and Social Measurement, 3, 353–371.

Scheler, H.-E. and Wiegand, J. (1987). A Report on Experiments in Fusion in the 'Official' German Media Research (AG.MA). In: Readership Research: Theory and Practice, ed. Henry, H., Amsterdam: Elsevier Science Publishers, 352–360.

Sims, C.A. (1972a). Comments. Annals of Economic and Social Measurement, 1, 343–345.

Sims, C.A. (1972b). Rejoinder. Annals of Economic and Social Measurement, 1, 355–357.

Wiedenbeck, M. (1995). Datenfusion in normalverteilten Populationen. Mannheim: Unpublished paper.

Woodbury, M.A. (1983). Statistical Record Matching for Files. In: Incomplete Data in Sample Surveys, 3, Proceedings of the Symposium, eds. Madow, W.G. and Olkin, I., New York: Academic Press, 173–181.

A Conditional Minimax Estimator for Treating Nonresponse

SIEGFRIED GABLER AND SABINE HÄDER

***Abstract**: Losses due to nonresponse may lead to systematic biases in the samples which result in biased estimates. A usual way to compensate for this bias is the adjustment of the net sample to known population data. We choose another approach. If auxiliary information is available for each individual of the gross sample we adjust the net sample to the gross sample. The advantage is - in contrast to the usual post-stratification - that each element of the net sample may get an „own" weight. For the construction of these weights the conditional minimax principle is applied. It determines optimal weights conditional on the selected net sample which minimizes the maximal loss under the assumption that a certain variance of the unknown population values in the gross sample is finite. The effects of our weighting procedure are shown for data of the German General Social Survey (ALLBUS) 1996.*

***Keywords**: ALLBUS, BLU estimator, conditional minimax, nonresponse, parameter space, representative estimator*

1 Introduction

The response rates in German social surveys declined during the last decades. Nowadays they have reached a level of about 60 percent. Typical for academic surveys is a wide range of nonresponse rates from about 20 to about 50 percent - depending on the conducting commercial institute and the subject of the survey (Schnell 1996). Unfortunately, we cannot assume on principle that the resulting net samples are unbiased. That means, that the losses of the gross samples are frequently not at random but systematic. A usual way to compensate for this nonresponse bias is the adjustment of the net sample to known population data (Little 1989, Deville and Särndal 1992, Deville, Särndal and Sautory 1993, Gabler and Häder 1997, Häder and Gabler 1997). Elliot (1991, 1996) discusses pros and cons of various procedures for post-stratification. We want to demonstrate another approach of correction for nonresponse bias.

The application of post-stratification means to classify the net sample into multivariate cells for which we have reference data of the population in the form of some marginal distributions.The weights are constructed to adjust the cell frequencies in the net sample to the reference data. However, this approach is connected with practical difficulties or restrictions resulting from the (non-) availability of appropriate population data (Elliot 1996, p. 2).

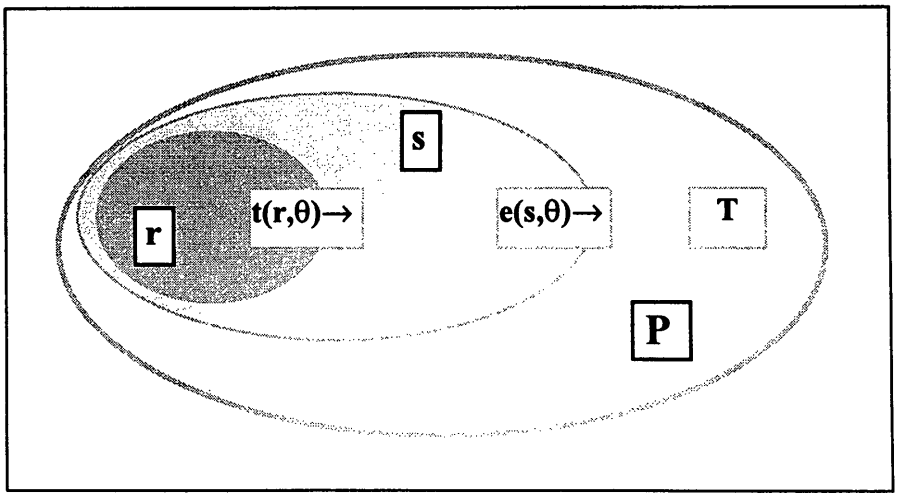
In our approach we estimate parameters of the population, for example the population total of a variable of interest, not directly from the net sample but by considering auxiliary information from the gross sample with the aim to adjust the net sample to the gross sample. The advantage is - in contrast to the usual post-stratification - that this information is available for each individual - that means each element of the net sample constitutes one cell and gets an "own" weight. For the construction of the weights the conditional minimax principle is applied. This way, some disadvantages of iterative procedures for the computation of weights can be avoided, for example the order of the variables in the adjusting process does not play a role. It is also possible to interpret our weights as BLU estimator in the framework of linear regression models.

The idea is demonstrated for data of the German General Social Survey (ALLBUS) 1996 whose response rate reached only 54%.

2 The conditional minimax principle

After having selected a sample s of a population $P=\{1,...,N\}$ the statistician estimates the population total $T= y_1+...+y_N$ of a variable of interest by an estimate $e(s,\theta)=\Sigma a_{si}y_i$ where a_{si} are real numbers with $a_{si}=0$ for $i\notin s$ and $\theta=(y_1,...,y_N)$. He will choose an estimate which

Figure 1: The estimation process



has good properties. Unfortunately, due to nonresponse the statistician does not get the y_i-values for all $i \in s$ but only for the units in a subset r of s. In many cases it cannot be assumed that r is a simple random sample of s. Thus the question remains how to estimate T using only the y-values in r. Often there exist x_i-values of auxiliary variables for all units in s, for example regional or demographic values. Our suggestion is to estimate T not directly but to estimate $e(s,\theta)$ by an estimate $t(r,\theta)=\Sigma b_{ri}y_i$ where b_{ri} are real numbers with $b_{ri}=0$ for $i \notin r$, and thus to estimate indirectly also T.

A decision theoretical approach for choosing an optimal estimator is given by the conditional minimax principle. This principle has been defined by Gabler (1988) and yields one possibility for computing $t(r,\theta)$. It can be also used for constructing cell weights, as Gabler (1991) shows. The conditional minimax principle says that in a given class D of estimators the optimal one in D minimizes the supremum of $L(t,\theta)$ where $L(t,\theta)$ denotes the loss of the estimate t at θ and the supremum is taken with respect to θ. We will consider only

$$L(t,\theta) = [t(r,\theta) - e(s,\theta)]^2$$

as loss function. Obviously, the loss function depends only on θ_s containing those y_i-values for which $i \in s$. Thus $L(t,\theta)=L(t,\theta_s)$, and we omit the suffix s in θ_s having in mind that θ now is an element of $\Re^n$ where n denotes the sample size of s. In general, the loss is unbounded on $\Re^n$. Therefore, we restrict ourselves to a subset Ω of $\Re^n$ known as parameter space. We consider only parameter spaces Ω on $\Re^n$ defined by quadratic forms in θ, i.e.

$$\Omega = \{\theta \in \Re^n: \theta' V \theta \leq c^2\}$$

where V is a nonnegative definite symmetric matrix and $c \neq 0$ is a given real number. Ω reflects the a-priori assumption about the y_i-values in s we have in mind. For example, Ω may be the set of all θ with bounded sample variance or more general

$$\Omega = \{\theta \in \Re^n: \sum_{i \in s} g_i (y_i - x_i \frac{ys}{xs})^2 \leq c^2 \}$$

where g_i and x_i are positive numbers and ys (xs) denotes the sum of all y-values (x-values) in s. An early paper dealing with such parameter spaces has been given by Wynn (1976). Minimax strategies are considered also by Bickel and Lehmann (1981), Chaudhuri and Stenger (1992), Stenger (1979) and Stenger and Gabler (1996).

If V is singular there exists a matrix X with VX=0. In this case the loss $L(t,\theta)$ is unbounded on Ω unless the linear estimator $t(r,\theta)$ satisfies $t(r,\xi)=e(s,\xi)$ for each column ξ

of X. Due to Hájek (1981) it means that t is a representative estimator for e with respect to X.

Thus our task is to look for a representative estimator $t(r,\theta)$ which minimizes the maximum loss on Ω.

The following Lemma can be derived from the Cauchy-Schwarz inequality and can be found in Gabler(1990, p. 111).

Lemma. Let W be a nonnegative definite symmetric n×n-matrix. For any n-dimensional vectors θ and α we have

$$(\theta' W\alpha)^2 \leq \theta' W\theta \cdot \alpha' W\alpha$$

Equality holds if and only if $A\theta$ and $ABW\alpha$ are proportional where $W=A'A$ and B is a symmetric reflexive g-inverse of W (see Rao and Mitra 1971).

Theorem. Let V be a nonnegative definite symmetric n×n-matrix of rank n-H and $VQ=0$, Q a n×H-matrix of rank H. We assume that R is a n×H matrix of rank H with $R'Q=I$, the identity matrix. Let U be the symmetric reflexive g-inverse of V with $UR=0$. The conditional minimax solution $t^c(r,\theta)=\Sigma b_{ri}y_i$ for $e(s,\theta)=\Sigma a_{si}y_i$ on $\Omega=\{\theta\in\Re^n: \theta' V\theta\leq c^2\}$ is given by

$$b^c = U_{rr}^{-1}U_{rs}a + U_{rr}^{-1}Q_r(Q_r' U_{rr}^{-1}Q_r)^{-1}(Q'-Q_r' U_{rr}^{-1}U_{rs})a$$

Proof. For the n-dimensional vector $\alpha = (b_{si} - a_{si})_{i\in s}$ we must have $\alpha'Q=0$ as shown above to get finite maximal loss. We define $\zeta=(I - QR')\theta$. Then $R'\zeta=0$ and

$$L(t,\theta) = (\alpha'\theta)^2 = (\alpha'\zeta)^2$$

For any such ζ there exists η with $\zeta=U\eta$. We get

$$(\alpha'\theta)^2 = (\alpha'\zeta)^2=(\alpha' U\eta)^2$$

From the lemma it follows

$$(\alpha'\theta)^2 = (\alpha' U\eta)^2 \leq (\alpha' U\alpha)(\eta' U\eta) = (\alpha' U\alpha)(\eta' UVU\eta) = (\alpha' U\alpha)(\zeta' V\zeta) = (\alpha' U\alpha)(\theta' V\theta)$$

In order to obtain a conditional minimax solution we have to minimize

$$\alpha' U\alpha \text{ with the constraint } Q'\alpha=0.$$

Defining the n-dimensional vector $a = (a_{si})_{i\in s}$ and the m-dimensional vector $b=(b_{ri})_{i\in r}$ where m is the size of the sample r, we have $\alpha'U\alpha = a'Ua - 2a'U_{sr}b + b'U_{rr}b$ and the minimum of $\alpha'U\alpha$ under the constraint $a'Q=b'Q_r$ is attained at

$$b^c = U_{rr}^{-1}U_{rs}a + U_{rr}^{-1}Q_r(Q_r'U_{rr}^{-1}Q_r)^{-1}(Q'-Q_r'U_{rr}^{-1}U_{rs})a$$

provided that all inverses exist.
In the case $U_{rs}a=0$, for example if $a=Rx$, the above expression simplifies to

$$b^c = U_{rr}^{-1}Q_r(Q_r'U_{rr}^{-1}Q_r)^{-1}Q'a$$

Examples for b^c such as the ratio estimator can be found in Gabler (1990, pp. 114-116).

Remark 1. Let $Y=(Y_i)_{i\in s}$ be a superpopulation model with expectation $Q\beta$, Q a $n\times H$ matrix of rank H and β a n-dimensional parameter vector, and variance-covariance matrix U, where U is a nonnegative definite symmetric matrix of rank n-H and R a $n\times H$ matrix with $R'Q=I$ and $UR=0$. Analogous to chapter 5.5 in Gabler (1990) it can be shown that

b^c is a BLU estimator with respect to the above model.

From $E[R'Y]=R'Q\beta=\beta$ and $var(R'Y)=R'UR=0$ we conclude $R'Y=\beta$ with probability one. Estimation of β is the same as estimation of $R'Y$.

Remark 2. Let $Y=(Y_i)_{i\in s}$ be singular multivariate normally distributed with expectation $Q\beta$, Q a $n\times H$ matrix of rank H, β a n-dimensional parameter vector, and variance-covariance matrix U, where U is a nonnegative definite symmetric matrix of rank n-H and R a $n\times H$ matrix with $R'Q=I$ and $UR=0$. Analogous to chapter 5.5 in Gabler (1990), after a suitable truncation of the distribution of Y to $\Omega = \{\theta\in\Re^n: \theta'V\theta\leq c^2\}$ and choosing

$$f(\beta) = 1 \text{ for all } \beta\in\Re^H$$

as prior distribution of β, it can be shown that

b^c is a Bayes estimator with respect to the above model.

If our viewpoint is the minimax principle we will start the computations with V and Q. If our viewpoint is the superpopulation approach we will start the computations with U and R.

3 Special cases

For real computations there is a drawback in the formula of b^c. If the sizes of the samples r and s are large, say 2000 and 4000, the matrices U_{rr} and U_{rs} are huge and b^c cannot be computed on a PC because of memory limitations. Fortunately, there exist special cases for which this problem can be eliminated.

If

$$V=(I-RQ')G(I-QR')$$

where G is a positive definite matrix, then

$$U = G^{-1} - G^{-1}R(R'G^{-1}R)^{-1}R'G^{-1}.$$

U depends on Q only via R'Q=I.

We assume in addition that G_{rt} is a matrix of zeros. Then the inverse of U_{rr} is

$$U_{rr}^{-1} = G_{rr} + R_r(R_t'G_{tt}^{-1}R_t)^{-1}R_r'$$

where R_r consists of the rows of R belonging to r.

Proof. From $U=G^{-1}-G^{-1}R(R'G^{-1}R)^{-1}R'G^{-1}$ it follows that

$$U_{rr} = G_{rr}^{-1} - G_{rr}^{-1}R_r(R'G^{-1}R)^{-1}R_r'G_{rr}^{-1}$$

since G_{rt} is a matrix of zeros (which implies $(G_{rr})^{-1} = (G^{-1})_{rr} =: G_{rr}^{-1}$).

The above formula for U_{rr}^{-1} results from the fact that

$$(A - BB')^{-1} = A^{-1} + A^{-1}B(I - B'A^{-1}B)^{-1}B'A^{-1}.$$

Defining t as complementary set of r in s and

$$C = R'G^{-1}R - R_r'G_{rr}^{-1}R_r = R_t'G_{tt}^{-1}R_t$$

we have

$$U_{rr}^{-1}U_{rs}a = a_r - R_rC^{-1}R_t'G_{tt}^{-1}a_t$$
$$U_{rr}^{-1}Q_r = G_{rr}Q_r + R_rC^{-1}R_r'Q_r$$
$$(Q'-Q_r'U_{rr}^{-1}U_{rs})a = Q_t'a_t + Q_r'R_rC^{-1}R_t'G_{tt}^{-1}a_t.$$

and it follows

$$b^c = a_r - R_r C^{-1} R_t' G_{tt}^{-1} a_t +$$

$$\left(G_{rr}Q_r + R_r C^{-1} R_r' Q_r\right)\left(Q_r' G_{rr} Q_r + Q_r' R_r C^{-1} R_r' Q_r\right)^{-1}\left(Q_t' a_t + Q_r' R_r C^{-1} R_t' G_{tt}^{-1} a_t\right)$$

If G is a diagonal matrix this formula helps to save memory for software packages, such as GAUSS, with matrix language and elementwise multiplication of matrices. For example, $G_{rr}Q_r$ may be programmed as $g_r .*Q_r$ where g_r is the vector of the diagonal elements of G belonging to r, and $.*$ is the elementwise multiplication operator. Thus all terms involved in b^c may be computed by linkage of relative small components.

The question is how restrictive the representation of $V=(I-RQ')G(I-QR')$ with G as diagonal matrix is. Our feeling is that this restriction is not a real one. We will demonstrate it for the case H=1 and $a=R\eta$.

Let w be a positive vector with $w'Q_r = a'Q = \eta$. Then there exists always a positive diagonal matrix G yielding w as conditional minimax solution.

Proof. Let G_{tt} be an arbitrary diagonal matrix and let g_r denote the vector consisting of the diagonal elements of G_{rr}. We define

$$g_r = [k \cdot w - R_r (R_t' G_{tt}^{-1} R_t)^{-1} R_r' Q_r]./Q_r$$

where k can be always chosen so large that g_r is a positive vector. The operator ./ means elementwise division. It follows

$$\begin{aligned} b^c &= U_{rr}^{-1} Q_r \left(Q_r' U_{rr}^{-1} Q_r\right)^{-1} \eta \\ &= \left(g_r .*Q_r + R_r (R_t' G_{tt}^{-1} R_t)^{-1} R_r' Q_r\right)\left(Q_r' [g_r .*Q_r + R_r (R_t' G_{tt}^{-1} R_t)^{-1} R_r' Q_r]\right)^{-1} \eta \\ &= \frac{k \cdot w}{Q_r'(k \cdot w)} \eta \\ &= w \end{aligned}$$

Thus the restriction to parameter spaces defined by

$$\Omega = \{\theta \in \Re^n : \theta' V \theta \le c^2\}$$

where

$$V=(I-RQ')G(I-QR')$$

with diagonal matrix G is not as restrictive as it seemed at first.

4 Application to the ALLBUS 1996

Now we will apply the above results to the data of the ALLBUS 1996. The ALLBUS is a bi-annually General Social Survey similar to the American General Social Survey and the British Social Attitudes Survey (see Davis et al. 1994). The ALLBUS is conducted as face-to-face interviews. Until 1992 the underlying sampling design of the ALLBUS theoretically selected German households with equal probabilities and one respondent within a chosen household proportional to the inverse of the number of the target people in that household. 1994 and 1996 the design was changed as follows: After stratification of the communities due to regional aspects, and selecting communities as primary units proportional to the number of the inhabitants belonging to the target population, the respondents were selected from local registers. This procedure implies equal inclusion probabilities for all persons of the target population. If all selected people in the gross sample were attainable and would answer the questions in the face-to-face interviews, the researcher could estimate the population total of a variable of interest by the sample total or by the ratio estimator each multiplied by the sampling fraction. Unfortunately, only a subset of the chosen people were willing or able to answer the questions. About 54%, i.e. 3518 persons were recorded in the official ALLBUS file distributed by the Zentralarchiv in Cologne.

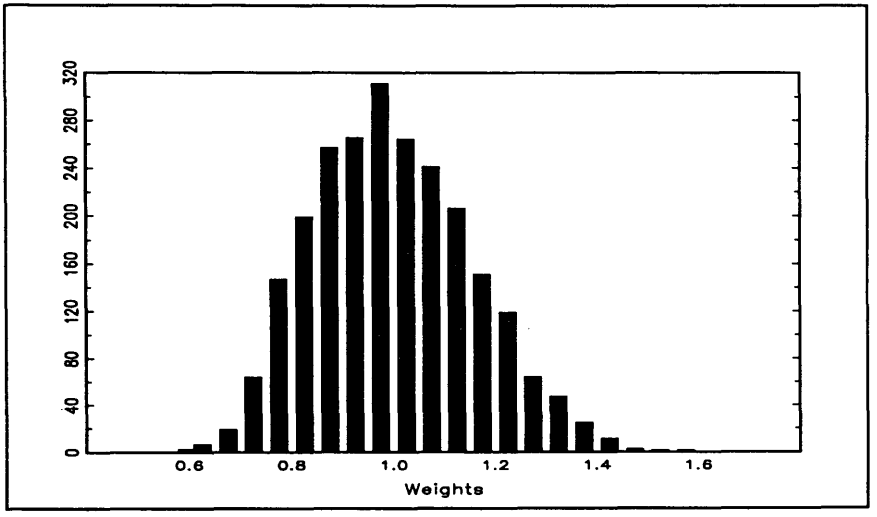
One of the advantages of this ALLBUS design is the knowledge of additional information for (nearly) all people of the gross sample. More specifically we obtained after some corrections 6488 selected persons from the local registers, 4430 in West Germany and 2058 in East Germany. Since West and East Germany are treated in the design separately we should apply our formulae also for both parts of Germany separately. We will present the results exemplary for West Germany with 2402 respondents.

As auxiliary variables we use age[1], gender, nationality, regional characteristics and the BIK variable which classifies the communities. Other variables - if available - could of course be used. For the gross sample the values of these variables are distributed similar to that in the population. For example, the portion of males in the gross sample is 48%, in contrast to 50% in the ALLBUS net sample. The columns of the matrix Q consist of the values of the variables which were categorized except age. Q has 4430 rows and 23 columns, 1 for age, 1 for gender, 1 for nationality, 10 for the federal states and 10 for the BIK variable. As R we used $Q(Q'Q)^{-1}$ and for a the vector consisting of ones. After computing b^c in the conditional minimax estimator we normed the weights in order to have sum equals 2402 which is the size of the ALLBUS sample in West Germany. The

[1] In Berlin the age was available only in classified form. We used the class mean as age. Sometimes the age value of the respondents differs from the age values of the local register. The reason could be that the interviewer asked the wrong person in the household.

following Figure 2 shows the distribution of the weights.

Figure 2: Frequencies of the optimal weights



The minimum, maximum, mean and standard deviation of the weights are 0.5889, 1.5543, 1.0000, 0.1562.

5 Results

There exist only minor differences in the distributions of the auxiliary variables age, gender, nationality, regional characteristics and the BIK-variable between the gross sample s and the net sample r. This is an indication for the relative high quality (unbiasedness) of the net sample with regard to these variables. Therefore, the adjustment results in only little changes - and mostly in the expected direction.

For several survey variables we compared the portions of the unweighted and the weighted sample. This way we show the effect of the weighting procedure. In general, the differences between the weighted and the unweighted variables are small, in most cases less than one percent (see Table 1) and therefore negligible.

Table 1: Unweighted and weighted distributions for selected ALLBUS variables

ALLBUS Variable	Unweighted	Weighted	Difference col.2-col.3
V10	Women should care for children		
Miss.	3.16	3.18	-0.02
Agree strongly	23.19	24.35	-1.16
Agree	26.23	26.55	-0.32
Disagree	27.52	26.69	0.83
Disagree strongly	19.90	19.23	0.67
V141	Sex		
Male	50.12	48.31	1.81
Female	49.88	51.69	-1.81
V142	Education		
Miss.	0.21	0.23	-0.02
No formal education certificate	2.33	2.33	0.00
Lower secondary school certificate	47.84	48.97	-1.14
Intermediate secondary school certificate	24.77	24.34	0.43
Specialized Abitur	5.75	5.63	0.12
Abitur	17.86	17.39	0.47
Other certificate	0.42	0.43	-0.01
Pupil	0.83	0.69	0.14
V155	Occupational status		
Full-time employed	48.79	46.02	2.78
Half-time employed	6.70	6.74	-0.04
Part-time employed (not main job)	4.75	4.78	-0.04
Not employed	39.76	42.46	-2.70
V183	Marital status		
Married, live together	63.16	63.63	-0.48
Married, live apart	1.71	1.68	0.02
Widowed	7.66	9.15	-1.49
Divorced	4.95	5.18	-0.23
Single	22.52	20.35	2.18
V263	Number of persons living in the household		
1 Person	16.78	17.94	-1.16
2 Persons	35.05	36.79	-1.74
3 Persons	21.23	20.33	0.91
4 Persons	17.99	16.74	1.25
5 Persons	6.00	5.58	0.42
6 Persons	1.87	1.67	0.20

Table 1 (continued):

Variable	Unweighted	Weighted	Difference col.2-col.3
7 Persons	0.79	0.70	0.09
8 Persons	0.17	0.16	0.01
9 Persons	0.08	0.07	0.02
18 Persons	0.04	0.04	0.00
V354	Number of inhabitants in community		
To 1.999 inh.	4.91	4.51	0.40
2.000 - 4.999 inh.	9.20	8.30	0.91
5.000 - 19.999 inh.	24.65	23.62	1.03
20.000 - 49.999 inh.	16.61	16.60	0.01
50.000 - 99.999 inh.	8.24	8.65	-0.41
100.000 - 499.999 inh.	21.11	21.87	-0.76
500.000 inh. and more	15.28	16.45	-1.17
V355	BIK - Type of community		
Type 1 (small urban areas)	4.91	4.51	0.40
Type 2	8.08	7.20	0.88
Type 3	16.57	15.49	1.08
Type 4	9.16	9.14	0.02
Type 5	1.04	0.93	0.12
Type 6	3.29	3.57	-0.28
Type 7	5.70	6.32	-0.62
Type 8	9.49	9.12	0.37
Type 9	12.53	11.76	0.77
Type 10 (centers of large cities)	29.23	31.96	-2.74
	Number of persons 18 years and older living in the household		
Miss.	0.12	0.10	-0.02
1 Person	66.74	69.12	-2.38
2 Persons	16.49	15.41	1.08
3 Persons	11.74	10.82	0.92
4 Persons	3.62	3.32	0.30
5 Persons	1.17	1.10	0.07
6 Persons	0.08	0.08	0
7 Persons	0.04	0.04	0

In the following we discuss only those differences between the unweighted and weighted portions of our variables of interest where the differences are greater than one percent. First we present for some demographic variables how the weights change their distributions.

The portion of

- men decreases
- people with lower secondary school certificate increases
- full-time employed persons decreases
- not employed people increases
- widowed people increases
- singles decreases
- people in households with one person increases
- people in households with two persons increases
- people in households with four persons decreases
- people in communities with 5.000-19.999 inhabitants decreases
- people in communities with 500.000 and more inhabitants increases
- people in rural areas (BIK Type 3) decreases
- people in the centers of urban areas (BIK Type 10) increases
- people in households with only one person older than 17 years increases
- people in households with two persons older than 17 years decreases

in the weighted sample compared to the unweighted net sample.

Most of these differences can be explained by the adjustment variables gender and BIK. As we have already mentioned in the net sample the portion of men is somewhat higher than in the gross sample. The adjustment of gender affects such variables as marital status, occupational status and household size because females in Germany are more frequently widowed, not employed, and living alone in a household than men. The fact that the portion of people living in households with two persons increased and people in households with two persons older than 17 years decreased simultaneously can be explained as follows: It seems that single adults living with one child (mostly women) were underrepresented in the net sample. It may be that those persons form a typical group of nonrespondents. A not desirable result is the decreasing of the portion of singles in the weighted sample because this portion is in the population higher than in the unweighted sample[2]. The reason for this is the substantially smaller portion of female singles in the net sample compared to male singles.

The adjustment of the variable BIK has the effect that the portion of people in large cities, especially in their centers, increased. People living in these areas are another group with a relatively high portion of nonrespondents.

The differences in other portions of variables like denomination, union membership and

[2] In 1995 the portion of single persons in West Germany was 23.5% for people 18 years and older (Statistisches Bundesamt, p.52). However, it has to be considered that this information of the Federal Statistical Office does not include foreigners.

German nationality between unweighted and weighted data are negligible. Therefore, we did not include the numbers in Table 1.

We also checked the differences between unweighted and weighted data for several survey variables which measure political and other attitudes. The portions differed only in one case by more than one percent: In the weighted data set the portion of people who strongly agree with the statement "It is much better for all persons concerned if the man preferred his job and the woman stayed at home and cared for children" is somewhat higher. That may be an effect of the adjustment of the variable age: In the gross sample the mean age of all respondents was 47.1 in contrast to 45.4 in the unweighted net sample. That means that after weighting the respondents are "older" and therefore more conservative. In all other variables (Left-Right-Scheme, Inglehart-Index, political interest, subjective social class) the differences are not worth mentioning.

6 Conclusions

The conditional minimax principle enables the researcher to treat nonresponse in a flexible way by using prior knowledge from the gross sample. He has several possibilities to incorporate the knowledge into his model. First of all he should summarize his knowledge into a matrix Q. The columns of Q may contain the values of a metric variable or a column of a design matrix of a categorical variable. If these values are not known for all respondents in the gross sample they should be estimated by a suitable procedure such as imputation to avoid missing values. In addition, Q must be of full rank. An important factor is the determination of the parameter space determined by a singular matrix V which describes the admissible variation of the values of a variable of interest. A class of such matrices V was given which involves a diagonal matrix G and a matrix R with R'Q as identity matrix. In order to take into consideration the probability of participation or the probability of contact for each respondent, the researcher could incorporate different diagonal elements into G. $R=AQ(Q'AQ)^{-1}$ is only one possibility for choosing R where A is an arbitrary non-singular matrix. Since the computation of b^c does not lead to nonnegative weights necessarily, the selection of A may help to avoid negative or other extreme weights. Further investigations are necessary to clarify this point.

Another important advantage of the presented idea is that it can be easily transferred to the estimation of parameters in subgroups.

The application to data of the ALLBUS 1996 yields different results: The weights assigned to most of the elements of the net sample are in the small range from 0.8 to 1.2 . The reason for this are the relative small differences in the distributions of the auxiliary variables between the net sample and the gross sample of the ALLBUS 1996. In cases

where the difference between the net sample and the gross sample is not so small the effects of weighting are stronger. In our application we stated only minor changes in the survey variables due to weighting - and in the expected directions.

On condition that the gross sample is drawn as a random sample that represents the population well, and auxilary information on the individuals of the gross sample is available, our suggestion for the estimation of population parameters seems to be promising.

References

Bickel, P.J. and Lehmann, E.L. (1981). A Minimax Property of the Sample Mean in Finite Populations. Annals of Statistics 9, pp. 1119-1122

Chaudhuri, A. and Stenger, H. (1992). Survey Sampling. New York: Marcel Dekker, Inc.

Davis, J.A., Mohler, P.Ph. and Smith, T. (1994). Nationwide General Social Surveys. In: Borg, I. and Mohler, P.Ph. (eds.): Trends and Perspectives in Empirical Social Research. Berlin: de Gruyter

Deville, J.-C. and Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association 87, pp. 376-382

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling. Journal of the American Statistical Association 88, pp. 1013-1020

Elliot, D. (1991). Weighting for Nonresponse - a Survey Researchers Guide. OPCS

Elliot, D. (1996). The Effectiveness of CALMAR in Reducing Nonresponse Bias on the UK Family Expenditure Survey. Paper presented at the 7th International Workshop on Household Nonresponse, Rome, Oktober 1996

Gabler, S. (1988). A Conditional Minimax Approach in Survey Sampling. Metrika 35, pp. 211-221

Gabler, S. (1990). Minimax Solutions in Sampling from Finite Populations. Lecture Notes in Statistics 64, New York: Springer Verlag

Gabler, S. (1991). Eine allgemeine Formel zur Anpassung an Randtabellen. ZUMA-Nachrichten 29, pp. 29-43

Gabler, S. and Häder, S. (1997). Wirkung von Gewichtungen bei Face-to-Face und Telefonstichproben. In: Gabler, S. and Hoffmeyer-Zlotnik, J.H.P. (eds.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag, pp. 221-245

Häder, S. and Gabler, S. (1997). Deviations from the population and optimal weights. In: Saris, W.E. and Kaase, M. (eds.): Eurobarometer – Measurement Instruments for Opinions in Europe. ZUMA-Spezial Band 2, pp. 37-52

Hájek, J. (1981). Sampling from a Finite Population. New York: Marcel Dekker, Inc.

Little, R.J.A. (1989). Survey Inference with Weights for Differential Sample Selection on

Nonresponse. ASA 1989 Proceedings of the Section on Survey Research Methods, pp. 62-69

Rao, C.R. and Mitra, S.K. (1971). Generalized Inverse of Matrices and its Applications. New York: John Wiley and Sons, Inc.

Schnell, R. (1996). Development of Nonresponse in Germany 1953-1994. Paper presented at the 7th International Workshop on Household Nonresponse, Rome, Oktober 1996

Statistisches Bundesamt (1995): Demographische Standards. Methoden-Verfahren-Entwicklungen

Stenger, H. (1979). A Minimax Approach to Randomization and Estimation in Survey Sampling. Annals of Statistics 7, pp. 395-399

Stenger, H. and Gabler, S. (1996). A Minimax Property of Lahiri-Midzuno-Sen's Sampling Scheme. Metrika 43, pp. 213-220

Wynn, H.P. (1976). Connected Finite Population Sampling Plans. Biometrika 63, pp. 208-210

CONTRIBUTORS IN ALPHABETICAL ORDER:

ALLEN, BRUCE
Westat Inc.
1650 Research Building
Rockville
Maryland 20850
USA

BENNETT, NIKKI
Social Survey Division
Office for National Statistics
1, Drummond Gate
London, SW1V, 2QQ
England

BRICK, J. MICHAEL
Westat Inc.
1650 Research Building
Rockville
Maryland 20850
USA

BÜCHEL, FELIX
Technical University Berlin
Uhlandstr. 4/5
10623 Berlin
Germany

CAMPANELLI, PAMELA
Survey Methods at SCPR
35 Northampton Square,
London, EC1V OAX
United Kingdom

CANTOR, DAVID
Westat Inc.
1650 Research Building
Rockville
Maryland 20850
USA

CARTON, ANN
Departement of Sociology,
K.U.Leuven
Van Evenstraat 2B,
3000 Leuven,
Belgium

CECCARELLI, CLAUDIO
National Institute of Statistics
Via Adolfo Ravà, 150 -
00142 Rome
Italy

CHRISTOPH, BERNHARD
Linienstr. 131
10115 Berlin
Germany

COCCIA, GIULIANA
National Institute of Statistics
Via Adolfo Ravà, 150 -
00142 Rome
Italy

CRESCENZI, FABIO
National Institute of Statistics
Via Adolfo Ravà, 150 -
00142 Rome
Italy

CUNNINGHAM, PATRICIA
Westat Inc.
1650 Research Building
Rockville
Maryland 20850
USA

DE HEER, WIM
Statistics Netherlands
P.O. Box 4481
6401 CZ Heerlen
Netherlands

DE LEEUW, EDITH
Plantage Doklaan 40
1018 CN Amsterdam
Netherlands

DOLSON, DAVID
Statistics Canada
Ottawa
Canada K1A 0T6
Canada

FLEISCHER, KARLHEINZ
University Erlangen-Nuernberg
Statistics and Econometrics
Lange Gasse 20
90403 Nuernberg
Germany

FREETH, STEPHANIE
Social Survey Division
Office for National Statistics
1, Drummond Gate
London, SW1V, 2QQ
England

GIAMBO, PAMELA
Westat Inc.
1650 Research Building
Rockville
Maryland 20850
USA

HARKNESS, JANET
Zentrum für Umfragen, Methoden
und Analysen (ZUMA)
P.O. Box 12 21 55
68072 Mannheim
Germany

HARRIS-KOJETIN, BRIAN A.
U. S. Bureau of Labor Statistics
2 Massachusetts Avenue N.E.,
Room 4915
Washington, D. C. 20212
USA

HAVASI, ÉVA
Hungarian Central Statistical Office
Keleti Károly u. 5-7.
P.O. Box 51
1525 Budapest II.
Hungary

HOX, JOOP J.
University Utrecht
Methodology & Statistics
Department
P.O. Box 80.140
3508 TC Utrecht
Netherlands

KANTOROWITZ, MALKA
Israel Central Bureau of Statistics
P.O.Box 13015, Hakirya Romema,
Jerusalem 91130,
Israel

KENNY, GENEVIEVE
Urban Institute
2100 M. Street NW
Washington D.C. 20036
USA

KING, JOHN
Office for National Statistics
1 Drummond Gate
London, SW1V 2QQ
United Kingdom

KUUSELA, VESA
Statistics Finland
Interview services
PB 5A
FIN-00022 Statistics Finland
Finland

LAAKSONEN, SEPPO
Statistics Finland
Stakes and University of Tampere
00022 Statistics Finland
Finland

LINDSTRÖM, HÅKAN L.
R&D Department
Statistics Sweden,
Statistics Sweden, U/PTU
701 89 Örebro
Sweden

LOOSVELDT; GEERT
Departement of Sociology,
K.U.Leuven
Van Evenstraat 2B,
3000 Leuven,
Belgium

LOZAR, KATJA
Faculty of Social Sciences
University of Ljubljana
Kardeljeva ploscad 5
1000 Ljubljana
Slovenia

MAHER, MARY P.
Survey Research Center
Institute for Social Research
University of Michigan
P.O. Box 1248,
Ann Arbor, MI, 48106
USA

MARTIN, JEAN
Social Survey Division
Office for National Statistics
1, Drummond Gate
London, SW1V, 2QQ
England

MARTON, ÁDÁM
Hungarian Central Statistical Office
Keleti Károly u. 5-7.
P.O. Box 51
1525 Budapest II.
Hungary

MOHLER, PETER PH.
Zentrum für Umfragen, Methoden
und Analysen (ZUMA)
P.O. Box 12 21 55
68072 Mannheim
Germany

PETRONI, RITA J.
U.S. Bureau of the Census,
Washington, D.C. 20233-0001
USA

PICKERY, JAN
Departement of Sociology,
K.U.Leuven
Van Evenstraat 2B,
3000 Leuven,
Belgium

RAESSLER, SUSANNE
University Erlangen-Nuernberg
Statistics and Econometrics
Lange Gasse 20
90403 Nuernberg
Germany

RENDTEL, ULRICH
University of Frankfurt/Main
Mertonstraße 17
60054 Frankfurt/M.
Germany

SCHNEID, MICHAEL
Zentrum für Umfragen, Methoden
und Analysen (ZUMA)
P.O. Box 12 21 55
68072 Mannheim
Germany

SINGER, ELEANOR
Survey Research Center
Institute for Social Research
University of Michigan
P.O. Box 1248,
Ann Arbor, MI, 48106
USA

SINGH, RAJENDRA P.
U.S. Bureau of the Census,
Washington, D.C. 20233-0001
USA

SLOBASKY, RENEE
Westat Inc.
1650 Research Building
Rockville
Maryland 20850
USA

SNIJKERS, GER
Statistics Netherlands
P.O. Box 4481
6401 CZ Heerlen
Netherlands

STURGIS, PATRICK
Survey Methods at SCPR
35 Northampton Square,
London, EC1V OAX
United Kingdom

SWAIN, LARRY
Statistics Canada
Ottawa
Canada K1A 0T6
Canada

TUCKER, CLYDE
U. S. Bureau of Labor Statistics
2 Massachusetts Avenue N.E.,
Room 4915
Washington, D. C. 20212
USA
United Kingdom

VAN HOEWYK, JOHN
Survey Research Center
Institute for Social Research
University of Michigan
P.O. Box 1248,
Ann Arbor, MI, 48106
USA

VEHOVAR, VASJA
Faculty of Social Sciences
University of Ljubljana
Kardeljeva ploscad 5
1000 Ljubljana
Slovenia

WHITE, AMANDA
Social Survey Division
Office for National Statistics
1, Drummond Gate
London, SW1V, 2QQ
England

ZALETEL, METKA
Statistical Office of the Republic of Slovenia
Vozarski pot 12
1000 Ljubljana
Slovenia

WAITE, PRESTON JAY
U.S. Bureau of the Census
Department of Commerce
Washington, D.C. 20233
USA

HUGGINS, VICKI J.
U.S. Bureau of the Census
Department of Commerce
Washington, D.C. 20233
USA

MACK, STEPHEN P.
U.S. Bureau of the Census
Department of Commerce
Washington, D.C. 20233
USA

GABLER, SIEGFRIED
Zentrum für Umfragen,
Methoden und Analysen
(ZUMA)
P.O. Box 12 21 55
68072 Mannheim
Germany

HÄDER, SABINE
Zentrum für Umfragen,
Methoden und Analysen
(ZUMA)
P.O. Box 12 21 55
68072 Mannheim
Germany

GESIS

ZUMA ist Mitglied der
Gesellschaft Sozialwissenschaftlicher
Infrastruktureinrichtungen (GESIS) e.V.

ISBN 3-924220-15-8