![ISER logo](INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH)

A REVIEW OF METHODOLOGICAL RESEARCH PERTINENT TO LONGITUDINAL SURVEY DESIGN AND DATA COLLECTION

Peter Lynn, Nicholas Buck, Jonathan Burton, Annette Jäckle, Heather Laurie

Institute for Social and Economic Research

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Studies Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It is responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national data sets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

Readers wishing to cite this document are asked to use the following form of words:

> **Lynn, Peter, Buck, Nicholas, Burton, Jonathan, Jäckle, Annette, and Laurie, Heather (December 2005) 'A Review of Methodological Research Pertinent to Longitudinal Survey Design and Data Collection',** *Working Papers of the Institute for Social and Economic Research*, **paper 2005-29. Colchester: University of Essex.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

We present here a review of methodological research regarding issues that are pertinent to surveys involving longitudinal data collection, i.e. repeated measurement over time on the same units. The objective of the review is to identify important gaps in our knowledge of issues affecting the design and implementation of such surveys and the use of the data that they provide. This should help to inform the development of an agenda for future methodological research as well as serving as a useful summary of current knowledge. The issues addressed relate to sample design, missing data (as a result of item and unit non-response and attrition) and measurement error (including panel conditioning).

Key words: attrition, bias, dependent interviewing, item non-response, longitudinal surveys, measurement error, panel conditioning, recall errors, sample design, seam bias, unit non-response

Contents

1. Background

The UK Longitudinal Studies Centre (ULSC) is concerned to develop and implement a broad programme of methodological research to support its wider objectives. Those wider objectives include helping to improve the quality and quantity of use made of UK longitudinal data resources, and helping to inform decision making regarding the design and implementation of longitudinal surveys. To inform the development of the methodological research programme, a review of previous research has been carried out and is summarised here.

Three broad areas of methodological research relevant to longitudinal surveys can be identified:

I.   Analytical methods that allow researchers to exploit the particular, and often complex, structures of longitudinal data;

II.   Research which contributes to an understanding of how data quality issues may influence the research that can be carried out using longitudinal survey data;

III.   Research which contributes to the improvement of longitudinal data collection methods and longitudinal survey design.

The review has initially been restricted to areas II and III. This is not to say that area I is in any way less important, merely that it was necessary for practical reasons to constrain the current review. Research into analytical methods for longitudinal survey data will be dealt with separately.

Within the two broad areas that form the subject of this review, we have identified eight research topics. These are not comprehensive and neither are they all entirely distinct from one another. However, we believe that they cover the main methodological issues that are either unique to longitudinal surveys or have

particular importance and/or distinctive characteristics in the case of longitudinal surveys. The topics are:

1. The nature of attrition, and bias due to attrition and non-response;

2. Methods to minimise attrition;

3. Methods to minimise item non-response

4. The nature of recall error;

5. Measurement error that is specific to repeated measures;

6. Panel conditioning;

7. Interviewing strategies;

8. Sampling for longitudinal surveys.

There are two further topic areas that could arguably be assigned to either area I or III as they concern the preparation of data for analysis. These are:

9. Weighting methods;

10. Imputation methods.

We do not deal with these two topics in this current paper, but we intend to review them separately.

For each of topics 1 to 8 identified above, we have attempted to summarise what is known (research findings) and what is not known. In conclusion, we identify promising areas for further research. Our focus has been on research knowledge that is relevant to a range of large-scale longitudinal social and economic surveys. In other words, we have concentrated on issues that are of generic interest to

longitudinal surveys and have excluded survey-specific issues. We think this should be seen as a strength rather than a limitation of our study, as in practice we found very few issues that were genuinely survey-specific.

However, longitudinal surveys come in many forms, so to avoid misunderstanding we should describe the type of survey that we have in mind. Our prime concern is with surveys of the nature of those that currently constitute the portfolio of important longitudinal survey data resources for the social and economic sciences in the UK, or which could potentially contribute to such resources. These include, but are not limited to:

- National Child Development Study (NCDS);

- 1970 British Cohort Survey (BCS70);

- Millenium Cohort Study (MCS);

- British Household Panel Survey (BHPS);

- English Longitudinal Study of Ageing (ELSA);

- Youth Cohort Study (YCS);

- Longitudinal Survey of Young People in England (LSYPE);

- Family and Children Survey (FACS);

- Integrated Household Survey (IHS, previously referred to as the Continuous Population Survey, CPS);

- British Election Panel Surveys (BEPS).

Several of these surveys currently receive substantial ESRC funding.

These surveys are characterised by large sample sizes of human populations and multiple waves of data collection over long periods of time (up to almost 50 years so far, in the case of NCDS). They all involve collecting data directly from the subjects of enquiry, in most cases by personal interview, but sometimes by postal self-completion questionnaire or by telephone interview. They vary in other respects, such as sample design, follow-up strategy, data collection modes and subject matter. We have inevitably had to apply judgement regarding the relevance of research knowledge that may be based on work carried out many years ago, in other countries, on rather specific study populations, etc. We have attempted to comment on the desirability or otherwise of replicating/ up-dating studies in the current UK context. Having said that, though this review concentrates on the UK context we believe that it contains much that will be of relevance to researchers in other countries.

For each topic, we have sought to identify relevant previous reviews (reviews of the literature, meta-analyses, systematic reviews, etc) and/or any seminal studies and to summarise the key conclusions. In some topic areas, there are many aspects that are not specific to longitudinal surveys (but are still important to longitudinal surveys) but some that *are* specific to longitudinal surveys. We have attempted to review both, but with an emphasis on the latter. Our suggestions for future research are limited to the latter category.

Section 2 of this paper summarises our findings in each subject area. The reviews upon which these summaries are based appear as sections 3 to 10.

2. Preliminary Findings

2.1 THE NATURE OF ATTRITION AND BIAS DUE TO ATTRITION AND NON-RESPONSE

A considerable amount of research has investigated the characteristics of non-respondents in general. A smaller body of research has addressed the characteristics of those who drop out of longitudinal surveys. More recently, there have been developments of the theory of how and why sample members respond or fail to respond, and this has been linked to empirical studies. In consequence, a lot is known about the nature of attrition (how and why it happens) and the implications for subsequent bias (what kinds of people are more or less prone to it), though there remain some gaps in our knowledge:

- Apart from the recently-developed theory of utilitarian individualism, little is known about the *reasons* for attrition due to refusal amongst previously co-operative respondents. It would be very useful to identify what distinguishes between those who choose to continue in a survey and those who do not in terms of factors that might be influenced by survey design and implementation.

- Little is known about the nature of non-ignorable attrition, particularly that related to geographical mobility. This is likely to be culture-specific to some extent, and UK research on this topic is lacking.

Direct assessment of bias due to attrition and non-response is only possible in special situations, such as where a very informative sampling frame is available, linkage to individual records is possible, or some other form of validation study can be carried out. Such assessment is rarely possible and is expensive. In consequence, research in this area is thin on the ground. The issue is, however, a major concern for researchers. As bias is specific to the measures of importance in particular surveys, it is difficult to generalise about the kind of information that would be of most value. It would be desirable to be alert to the possibility of obtaining auxiliary and validation data on all longitudinal surveys.

5

2.2 METHODS TO MINIMISE ATTRITION

Again, there is a lot of research knowledge about the effects of specific design and implementation features on attrition rates with particular survey populations, though important knowledge gaps remain. These relate to the interactions between the various design features; to the generalisability of effects across populations and subgroups; and to effects on non-response bias, rather than just response rates. Any or all of these features could usefully be built in to any future research in the areas outlined below. Some potential areas for future research include:

- The potential gains from better harnessing a variety of methods for keeping in touch with sample members, including mobile phone numbers, SMS, email, web-based methods, etc;

- The potential of tailoring the approach to sample members based on information provided at earlier waves;

- Alternatives ways of introducing the nature of panel membership, and the timing of such introductions;

- Alternative levels, forms and administration methods for respondent incentives;

- Effects of 'matching' of interviewers to respondents in a variety of ways.

2.3 METHODS TO MINIMISE ITEM NON-RESPONSE

Some important areas in which there are clear information gaps include:

- The effect of interviewer continuity on item non-response rates (existing evidence is restricted to non-experimental designs);

- Effects of respondent commitment and methods for attempting to increase commitment, such as variants of respondent "contracts" or instructions;

- The use of indirect methods for certain sensitive items, a good example being the use of 'unfolding brackets' for income items. This and similar approaches might usefully be extended to other types of survey items;

- The impact of in-interview computer-assisted coding for items with complex code frames (occupation, industry, medicaments, etc).

2.4 THE NATURE OF RECALL ERROR

Some of the most fruitful research in this area seems to have arisen from collaborations between psychologists and survey researchers, of which there have been very few in the UK in the past couple of decades. Fortunately, this is likely to be an area where results from other developed countries are broadly applicable to UK too. Important questions remain in some areas, however:

- Processes underlying the recall and reporting of dates, for events of varying saliency and impact;

- Methods for improving the recall of dates and of sequencing of events, including the effects of life events calendars, parallel theme calendars, and other prompting devices.

2.5 MEASUREMENT ERROR THAT IS SPECIFIC TO REPEATED MEASURES

Major questions remain regarding:

- The nature of correlations (over time, within sample units) in measurement error and the consequent impact on estimates of change;

- The causes of correlations in measurement errors;

- The nature of the processes and circumstances that lead to 'seam effects' in longitudinal surveys that aim to obtain continuous status histories (e.g. of

employment, education and other activity, receipt of income from various sources, partnership formation and dissolution, household composition, etc);

- Evaluation of methods for analysis in the presence of seam effects;

- Potential impacts of various types of dependent interviewing techniques for a range of different types of survey items.

Particularly important to the ESRC longitudinal studies are the implications of the above issues - and those listed under recall error - for optimum intervals between data collection waves and for consideration of alternative modes of data collection, including multi-mode options.

2.6 PANEL CONDITIONING

This is an area where relatively little is known. There is much speculation about the nature of conditioning effects and some studies claim to have observed such effects, though these are usually confounded with other effects such as attrition. There are no convincing explanations for the cause of the effects. In consequence, little is known about how best to control any detrimental conditioning effects or how to encourage any positive conditioning effects.

However, this is also an area where empirical research is particularly difficult and/or expensive. We have not been able to identify any obvious research questions that could be addressed by any study on anything other than a prohibitively large scale. We are therefore hesitant to recommend methodological research in this area.

We would note, however, that there is some potential for research based on the Scottish and Welsh boost BHPS samples, and on BEPS in election years, where there are new (unconditioned) and old (conditioned) samples running side-by-side. However, even in these cases, conditioning effects may still be confounded with attrition effects and there are no special features incorporated into the designs to facilitate the study of conditioning effects. Another opportunity might arise with a possible BHPS2, if that were to run in parallel with the original BHPS sample at least

8

for a while. It would be worth considering carefully whether there are special design features that could be incorporated to facilitate the study of panel conditioning.

2.7 INTERVIEWING STRATEGIES

Complex dependent interviewing (DI), harnessing the full potential of Computer Assisted Personal Interviewing (CAPI) and Telephone Interviewing (CATI), is a relatively new innovation. Understanding of the effects of DI has been considerably advanced by recent work at ISER but there is still a need for research on some issues. Important areas include:

- Effects of DI on a range of important estimation problems, including modelling of transitions;

- Alternative ways of asking dependent questions about things that may accrue or decline over time (e.g. qualifications, savings) – viz. asking explicitly about changes (e.g. new qualifications) versus asking, perhaps periodically, about current status (all qualifications held);

- Methods of implementation of DI, including software implications;

- Interviewer perceptions and impacts of DI on interviewer effects and on interviewer-respondent rapport.

The effects of interviewer stability have been studied only on response rates, and this research is limited. Further research on interviewer stability could usefully address:

- Sample subgroups and circumstances in which effects on response rates may be positive (or negative);

- Effects on social desirability bias;

- Effects on satisficing and respondent commitment;

- Effects on correlated interviewer error.

We note that there exists an opportunity to study some of these issues by taking advantage of the BHPS wave 2 interpenetrated design.

2.8 SAMPLING

Aspects of sampling that are specific to longitudinal surveys arise mainly from one or other of two considerations. The first is the opportunity for sample composition to change over the life of a survey, for example in terms of geographical location or other non-permanent characteristics. The second is the particular sample structure requirements of some analysis methods that are appropriate for longitudinal data. Additionally, while it may be clear what population their initial sample represents, it may be much less clear what population is represented as the sample evolves over time. Important issues include the following:

- While simple random sampling (SRS) is held up as an ideal, and there may be particular reasons from a modelling perspective for preferring it for longitudinal surveys, in practice surveys tend to chose more complex designs on cost grounds. However the balance of advantage may be less clear when the spatial declustering of the sample is taken into account. Research is required into the effects of declustering on both data collection costs and design effects.

- There is often a substantive case made for various forms of departure from SRS – e.g. over-sampling of particular groups, or spatial clustering for substantive reasons. It may be that there could be more general guidance on these issues. Again, geographical mobility and change of status over time are particular considerations for longitudinal surveys in this context.

- Longitudinal survey designs are not typically good at capturing the effects of international migration processes (or other movements in and out of the eligible population). There is a need for better information about how much this matters, and what might be done about it.

- For non-monotone sample designs, processes of change in unit composition over time (where units may be households, businesses, schools, etc) generate considerable complexity and uncertainty over eligibility and inclusion probabilities. Research is needed into the nature of the problem and potential data collection and analysis solutions, so that researchers concerned with longitudinal surveys of this sort can have a better understanding of the implications.

3. Attrition and Non-response: Causes and Consequences

The mechanisms that lead to non-response and attrition in panel surveys are, to a certain extent, similar to those that operate on standard cross-sectional surveys. The processes which influence the decision to participate, once re-contacted, have much in common with those that resulting in the decision to participate at the original survey request. However, there are also some specific influences on contact and participation for a longitudinal panel survey. This review considers non-response (refusal or non-contact) at the first wave of a longitudinal survey and attrition (refusal or non-contact at the second and subsequent waves). We examine the causes (sections 3.1 to 3.4), the characteristics of non-respondents (section 3.5) and the effects on survey estimates (sections 3.6 and 3.7).

3.1 CAUSES OF REFUSAL: GENERAL

Groves and Couper (1998) examined the influences on household survey participation in general. Two types of influence are exogenous to the survey: social environmental influences (such as the survey-taking climate), and knowledge and social psychological attributes of the sample persons (e.g., civic duty, interest in the topic of survey). Two influences are features over which the survey organisation has some control: survey protocols (e.g., mode of data collection, incentives, burden), and the selection and training of interviewers. Groves and Couper highlight the importance of the first few moments of interaction between the interviewer and the respondent (see also Dijkstra and Smit (2002), who discuss the influence of the respondent-interviewer interaction on nonresponse in the case of telephone surveys). Linked to these are other concepts and possible influences on survey participation. The "social exchange" concept uses the norm of reciprocation: small gestures on the part of the survey organisation (including incentives) attempt to promote trust and engender in the potential respondent a reciprocating effect - respondents feel that they ought to co-operate. The content of the questionnaire - and the way it is introduced by the interviewer - is important in the "topic saliency" concept. This speculates that when the respondent is interested in the topic of the survey, or they believe that their group might be advantaged by the information, or

they enjoy the chance to exhibit their knowledge about the subject, they are more likely to participate. The role of the interviewer is important in the "interviewer effects" hypothesis. Since interviewers achieve different response rates the suggestion is that some interviewers are better at persuading respondents to participate through tailoring their approach to each individual (Groves, Cialdini and Couper, 1992; Dillman *et al*, 2002).

The "Opportunity costs" theory posits that a sample member weighs up the costs and benefits of participating in a survey. If respondent burden is too high relative to any benefits, then they are likely to refuse the survey request (Groves and Couper, 1998; Groves, Singer and Corning, 1999). On the negative sides are such factors as length (Groves and Couper, 1998; Morton-Williams, 1993) or complexity of interview (Baldinger *et al*, 1995; Meier, 1991; Sharpe and Frankel, 1983; Smith, 1995) and concerns about privacy and confidentiality (Cialdini *et al*, 1993; DeMaio, 1980; Frey, 1986; Goyder and MacKenzie Leiper, 1985; Olson and Klein, 1980; Singer *et al*, 1992; Singer *et al*, 1995; Singer, 1984; but also see US National Research Council, 1979 and Reamer, 1979). On the positive side the respondent may be interested in the topic of the survey, there may be an incentive to participate, or the respondent may have strong feelings of civic duty. Civic duty is relevant, it is argued, because surveys are often conducted by major institutions of the society, and they often seek information that will be freely shared and is of potential value to the society as a whole (Dillman 1978; Goyder 1987). Surveys which are of interest to the respondent – have a high saliency – are likely to be more positively received. Heberlein and Baumgartner (1979) and Goyder (1982) have found that in mail surveys salience is the second most important influence on response rates, after the use of reminders. Salience may be characterised as either the enjoyment of taking part in the survey or the perceived importance of the survey, where the respondent can see the benefit in co-operating.

3.2 CAUSES OF REFUSAL IN THE SECOND AND SUBSEQUENT WAVES OF A PANEL SURVEY

At the second and subsequent waves of a panel survey the sample member has additional information which they did not have at the first encounter because they have had experience of participation in that survey.

In a longitudinal survey, therefore, it is likely that the respondent will use previous waves as a guide to decide whether or not to participate. Hill and Willis (2001) found that, in a health study, less than three-quarters of wave 1 respondents who did not enjoy the interview "at all" responded at wave 3, compared to 90.3% of those who reported having enjoyed it "a great deal". Taylor, Heath and Lynn (1996) found a similar pattern on the British General Election Panel Survey. Hill and Willis found that the largest and most significant factor which predicted response at a future wave was having the same interviewer at each wave. Once the fixed interviewer effects and other variables were controlled reassigning the same interviewer to a sample unit is associated with roughly a six percent increase in response rates (see also Waterton and Lievesley, 1987; Rendtel, 1990 but also see Campanelli and O'Muircheartaigh, 1999 and 2002 who find that any interviewer effects disappear once area effects are controlled). When looking at respondent burden in a longitudinal sense there is an additional component to the questionnaire length; the impact of the perceived cost of future survey participation. Apodaca, Lea and Edwards (1998) found that the effect of the perceived longitudinal burden on the survey participation resulted in a 5 percent decrease in the response rate. Lynn, Taylor and Brook (1997) found that explicitly telling sample members that the survey was longitudinal, but only at a later wave, resulted in slight reduction in response at that wave, but an overall improved net response at subsequent waves.

Bogen (1996) reviews research on the effect of questionnaire length on response rates for the first and subsequent waves. There was some evidence (Frankel and Sharp, 1981; Sharp and Frankel, 1983; Taylor and Lynn, 1996) that interview length did not have a significant effect on subsequent response rates in a panel survey. According to a study by Branden, Gritz and Pergamit (1995), interview length (time) or questionnaire length (number of questions) had either no effect or a positive effect

14

on sample retention rates. However, Zabel (1994) reported that attrition rates were reduced after a decrease in interview length. Kantorowitz (1998) found no increase in refusal rate, either at current or subsequent waves when the Israeli Labour Force Survey was lengthened. She also found no relationship between interview length and response rate when the supplementary questionnaire was drastically shortened. Galvin and King (2000), using the SIPP, found that people whose interviews were short were much more likely to drop out of the panel than people whose interviews were longer. There is, however, very little experimental work in this area.

3.3 CAUSES OF NON-CONTACT: GENERAL

Non-contact in a survey occurs when a sample member is not interviewed because they cannot be located. It may be that the sample member is simply not present at the time the interviewer calls. There may also be some obstacles in the way of the interviewer such as locked central entrances to flats or houses, locked gates and intercoms. For telephone surveys the obstacles include answering phones or services, caller identification features and ex-directory numbers - although this latter obstacle can be overcome using Random Digit Dialling (Bennett and Steel, 2000; Dillman *et al*, 2002; Groves and Couper, 1998; Nicolaas and Lynn, 2002). In cases where the sample member is simply absent from the house on particular occasions, the chance of eventual contact is affected by the interviewer's calling strategy (Purdon *et al* 1997; Swires-Hennessy and Drake 1992; Massey et al, 1996; Weeks *et al*, 1987). Those interviewers who only make calls during working hours on Monday-Friday are unlikely to contact those sample members who are out at work during that time.

3.4 CAUSES OF NON-CONTACT IN THE SECOND AND SUBSEQUENT WAVES OF A PANEL SURVEY

In the second and subsequent waves of a panel survey there is an additional possible source of non-contact over those found in a cross-sectional survey. As well as being temporarily absent or out of the house when the interviewer calls, the sample member may have moved house and changed addresses. The chance of

locating the new address of the sample member depends partly on the efforts of the survey organisation to collect tracking information in the previous wave and partly on the interviewer on the door-step getting information from those currently at the address. There is a difference between addresses where the whole household has moved and those where one (or more) of the original members have moved but some original members remain. In the latter case there is usually a greater chance of finding out the new address of those who have moved (Laurie *et al*, 1999; Lepkowski and Couper, 2002). The geographical mobility of the population under study is likely to have a considerable influence on the extent of non-contact. For example, in the UK panel surveys of young people are likely to be more affected by non-contact than surveys of elderly people. Features of the survey design can also affect the possibility of locating respondents. Such features include the length of the panel, the length of the gap between interviews and the extent and nature of contact between the survey organisation and sample member between waves (Laurie *et al*, 1999; Lepkowski and Couper, 2002; Lynn, 2003; Taylor and Lynn, 1996).

3.5 CHARACTERISTICS OF REFUSERS AND NON-CONTACTS

There is a large literature on the characteristics of non-responders. The two main components of nonresponse are refusal and non-contact and particular demographic characteristics may have different effects on propensity to be contacted and willingness to participate once contacted. Groves and Couper (1998) hypothesise that these demographic variables are indirect measures of social psychological constructs, rather than direct causal influences on participation.

Those more likely to refuse include:

- Elderly (Brehm, 1993; Goyder, 1987; Groves and Couper, 1998; Hawkins, 1975; Cheesbrough, 1993; Foster and Bushnell, 1994; Lepkowski and Couper, 2002)

- Lower incomes (DeMaio, 1980; Allen *et al*, 1991; Ross and Reynolds, 1996; Goyder, 1987; Brehm, 1993; GSOEP; Fitzgerald *et al*, 1998; Iyer, 1981; Nathan, 1999)

- Less educated (Robins, 1963; Dunkelberg and Day, 1973; Wilcox, 1977; Dillman, 1978; O'Neill, 1979; Goyder, 1987; Cheesbrough, 1993; Foster and Bushnell, 1994; Green, 1996; Loosveldt, 1995; Foster, 1998; Lepkowski and Couper, 2002; Fitzgerald *et al*, 1998; Lillard and Panis, 1998; Iyer, 1981; Nathan, 1999; Lynn *et al*, 1994)

- Single (not married) people (Foster and Bushnell, 1994; Goyder, 1987; Foster, 1998; Fitzgerald *et al*, 1998; Lillard and Panis, 1998; Nicoletti and Peracchi, 2002)

- Ethnic minorities (Allen *et al*, 1991; O'Neil, 1979; Foster, 1998; Fitzgerald *et al*, 1998; Iyer, 1981; Lynn and Clarke, 2002; Nathan, 1999)

- Households with high mobility (Goyder, 1987)

- Those in urban areas (Goyder, 1987; Smith, 1995; Steeh, 1981; Couper, 1991; Groves and Couper, 1998; Marton, 1995; Lyberg and Lyberg, 1991; Foster *et al*, 1993; Foster and Bushnell, 1994; DeMaio, 1980; Foster, 1998; GSOEP)

Those more likely to be non-contacts

- Elderly (Hawkins, 1975; DeMaio, 1980; Dohrenwend and Dohrenwend, 1968; Cannell *et al*, 1987; Goyder, 1987; Groves, 1989; Cheesbrough, 1993; GSOEP; Lillard and Panis, 1998)

- Young (Foster and Bushnell, 1994; Foster, 1998; Groves and Couper, 1998; Lynn and Clarke, 2002)

- Men (Brehm, 1993; Goyder, 1987; Foster, 1998; Lepkowski and Couper, 2002; Lynn *et al*, 1994)

- Higher Incomes and/or in employment (Dunkelberg and Day, 1973; Goyder, 1987; Cheesbrough, 1993; Foster and Bushnell, 1994; Lynn and Clarke, 2002)

- Single-person households (Brown and Bishop, 1982; Wilcox, 1977; Goyder, 1987; Couper, 1991; Groves and Couper, 1998; Kordos, 1994; Lyberg and Lyberg, 1991; Kemsley, 1976, Foster *et al* 1993; Foster and Bushnell, 1994; Foster, 1998; Nicoletti and Peracchi, 2002)

- Households with high mobility (GSOEP; Lepkowski and Couper, 2002)

- Those in urban areas (Goyder, 1987; Smith, 1995; Steeh, 1981; Couper, 1991; Groves and Couper, 1998; Marton, 1995; Lyberg and Lyberg, 1991; Foster *et al*, 1993; Foster and Bushnell, 1994; DeMaio, 1980; Foster, 1998; GSOEP)

3.6 BIAS DUE TO UNIT NON-RESPONSE AND ATTRITION

If non-response was completely random then the only concern with lower response rates would be smaller sample size and the associated increase in the *variance* of estimates. Non-response is an issue because it may introduce *bias* to survey estimates, because units that fail to respond may differ from those that do respond and analysis and conclusions can be based only on the latter (Lynn, 2006). For example, a survey which only interviewed respondents at home during working hours on Monday to Friday would have fewer people who work during the day than the population, while the elderly and those with young children would be disproportionately represented.

Non-response bias is only defined relative to a desired estimate or analysis. It is not an inherent property of a respondent sample. In other words, it is quite possible that the same sample could be unbiased with respect to one estimate but biased with respect to another. It depends on the association between the variables from which the estimate is constructed and the non-response propensity of sample units. If an estimate is not associated at all with non-response propensity then the data are said to be *missing completely at random* (MCAR); if the association between an estimate and non-response propensity is completely explained by observed variables the data are said to be *missing at random* (MAR); if the association cannot be explained by observed variables, the data are *missing not at random* (MNAR) (also referred to as

18

non-ignorable non-response). If data are missing at random, then relatively simple techniques such as weighting can be used to eliminate the biasing effects of non-response. It is typically difficult to demonstrate convincingly that data are MAR, but the assumption may be more likely to hold on a longitudinal survey, at least for non-response subsequent to the first wave, due to the wealth of data typically available for non-respondents from previous waves.

One way of assessing the characteristics of those who are missing from surveys is to compare the profile of respondents with some reliable external data source relating to the survey population. The Office for National Statistics is able to match sample members on their government surveys with census data. This method of validation can be used to identify what biases may exist in the survey sample. The Labour Force Survey User Guide (Volume 1: Background & Methodology) reports that the survey under-represented those living in London, those renting (from a Housing Association or with a job), those living in converted or shared accommodation, households where there were three or more cars or vans available, households consisting of just one adult aged 16-19, households where the head is single and where the head was born in the New Commonwealth. In contrast, households which were over-represented were those with three or more children in the household, households where there was a child aged 0-4, couples with dependent children, and where the head of household had a degree or equivalent. Foster (1998) has analysed the first three Census match studies and found that an increase in the propensity to refuse is associated with lack of academic qualifications, London residents, households with an older head and ethnic minorities.

Such studies are limited in that bias can only be assessed with respect to the types of variables that typically are not themselves used to construct key estimates, but rather are used as explanatory variables. For longitudinal surveys, key estimates are typically dynamic measures such as rates of change or measures of association between other variables and rates of change. It is rare to be able directly to assess the nature of non-response bias with respect to such estimates. Rather, various modelling approaches have been employed to study the nature of attrition bias. Rendtel (2002) concludes that the likely effect of panel attrition is "an over-

representation of stability". However, he suggests that the potential bias of attrition will be small because above average risks of attrition are restricted to small sub-groups of the population. Sisto and Rendtel (2004) find bias due to non-response at wave 1, but very little attrition bias, in terms of parameters of income distribution. Similarly, Behr et al (2004) conclude that attrition in the ECHP introduces very little bias into estimates of income inequality and change and find that "conclusions about the ranking of European countries are almost unaffected by attrition, which is a positive result for comparative analysis." Crouchley et al (2005) use double selection modelling of BHPS data to investigate the effect of attrition on employment participation and wages and compare the findings with those from OLS and Heckman selection modelling. They find that data are MNAR but that this has little substantive impact on their analyses. Jones et al (2006) use variable addition tests and inverse-probability weighting for non-response to examine the effects of attrition on estimates of dynamic models of the association between socio-economic status and self-assessed health, using both BHPS and ECHP data. They conclude that, "on the whole, there are not substantive differences in the average partial effects (APE) of the variables of interest."

While the general conclusion from the studies mentioned here might be that attrition rarely seems to introduce substantively-important bias, it is clear that the effects can be analysis-specific. For this reason, analysts are increasingly incorporating into their estimation a range of techniques to control for attrition, or at least to explore the sensitivity of their results to assumptions about attrition. In addition to the various selection modelling approaches mentioned above, and estimate-specific weighting approaches, these can include forms of multiple imputation and doubly-robust estimation (Carpenter et al, 2006). Such techniques are not reviewed here as they may be considered to fall within area I as outlined in section 1 of this document.

3.7 BIAS DUE TO ITEM NON-RESPONSE

Item non-response can be a serious source of non-sampling error (Pohl and Bruno, 1978). Those who participate in the survey but then refuse to answer particular questions may differ in some way from those who answer every question. Those

who have item nonresponse are likely to have lower involvement with the topic of the survey (O'Neil, 1979; Singer, 2002; Stinchcombe *et al*, 1981). It has been found that in medical surveys those who don't respond and those who respond but have high levels of item nonresponse are more likely to have poor health (Colsher and Wallace, 1989; Grotzinger *et al*, 1994). Lillard *et al* (1986) studied item non-response in the US Current Population Study. They found that those who did not answer a number of questions were more likely to have low income, whilst those who only refused the income question and no others were more likely to have high income (also Biewen, 2001). Duncan and Mathiowetz (1985) found that, in a validation study of the PSID, those who did not respond at all (unit non-response) had earnings that were higher than the average whilst those who responded, but refused the income questions (item non-response) had lower than average earnings. In addition, workers are likely to under-report time spent unemployed and over-report working hours. Souza-Poza and Henneberger (2000) found that, in telephone surveys, there was greater item non-response for those respondents with lower education, and among the self-employed and home owners.

3.8 FUTURE RESEARCH

Promising areas of future research might include the following:

Understanding the reasons for attrition due to refusal – especially identifying experiences which can prompt a refusal at the subsequent wave and other relevant factors which can be infuenced by design and implementation;

Better understanding the nature of non-ignorable attrition due to geographical mobility, for example through obtaining auxiliary or validation data.

21

4. Methods to Minimise Attrition

Attrition in panel surveys is caused when a sample unit fails to participate in the survey at the second or subsequent wave. This section examines the survey design and implementation features that have been found to influence attrition. The most common reasons for this are that the sample unit cannot be contacted, or that the sample unit is contacted but decides not to co-operate with the survey request.

4.1 CONTACTING THE SAMPLE: GENERAL

The elements of survey design which appear to have most effect on contact are the interviewer and the procedures established to track respondents between waves. To keep non-contact rates at a minimum it is important for the interviewer to have an effective calling strategy. In the case of face-to-face surveys, studies by Campanelli *et al* (1997), Swires-Hennesy and Drake (1992) provide indications of the elements of calling strategy that affect non-contact rates. For similar evidence in the case of telephone surveys, see Bennett and Steel (2000), Dennis *et al* (1999) and Massey et al (1996). Furthermore it is important that an interviewer is willing to make multiple calls to the same address (Groves and Couper, 1998; Lynn *et al*, 2002). Survey organisations tend to require that their interviews make a minimum number of calls, with at least one weekday evening and one weekend call (Dillman *et al*, 2002; Lievesley, 1986; Morton-Williams, 1993; Steeh, 1981; Weeks *et al*, 1987). The monitoring of fieldwork has been shown to have a positive affect on contact rates (de Leeuw and de Heer, 2002) and also gives the survey organisation an accurate picture of the progress of fieldwork. A clustered sample is beneficial for contact rates, since an interviewer is likely to have a number of addresses in the same area. In this case, calling on sample households on the way to or from an appointment can be efficient (Lynn, 2003; Morton-Williams, 1993).

**4.2 CONTACTING THE SAMPLE IN THE SECOND AND SUBSEQUENT WAVES OF A
LONGITUDINAL SURVEY**

Contacting – or re-contacting – a sample member in the second and subsequent
waves of a panel survey is slightly different to contacting them at the first stage. The
interviewer may have useful information as to the best times or days to call to
contact the sample member (Lepkowski and Couper, 2002). However, the
interviewer is looking for a named person rather than just whoever is living there at
the time and so if the sample member has moved house the survey organisation
must attempt to trace them and interview them at their new address.

In the UK it is estimated that around 10% of households change addresses each
year (Laurie *et al*, 1999). To minimise the attrition that this causes, a survey
organisation should have procedures in place to track respondents who move
between waves otherwise a survey would be likely to lose the more mobile sections
of the sample. It is standard practice to acquire contact details of the respondent and
someone outside the household at the end of each interview. Between waves of the
survey it is useful to keep in contact with the respondent through the use of thank-
you letters, respondent reports, birthday cards or change-of-address cards (Burgess,
1989; Laurie *et al*, 1999; Lepkowski and Couper, 2002; Schousa-Glusberg and Hunt,
1992). The Centre for Longitudinal Studies used a number of additional databases
when they were tracking respondents on the NCDS and the BCS70. The additional
sources of information they used included: telephone number databases, postcode
databases, electoral register, NHS Central Register, emigrations and deaths
registers, Health Authority address records, Driver and Vehicle Licensing Agency
(DVLA), Ministry of Defence records and media appeals. The extent of these
tracking procedures reflects the increased length of time between contacts.
Increasing the gap between contacts has been found to affect the contact rate
negatively (Burgess, 1989; Cantor, 1989; Lepkowski and Couper, 2002; Taylor and
Lynn, 1996).

4.3 GAINING CO-OPERATION: GENERAL

When it comes to gaining co-operation the general philosophy is to increase and emphasise the positive aspects of participation whilst decreasing and de-emphasising the negative aspects (Groves *et al*, 2000; Lynn, 2003; Morton-Williams, 1993). The survey should be introduced as being interesting, enjoyable and serving a useful purpose (social rather than market research). The topic of the survey may also affect the sample member's decision to co-operate, with a non-responder to one type of survey being a responder to a different type of survey (Baldinger *et al*, 1995; Carton *et al*, 1996; Donald, 1960; Goyder, 1987; Hox *et al*, 1995; Kojetin *et al*, 1993; Lyberg and Lyberg, 1991; Lynn and Clarke, 2001; McDaniel *et al*, 1987). An advance letter serves to add to the legitimacy of the survey request and may be used to provide assurances of confidentiality and to provide the respondent with a method of contacting the survey organisation to check the validity of the survey (Dillman *et al*, 2002; Groves *et al*, 1992; Groves and Couper, 1998; Luppes, 1994; Morton-Williams, 1993). The interviewer must be able to address legitimate concerns of the sample members and to tailor their approach to the individual respondent (Groves and Couper, 1998). The interviewer should also be as flexible as possible with respect to the time of day, or day of the week, the interview takes place and should be prepared to make more than one visit to get an interview. Interviewers should also be willing to retreat from a household before a soft refusal hardens (Morton-Williams, 1993).

There is a great deal of experimental literature concerning the use of incentives in surveys. The balance of research suggests that pre-paid incentives are more effective than promised incentives and that cash is more effective than gifts (Berk *et al*, 1987; Berlin *et al*, 1992; Church, 1993; Goyder, 1987; Groves and Couper, 1998; James, 1997; Mack *et al*, 1998; Singer, 2002; Singer *et al*, 1999). Incentives appear to work directly on respondents, rather than their effect being mediated through the interviewer (Lynn, 2001; Singer, 2002; Singer *et al*, 2000). Singer (2002) notes that incentives increase response rates by persuading those who would not normally take part to participate. The effect is especially strong for those in lower income groups and those for whom the subject matter of the survey is not of interest. Singer

concludes that "dependent variables would be seriously mismeasured if incentives had not been used".

4.4 GAINING CO-OPERATION ON THE SECOND AND SUBSEQUENT WAVES OF A LONGITUDINAL SURVEY

Sample members at the second or subsequent waves of a longitudinal survey are in a different position to those at the first wave. The sample member has had experience of the previous survey and so is in a position to know what to expect. They can make an evaluation of whether to participate in the survey based on how they felt about the previous survey. If the first survey was too long, the questions too personal (in the judgement of the sample member) or they just did not like the interviewer, then this is likely to reduce the chance of participating in the second wave (DeMaio, 1980; Hill and Willis, 2001; Kalton *et al*, 1990; Laurie *et al*, 1999; Nederhof, 1987; Schleifer, 1986). Survey length and complexity has been found to be a common complaint from respondents who are reluctant to participate in a survey (Baldinger *et al*, 1995; Meier, 1991; Sharpe and Frankel, 1983; Smith, 1995). In general longer interviews are related to respondents being less likely to co-operate at a future interview. However, there is some evidence that a very short questionnaire may actually reduce response rates on a panel survey (Lynn, 1993; Taylor and Lynn, 1996). In a panel survey, then, the details of the survey (length, subject and so on) affect not only the response at the present wave but also response at future waves.

Sample members in a panel survey also experience a greater respondent burden, since they are asked to participate in a series of surveys (Lepkowski and Couper, 2002). There is, however, very little research about the timing and the introduction of the survey as a panel survey and the effect that this has on initial and subsequent response rates (Lynn *et al*, 1997).

4.5 FUTURE RESEARCH

Whilst there has been a great deal of research into refusal and non-contact in surveys this has, in the main, focused on cross-sectional one-time surveys. Much of this work is also applicable to longitudinal panel surveys in that sample members still need to be contacted and, once contacted, persuaded to participate. There is less research looking at aspects specific to longitudinal surveys and much of what exists deals with contact and tracking from a fieldwork administration perspective. The area of research which seems to be under-developed is that of the effect of a longitudinal survey – rather than a cross-sectional survey – on a sample member's willingness to participate. Relevant issues here include the use of "tailored" approaches to the request for participation in each wave, based on information provided by the respondent in previous waves, "matching" of interviewers to respondents, ways of introducing the nature of panel membership, and better understanding which aspects of the interview experience affect future willingness to co-operate. Regarding contact, there is scope for investigating methods of better harnessing a variety of possible methods for keeping in touch with sample members between waves, again including tailored methods, perhaps based on model predictions of the likelihood of a change of address.

5. Methods to Minimise Item Non-response

5.1 WHAT IS ITEM-NON-RESPONSE?

When a sample member does not participate in the survey at all, either through being a non-contact or refusing, this is called "unit non-response". It is possible that someone who is willing to take part in the survey will refuse to answer one or more particular questions; this is "item non-response". In the main, item non-response is not as serious as unit non-response because much or most of the survey has been answered. However, if the sample member has refused on a question then any analysis which uses that question will be affected. If those sample members who refuse one particular question, for example income, are different to those who are happy to answer then there is a chance that any analysis which includes income will be biased. As Dillman *et al* (2002) note, the definition of item non-response can be complicated when it comes to considering what answers constitute non-response. For example "don't know" responses may count as non-response if it is considered by the researcher that the respondent has the necessary information to answer. However there are certain questions where "don't know" is a valid response (deLeeuw, 1999; Krosnick, 2002).

5.2 CAUSES OF ITEM NON-RESPONSE

Lynn and Thomas (2006) identify the reasons for item non-response as follows:

- *Unwillingness to answer:* The sample member does not want to answer the question, perhaps because the information is perceived as sensitive;

- *Inability to answer:* The sample member can not answer the question, typically because the answer is unknown to them;

- *Accidental omission:* The sample member misses a question by accident, perhaps because they skipped it, intending to come back to it at the end, but then forgot to return to it.

- *Partial response:* The respondent may not have any problem with a particular item as such, but may simply have got tired of, or otherwise had to give up, answering questions and abandoned the questionnaire or the interview, or a whole section of it.

Beatty and Herrmann (2002) suggest that item non-response depends on the outcome of two decisions made by the respondent when confronted with a question: whether they *can* answer (cognition) and whether they *will* answer (motivation) (see also deLeeuw, 1999; Sudman *et al*, 1996). Questions which require a great deal of thought or effort to answer, or questions which the respondent feels are too sensitive or intrusive or just uninteresting may be more likely to be refused or given a "don't know" response (Cannell *et al*, 1981; Krosnick, 1991).

Research has shown that item non-response is more likely to occur when the survey is self-administered rather than interviewer-administered (Dillman, 2000; Tourangeau *et al*, 2000). Lillard *et al* (1986) found that income was more likely to be reported in face-to-face rather than telephone interviews (also Groves, 1989; Jordan *et al*, 1980; Schräpler, 2002). The instructions to interviewers may also affect item non-response, particularly in the use of cues, prompts and probes. An interviewer who gently probes a "don't know" response may get a lower item non-response rate than an interviewer who accepts the "don't know" and carries on. There is also some empirical research to suggest that item non-response rates tend to be slightly higher for female interviewers (Riphahn and Serfling, 2002; Schräpler, 2001; Sousa-Poza and Henneberger, 2000). The presence of people other than the interviewer and respondent during the interview may also affect item non-response, especially if the third person is involved in the topic of the question (Schräpler, 2001).

There are some sorts of questions which are more likely to attract item non-response than others. Tourangeau *et al* (2000) note that questions about personal or household income, sexual behaviour, alcohol consumption and illegal activity are all more likely to elicit item non-response (see also Bradburn *et al*, 1978; DeMaio, 1985; Turner, 1982). Part of this may be due to concerns about confidentiality and privacy of the data and the interview situation (Schräpler, 2002; Singer *et al*, 1993). Another

concern is social desirability – the desire to appear in a positive light to the interviewer, the need to conform to social standards or the wish for social approval. For example, in a society which places a great deal of attention on material wealth there may be some reluctance for those at the lower end of the income scale to reveal their earnings (Schräpler, 2002; Smith, 1991; Tourangeau *et al*, 2000; Wagner and Motel, 1996). There is a danger that item non-response to sensitive questions that are strongly related to the topic of the survey as a whole may lead to (or at least be related to) unit non-response at a future wave of a longitudinal survey (Bollinger and David, 2001; Burton *et al*, 1999; Neukirch, 2002; Schräpler, 2002).

Certain types of people are also more likely not to respond to a particular item. Dillman *et al* (2002) identify the elderly and those with less education as being less likely to provide responses. Schräpler (2002) supports this picture, but also notes that there are differences between those who refuse to answer income questions and those who give "don't know" answers. In general, those who refuse are more likely to be men without dependent children in their household whilst those who say they "don't know" are more likely to be women, those in low or middle occupational states and those who work irregularly. Those who are reluctant to participate in the survey are also less likely to respond to every question (Loosveldt *et al*, 1999; Mason *et al*, 2002; Triplett *et al*, 1996). This latter finding implies that additional efforts to reduce unit non-response may result in increased item non-response.

5.3 PROCEDURES FOR REDUCING ITEM NON-RESPONSE

When dealing with potentially sensitive questions researchers and fieldwork agencies have usually looked at changing the structure, mode of administration, or the wording of the question. Shoemaker *et al* (2002) suggest that to decrease the number of refusals researchers should pay attention to cognitive effort required as well as sensitivity whilst the number of "don't know" responses are affected mainly by cognitive effort alone. Using income bands instead of asking for exact figures and using close-ended rather than open-ended questions have both been found to reduce item non-response (Dillman, 2000; Locander and Burton, 1976). Incentives may be used to increase unit response rates but also to reduce item non-response.

Singer (2002), in her review of the literature on incentives, finds that respondents who receive incentives may have less item non-response data and provide longer open-ended responses (Baumgartner *et al*, 1998; James and Bolstein, 1990; Singer *et al*, 2000; Shettle and Mooney, 1999 but see also Tzamourani and Lynn, 2000). Hill and Willis (2001) find that reassigning interviewers to the same respondents in a panel survey reduced item non-response as a level of trust is built up between interviewer and respondent (also see Schräpler, 2002).

5.4 FUTURE RESEARCH

There are several areas in which there are clear information gaps. One concerns the effect of interviewer continuity on item non-response rates. Though there have been some studies of this issue, they have been restricted to non-experimental designs. Another potential area for further research concerns the effects of respondent commitment and methods for attempting to increase commitment. An example of such a method is the respondent 'contract' proposed by Cannell and Henson (1974) and Cannell et al (1981).

For certain sensitive questionnaire items, a better understanding of the use of indirect questioning methods is required. An example of such a method is the use of 'unfolding brackets' for income items. This, and similar approaches, might usefully be extended to other types of survey items. Another technique that could profitably be used to reduce rates of item non-response is dependent interviewing. This is discussed below in section 9.

Finally, one other source of item missing data is the inability to code answers to open-ended questions in which answers are recorded verbatim. In the case of items with complex code frames, the impact of in-interview computer-assisted coding methods should be investigated.

6. The Nature of Recall Error

Recall error is a term often used to refer to a failure by respondents to accurately report events or circumstances from their past. In reality, an observed failure to report accurately could be caused either by an error of recall or by an error in reporting or recording the recalled information. Using repeated measures designs shortens the recall periods, thereby improving the accuracy of life event data (Menard, 2002). Nonetheless, retrospective information is required to fill the gap between interview points. Additionally, longitudinal surveys may collect retrospective life history information at some point (often in the first interview), in which case the data will suffer from the same problems as cross-sectional surveys collecting this type of data.

The development of new and sophisticated methods of investigating dynamic causal relationships of life events has increased the interest in the causes of recall errors and their implications for the accuracy of information on the timing and sequencing of past events and experiences (Freedman *et al*, 1988). Some of the most fruitful research in this area stems from collaborations between cognitive psychologists and survey researchers. This section provides an overview of different types of recall errors, factors influencing these, and techniques used in the design of surveys to reduce these sources of errors.

6.1 TYPES OF RECALL ERRORS

Respondents may simply forget that an event occurred, or not remember details regarding the circumstances or timing of an event. Forgetting may occur because the "memory trace", the original means for accessing the information, no longer exists and/or because the retrieval cue was not appropriate (Eisenhower *et al*, 1991). The path of forgetting autobiographical life events suggests a linear decay with time (Linton, 1982; Thompson *et al*, 1996; Wagenaar, 1986). In general, recall decay is thought to have a large impact when the reference period is long or when the events of interest are frequent routine events. While trace-dependent forgetting cannot be affected by the survey conditions other than the recall period (which typically

coincides with the interval between waves), cue-dependent forgetting implies scope for improvement of survey instruments.

Telescoping refers to the tendency of respondents to report events as having occurred more recently (forward telescoping/time compression) or earlier in time (backward telescoping/time expansion) than is actually true. Evidence on telescoping for consumer expenditures is found by a number of studies in the 1950s and 1960s (Cole and Uttinger, 1956; Ferber, 1955; Jaeger and Pennock, 1961). As Dex (1995) notes many of these studies use the apparent overestimation of the frequency of events to demonstrate telescoping – rather than errors in the dates. Neter and Waksberg (1964) were among the first to measure the direction and magnitude in which household respondents tend to shift the timing of expenditures. They found net forward telescoping of expenditures, in other words, more expenditures were moved forward in time than to earlier time periods. Additionally, the telescoping effect was found to increase with the size of expenditures. In other studies, telescoping is predominantly observed for short reference periods or non-routine events (Eisenhower *et al*, 1991). The use of the term 'telescoping' suggests that events simply tend to seem nearer in time than they actually are, implying an imperfect translation from real time to subjective time (Sudman and Bradburn, 1974). However, the telescoping effects observed in many studies need not be caused by systematic error in the dating of events (Baddeley *et al*, 1978; Huttenlocher *et al*, 1990; Rubin, 1982; Rubin and Baddeley, 1989). Friedman (1993) concludes that memory for time does not consist of special temporal codes or a chronologically organised memory store. Instead a chronological sense of the past results from ongoing reconstruction based on contextual information. Memories associated with the event are combined with separate knowledge of time patterns to infer the likely date of the event. Larsen and Thompson (1995) conclude that the memory of the personal context is more important for the accuracy of dating events, than the memory of the core of the events. Huttenlocher *et al* (1990) and Bradburn *et al* (1994) suggest that for certain types of events, respondents are more likely to think in terms of elapsed time and then translate this into calendar time. Elapsed time is commonly rounded to convenient (prototypic) units, where the distance between adjacent rounded values increases with elapsed time. This leads to asymmetrical

rounding, contributing to net forward telescoping[1]. The literature on telescoping has focused on the timing of individual events. However, as Bound *et al* (2001) note, errors in the timing of different events reported by a respondent are likely to be correlated, creating a potentially spurious coincidence of events. For event history analysis this is likely to be a serious issue.

Autobiographical survey questions often ask the respondent to recall frequencies of events or make 'mathematical' judgements (e.g. number of visits to GP in a given period, average expenditure on food) (Eisenhower *et al*, 1991). In answer to such questions, the respondent may attempt to recall the frequencies/amounts, or use an estimation strategy. Recent investigation suggests that respondents use a retrieval strategy which is a function of the true behavioural frequency (Bound *et al*, 2001). Blair and Burton (1987) and Burton and Blair (1991) indicate that respondents count events or items (episodic enumeration) if the frequency of the event/item is low, and choose to estimate more frequently occurring events. However, the point at which the respondent switches from episodic counting to estimation varies both with the characteristics of the respondent and the characteristics of the event. The authors discuss various aspects of survey design and the survey process that might influence the retrieval strategies used by respondents. However, as Bound *et al* (2001) point out, they focus on strategies that are likely to apply only to events with relatively high frequencies. Sudman *et al* (1996) criticise that no studies have investigated the relationship between individual characteristics, such as intelligence, education, or preference for cognitive complexity, and the choice of counting or estimation, given a number of events.

An event may be recalled but misclassified, or remembered inaccurately. As an example, individuals may give inaccurate information regarding their status at a particular date in the past (Dex, 1995). Similarly, they may report an event although it

---

[1] If, say, "about 6 months ago", "about a year ago", and "about 2 years ago" are used as common rounded values, events that occurred between 9 and 18 months ago will be reported as being "1 year ago". As a result, 6 months of events are misplaced forwards and only 3 months of events are misplaced backwards in time.

doesn't meet the criteria for an eligible event, or fail to report an eligible event because they did not think it was eligible (Lynn, 2003).

Finally, there are panel-specific errors in dating events. Seam effects refer to the phenomenon that a disproportionate number of changes in a respondent's status are reported (or inferred to have taken place) at the 'seam' between the end of the reference period for wave *t* and the start of the reference period for wave *t+1* of a panel survey. Section 7 of this paper provides an overview of the literature on seam effects and dependent interviewing techniques used to reduce this source of error.

6.2 FACTORS INFLUENCING THE EXTENT AND NATURE OF ERRORS

Many studies comparing the characteristics of events which are reported with those not reported, identify time as a fundamental factor: the longer the recall period, the greater the expected bias caused by respondent retrieval and reporting error. Bound *et al* (2001) refer to studies of the impact of time on the reporting of consumer expenditures and earnings, hospitalisations, visits to physicians, health conditions, motor vehicle accidents, crime and recreation. However, the literature also offers several examples in which the length of the recall period had no effect on the magnitude of response errors (e.g. Mathiowetz and Duncan, 1988; Schaeffer, 1994), or where the recall is found to improve with the elapse of time (Mason *et al*, 1976). Bound *et al* (2001, p.3744) conclude that these "investigations point to the importance of the complexity of the behavioural experience over time, as opposed to simply the passage of time, as the factor most indicative" of recall error.

The saliency of events or behaviours to be retrieved is a further essential component. In survey and psychological research, 'saliency' of an event is often measured by the emotion associated with it or the extent to which the event differs from related schema or the respondent's life in general (Eisenhower *et al*, 1991). Sudman (1975) distinguishes the saliency of an event according to its uniqueness, its economic or social costs and benefits, and ongoing consequences. The salience is likely to affect the strength of the memory trace and thereby the effort involved in retrieving the information from long-term memory: the stronger the trace, the lower

the effort required to locate and retrieve the information. Evidence that more salient events are recalled more accurately is, for example, found by Cannell and Henson (1974), Mathiowetz and Duncan (1988), Cannel, Fisher and Bakker (1965) and Waksberg and Valliant (1978). However, although salient events or behaviour may be subject to lower levels of errors or omission, there are also indications that salience can lead to overestimation on the part of the respondent (e.g. Chase and Harada, 1984). Pleasant events are recalled more accurately than unpleasant events (Skowronski *et al*, 1991), while both positive and negative events are recalled with more accuracy than neutral ones (Thompson, 1985). Correspondingly, divorce histories are recalled less accurately than marital histories (Peters, 1989). The saliency of the respondent's mood during the event or about a category of events also affects the respondent's recall ability (Eisenhower *et al*, 1991).

Other studies find that what happens to the recall ability over time depends on what occurs in the gap (McGeoch, 1932). As a respondent experiences an increasing number of similar or related events, the probability of recalling any one of these events declines (Eisenhower *et al*, 1991; Himmelweit *et al*, 1978). The respondent may lose the ability to distinguish between related events, or recall less detail of any one event. Interference can also take the form of receiving new information related to the event of interest, or learning the outcome of an event.

According to Menon (1994), retrieval strategies are also determined by the degree of regularity and similarity among events. Events which are dissimilar and occur irregularly require more cognitive effort. Similarly, the more habitual the response, the more accurate the recall is found to be (Tourangeau, 1984). Bound *et al* (2001) comment that much of the assessment of the quality of household-based survey data regards the reporting of discrete behaviours (i.e. events). However, many of the measures that are subject to survey inquiries could be considered trait phenomena. The cognitive processes of formulating answers to questions on discrete behaviours

are likely to be very different from those used in response to queries on traits[2]. In addition, individuals are likely to vary regarding their classification of phenomena. For example, a question on current wage may be an easy question for a respondent for whom income is a self-defining characteristic. For others, reporting wages might be linked to a discrete behaviour (the last paycheque), a recent rehearsal of the information (reporting wages for an application) or the construction of an estimate based on the retrieval of relevant information. Error is further found to increase with the difficulty of the task. Analysing unemployment spells, Mathiowetz and Duncan (1988) find that errors become frequent as the task becomes more difficult, either because there are more frequent spells or shorter spells of unemployment. Finally, the characteristics of respondents can also have significant impact on recall. Women have been found to be better at dating personal and other events (Skowronski *et al*, 1991; Cannell *et al*, 1965; Mathiowetz and Duncan, 1988). Race was found to affect recall in a study of marital histories and educated women were found to report marital history information more consistently (Peters, 1989), while those with less education have been found to underreport crimes (Sparks, 1981).

6.3 TECHNIQUES USED TO REDUCE THE IMPACT OF RECALL ERROR

The length of the reference period may be varied in order to balance the risk of recall decay versus telescoping. Respondents expected to have experienced many events can be asked to recall them for shorter recall periods than respondents with few events (Eisenhower *et al*, 1991). Gray (1955) concludes that reference periods ending on the day before the interview are to be preferred, since this at least rules out misplacement over the boundary at one end of the period.

Interviewing the respondent as close to the event of interest as possible is the best way of avoiding bias through intervening events. Alternatively, trying to recreate the mood and context the respondent experienced at the time of the event, and slowing

---

[2] As an example, the cognitive process by which a respondent formulates an answer to a question on current occupation is different from the process related to reporting the number of times they have consulted a doctor during the last year.

down the pace of the interview encourages respondents to think back and distinguish between intervening events and the event of interest. Using memory aids and dividing the period into shorter periods may also help reduce bias due to intervening events (Eisenhower *et al*, 1991).

Sudman and Bradburn (1973) state evidence that face-to-face interviews improve memory, but increase telescoping compared to self-administered interviews. Dillman and Tarnai (1991) additionally find that recall is slightly better in mail than in telephone surveys. There is also some evidence that recall is better when open-ended rather than closed questions are used (Bradburn and Sudman *et al*, 1979), and when the survey design encourages respondents to think about the context (Fisher and Quigley, 1990). Respondent recall can further be improved by asking questions in narrative style, or in a way which is compatible with the retrieval strategy used by the respondent (Means *et al*, 1991), by reducing the difficulty of the question and using a logical and natural sequence in their order (Tourangeau, 1984), and by asking questions about specific actions before asking about a general action (Reiser *et al*, 1985).

Furthermore, the greater the threat of the interview or question, the greater the errors of omission and the more negative the response effect as the recall period increases (Sudman and Bradburn, 1973).

In addition to the general questionnaire design, different tools have been used to stimulate respondents' memory. Aided recall refers to the use of specific cues, such as lists (e.g. of income types or qualifications), pictures or maps. Sudman and Bradburn (1973) find that aided recall techniques reduce the omission of activities, but may increase telescoping effects, leading to potential overstatements. Aided recall is therefore most helpful for less important events and for longer recall periods when telescoping effects are small. Additionally, aided recall is found to become more beneficial as the reference period increases[3]. Using records of the dates of

---

[3] They have insufficient data to confirm this effect for records.

events to support respondent recall has opposite effects: telescoping effects are eliminated, while omissions are hardly reduced. Records are consequently most appropriate for major events where omissions are unimportant and telescoping is the major source of error. However, the authors point out that even with aided recall the omission rate may be high and that other methods, such as diaries, should be used.

Bounded recall refers to the use of memory-bounding techniques to prevent the shifting of events in time, thereby reducing the impact of telescoping (Sudman and Bradburn, 1973). Neter and Waksberg (1964) use bounded recall procedures to reduce the telescoping of house repairs and alterations. At the start of the bounded interview, which is the second or later interview, the interviewer tells the respondents the expenditures which were reported during the previous interview, asking for expenditures since then. During and after the interview the interviewer checks for and queries duplicate reports of expenditures. Comparing different treatment groups, the authors find that the telescoping effect increases with the size of expenditures, and are reduced in the bounded interviews. Bounding techniques are employed by the NCS (US National Crime Survey), which doesn't use the first interview to estimate rates of victimisation (Menard, 2002)

Dependent interviewing (DI) is an extension of bounding techniques, used to reduce the impact of misplacement in time by respondents, for example on 'seam effects', the disproportionate number of changes in status reported at the 'seam' between two interviews. The method involves 'feeding forward' information collected at a previous interview and using it either in formulating the questions or in an in-interview post-response edit check. Section 7 of this paper reviews the literature on seam effects and independent interviewing techniques.

A further set of techniques involve the use of landmark events, calendars and the construction of temporal reference systems. These techniques improve the accuracy of reported events, because they stimulate memory and the personal 'landmarks' act as an anchor to increase the accuracy of dating other events (Baddeley *et al*, 1978; Means *et al*, 1988; Thompson *et al*, 1987; Belli *et al*, 2001; Belli *et al*, 2004). Loftus and Marburger (1983) found that using the eruption of Mt. St. Helens as the

landmark event for different studies significantly reduced the level of forward telescoping of crime victimisation. Van der Zouwen *et al* (1993) found that using a simplified personal time-line on which respondents marked personal events (birthdays, vacations) improved the recall of periods of schooling. Means *et al* (1989) similarly found improvements in the accuracy of recalled dates. Alternatively, temporal reference systems to which the events being recalled relate (e.g. term structure of the academic year) can be used (Sehulster, 1989; Robinson, 1986; Thompson *et al*, 1988).

Life or event history calendars (LHC/EHC) have been used increasingly due to the growing interest in the analysis of life course events and the development of more sophisticated statistical methods of analysis. A LHC is usually a separate document that combines the charts for a number of different event histories. The calendar format is generally a large grid, one dimension of the matrix being the behavioural patterns investigated, and the other dimension being time units. Freedman *et al* (1988) argue that LHCs "can improve the quality of the retrospective data by helping the respondent to relate, both visually and mentally, the timing of several kinds of events. Events more readily remembered, such as marriages, births, and changes in geographical residence, provide important reference points for recalling less salient events, such as details of employment and living arrangements." Additionally, the calendar display brings inconsistencies to the attention of both respondents and interviewers. Belli (1998) illustrates that the LHC reflects the structure of autobiographical memory. Compared to traditional survey questions that tend to segment life course domains and do not reflect the interrelatedness of events, LHCs improve respondent recall by encouraging retrieval of past events via multiple pathways that act top-down in the hierarchy, sequentially within and in parallel across life themes. However, the use of LHCs holds specific problems of coding, detection and resolution of data inconsistencies, construction of variables for analysis from detailed data, and possible additional costs.

Freedman *et al* (1988) use a LHC to collect retrospective life course data in the context of a panel survey. Comparing the LHC data with current activities reported five years earlier, the authors argue that the LHC yields accurate retrospective data.

They find very few non-response items and a high correspondence between the 1980 and 1985 responses. The authors conclude that the degree of volatility of the activity patterns seems to be important for the accuracy of recall data: highly variable events (e.g. employment) are likely to be measured less accurately than other events. However, since the LHC allows linking volatile events to less varying activity patterns, it nonetheless improves the accuracy of the data. Belli *et al* (2001) report the first experimental comparison between retrospective data obtained from a LHC and from traditional question lists. Respondents were interviewed by telephone and asked to report social and economic events that had occurred during the two previous years. The data was compared with information about current situations and activities obtained from the same respondents one year earlier. The results show that overall the LHC improved the quality of retrospective data compared to the question lists.

6.4 CONCLUSION

While the literature provides some general guidelines on how to design questionnaires in order to minimise the impact of recall errors, most measures attack one potential source – and may have adverse effects on other types of errors. It is therefore crucial to understand which types of errors are likely to affect the response to a particular question and to assess the relative importance of different sources of errors. Important questions remain regarding the processes underlying the recall and reporting of dates for events of varying saliency. Further research is also necessary to investigate the effects on survey estimates of techniques used to reduce recall error for different recall periods and interview modes.

Some of the more recent advances in calendar and aided-recall methods, particularly their computer-assisted versions, have not been tested in the UK.

7. Measurement Error that is Specific to Repeated Measures

Each wave of a panel survey is affected by measurement error in the same way as any cross-sectional survey. Depending on how the panel data are used, the errors may, however, manifest themselves and impact on analysis in different ways. Panel data can be used to derive repeated measures (for example, of the respondent's employment situation, health condition or attitudes at the time of each interview) or to derive continuous histories (for example, of education, receipt of income from various sources or partnership formation and dissolution). This section reviews what little is known about the nature of measurement error in repeated measures and then focuses on the effect of measurement error on continuous histories. We provide an overview of the processes and circumstances that lead to so-called seam effects, review methods of analysis in their presence and evaluate the potential of dependent interviewing techniques to reduce measurement error in continuous histories.

7.1 MEASUREMENT ERROR IN REPEATED MEASURES

The main advantage of repeated measures over cross-sectional data is the possibility of estimating change at the individual level. The observed change in a variable is, however, a weighted combination of the change of true scores and the change of measurement errors and, depending on the relative changes of these components, may be biased upward or downward (Rodgers, 1989). A second advantage of repeated measures is their potential to correct for bias due to correlation between individual unobserved characteristics and other observed explanatory variables by differencing or fixed effects modelling. Although differencing eliminates bias due to such omitted variables, it may increase the inconsistency due to measurement error, if the explanatory variables are more strongly correlated across times than the measurement errors. If, on the other hand, data are consistently misreported across time, then differencing could increase the reliability of longitudinal data (Bound *et al,* 1994; Bound and Krueger, 1991). As Solon concludes, "serial correlation in measurement error is crucial in determining the direction and magnitude of inconsistency for panel estimators that do not account for measurement error" (Solon, 1989, p. 493). On the other hand, the reliability of

measurement error correction models relies on assumptions about the correlation structure of errors (Bound, Brown and Mathiowetz, 2001). Despite the practical importance, extremely little is known about the properties of measurement errors across waves, since most validation studies are only based on one wave of interviews (for an overview, see Bound, Brown and Mathiowetz, 2001). The few existing multi-wave validation studies found positive correlations in measurement errors for employment related variables and reporting of voting behaviour.

Bound and Krueger (1991) examined annual earnings reports from the March Current Population interviews in 1977 and 1978, matched to Social Security payroll tax records. The authors reported correlation coefficients for measurement errors in annual earnings across the two years of .40 for men and .10 for women. Rodgers, Brown and Duncan (1993) and Bound *et al* (1994) analysed data from the Panel Study of Income Dynamics Validation Studies, for which a PSID-like survey was carried out with workers from a large manufacturing company and reports were matched to the employer's records. Workers were interviewed in 1983 and again in 1987. Both papers report positive correlations in measurement errors of annual earnings, annual hours worked and annual hourly pay (derived from annual earnings and annual hours) in the range of .06 to .23. (The correlations are presumably lower than those reported by Bound and Krueger because of the longer interval between interviews.) Apart from examining correlations of measurement errors from repeated interviews, Rodgers, Brown and Duncan (1993) also examined correlations of errors in retrospective reports from one interview about each of the previous years for which record data was available. Unsurprisingly correlations of measurement errors for repeated measures from one interview were higher, ranging from .36 to .49. Finally, Bound *et al* (1994) reported that errors in measures of change in annual earnings are negatively related to true changes, as is the case for errors in levels.

Presser and Traugott (1992) examined the nature of measurement error in reporting on voting behaviour, by matching the 1972-1974-1976 Michigan Election Study to administrative records of voting turnout. They found that misreporting about voting in one survey year was correlated with misreporting about voting in an earlier survey

year: only about half of those who incorrectly reported voting in one interview gave an accurate report in another interview.

None of the above studies address the causes of correlated errors, although Presser and Traugott explain the fact that misreporters of voting typically resemble actual voters by saying that the "better educated and more [politically] interested may feel more pressure to misreport because their naïve theories about politics tell them that they are the kinds of people who vote (or, alternatively, ought to vote)" (Presser and Traugott, 1992, p. 85). This suggests that correlations in measurement errors over time could be explained by a stable association between relatively stable respondent characteristics and parameters of measurement error. The association could either be with the systematic component of the measurement error, as in the voting example above, or with the random component.

7.2 MEASUREMENT ERROR IN CONTINUOUS HISTORIES

Continuous histories derived by interviewing respondents at different points in time, often suffer from longitudinal inconsistencies: the number of status changes is typically "much greater between months for which the data are collected in different waves than between months for which the data are collected in the same wave" (Kalton, Miller and Lepkowski, 1992, p. 13). The concentration of transitions between reference periods, known as the 'seam effect', affects all panel surveys and can be substantial (see, for example, Burkhead and Coder, 1985; Hill, 1987; Kalton and Miller, 1991; Lemaitre, 1992; Moore and Kasprzyk, 1984).

Causes of seam effects

A common misconception is that the concentration of transitions at the seam is the result of misdating of changes, and that respondents have a tendency to *report* more transitions at the seam. In reality seam effects are the result of combining data from repeated panel interviews in the presence of measurement and processing errors. These underlying errors are not specific to repeated panel studies – although their visibility as seam effects is. To illustrate, picture a survey with a reference period of

m months, where in the first interview, in month *m*, the respondent is asked about his situation in months 1, 2, …, *m*. In the second interview, in month 2*m*, he is asked about his situation in months *m*+1, *m*+2, …, 2*m*. When the information from both interviews is combined to create a continuous history, one typically attributes a disproportionate number of changes in status to months *m* and *m*+1, the 'seam' between reference periods.

The analyst may *infer* a change at the seam, if the status report from the first interview for month m does not match the retrospective report from the second interview for month *m*+1. Such mismatches may occur for several reasons (see Martini, 1989; Young, 1989):

1. *omission or under-reporting*: if the respondent correctly reports receipt of an income source in all but one interview, there will be an apparent movement off the source in month *m*+1, followed by a move back in month 2*m*+1;

2. *misclassification or re-definition* of past information: for example, if respondents retrospectively report their labour market activity for *m*+1 as 'looking after family', although at time *m* they reported being unemployed;

3. *misplacement of events in time*: if respondents correctly report on their current situation in 2*m*, but then report the same situation for all previous months of the recall period until *m*+1, 'telescoping' changes back to the seam and producing 'constant wave responses' (Rips, Conrad and Fricker, 2003; Young, 1989);

4. *keying errors* on the part of the interviewer;

5. *coding errors*, especially for items coded to complex coding frames, such as industry and occupation. Spurious changes can occur because respondents use different words to describe the same occupation or industry, because of ambiguous descriptions or coding errors.

The analyst may also *impute* a change at the seam, if the date of a status change is missing. Imputed dates are often set to half-way between previous and subsequent events. If the imputed date falls into the previous reference period, the date is by

default set to the start of the reference period (i.e. the seam), instead of over-riding information from the earlier interview. The reasoning is that information from the earlier interview is closer to the actual events and therefore considered more reliable (Halpin, 1998).

Analysis in the presence of seam effects

The seam effect can lead to bias in estimates of gross flows (Martini, 1989) and of spell durations (Boudreau, 2003). In the SIPP, for example, the length of many unemployment spells corresponds exactly to the equivalent of one or two reference periods (Chetty, 2004). The literature suggests different ways of dealing with the seam effect.

At the micro-level, different authors have suggested editing rules to reduce the appearance of seam effects, for example, by:

1. *imputing transition dates as random dates between interviews*. According to Halpin (1998), this practice has good properties for modelling durations. However, this approach assumes that the transitions at the seam are real but misplaced in time (for example constant wave responses) and does not deal with spurious transitions caused by omissions or mis-classifications;

2. *relaxing the rules for matching spells around the seam*. Strict criteria for matching employment spells at the seam may, for example, be based on the correspondence of industry and occupation codes, while relaxed criteria may consider spell type (full-time, part-time or self-employed) sufficient for matching (Halpin, 1998; Paull, 2002);

3. *editing reports of income sources to correct mis-classifications*, by using additional information about eligibility criteria (Marquis, Moore and Huggins, 1990).

At the macro-level the most prevalent correction strategies include:

4. *acknowledging seam problems* in a footnote, but ignoring them in the analysis;

5. *including a dummy variable for the seam month* in duration analysis (see, for example Brandon, 1995; Chetty, 2004; Halpin, 1998; Hill, 1994). The parameter estimates for the seam dummy are, however, not always shown or their effects discussed. Where the effects are discussed, the findings are conflicting. The seam dummy may significantly alter the effect of other explanatory variables, but this is not always the case. Long remarked that the seam dummy is only a rough correction because "it will not capture any existing correlation between the response errors that result in the bias toward the seam and the outcome variable or the other explanatory variables in the model" (Long, 1990, p. 54).

6. *using only the most recent month of each reference period*. Grogger, Haider and Klerman (2003), for example, use the fourth month of each wave from the SIPP to calculate tri-annual entry and exit rates from benefits and smooth these to obtain monthly estimates. Ham and Shore-Sheppard (2005) use the same approach to estimate duration models. They comment that although "this approach has the disadvantage that information on the timing of transitions reported to occur between months other than at the seam is lost, the advantage is that the data in the fourth month of each wave are the most likely to be accurate since it is closest to the time of interview" (Ham and Shore-Sheppard, 2005, p. 12).

Less common strategies for corrections at the macro-level include:

7. *weighting to correct for mis-placement of events in time*. Miller, Lepkowski and Kalton (1992) estimate duration models by assigning weights, so that the observed number of starts/ends in each month matches the expected number, given a uniform distribution. This approach assumes that within-wave transitions are correct, but some proportion of the seam transitions should have been reported in a different month. Similar to random imputation of start

dates at the micro-level (1. above), this approach does not account for seam effects caused by spurious transitions due to omissions or mis-classifications;

8. *latent class analysis.* Bassi *et al* (2000) use latent class models to correct for seam effects in labour market status information from the SIPP;

9. *estimating error probabilities.* Romeo (2001) defines consistent sequences of labour market activity reports based on the questionnaire and estimates the probability of correct and erroneous transitions at the seam. The error probabilities are then used to correct for seam errors in duration models.

Data collection methods to prevent seam effects: dependent interviewing (DI)

With the development of computer assisted interviewing technologies, it has become possible to feed forward substantive answers from previous interviews, which can be used to improve data quality (Brown, Hale and Michaud, 1998). Previous answers can be used to formulate questions or to determine routing (proactive DI) or to prompt post-response edit checks (reactive DI). (For a review of applications of dependent interviewing, see Mathiowetz and McGonagle, 2000. See also section 9.1 of this paper) This differs from traditional independent interviewing, where respondents are typically asked the same questions about their situation at different points in time, without reference to previous answers.

Dependent interviewing can reduce the occurrence of non-corresponding status reports at the seam by using substantive information from the previous interview

1. to verify apparent changes through post-response edit checks if the answer is inconsistent with the previous answer (this is mainly to catch keying errors, but also includes reminding respondents of items mentioned in the previous but not the current interview to reduce under-reporting);

2. to remind respondents of previous reports. Providing a memory aid and temporal boundary is likely to improve respondent recall and thereby reduce omissions and misclassifications;

47

3. to ask the respondent whether his situation is still the same, instead of inferring change from different status reports;

4. to route around follow-up questions if circumstances have not changed, in which case previous data can be brought forward. For coded answers to open-ended questions, such as industry and occupation, this reduces coding variability.

Dependent interviewing has been shown to reduce errors in the reporting of income and employment related information. Compared to administrative data, DI substantially reduces under-reporting of *income sources*, especially for sources which are relatively common or easy to forget (Dibbs *et al,* 1995; Lynn *et al,* 2006). In a study reported by Lynn *et al* (2004), for example, 23% of respondents who received child benefit according to administrative records forgot to report this in the survey. Under-reporting was reduced to 4% with proactive DI and to 8% with reactive DI. Income reporting also suffers from constant wave responses: income receipt only in the early part of the recall period is associated with an increased risk of under-reporting (Goudreau, Oberheu and Vaughan, 1984; Lynn *et al,* 2004).

Dependent interviewing also reduces spurious change in *employment characteristics* across waves. If respondents are asked to report on the characteristics of their employment every year, without reference to previous information, the answers imply an implausible amount of change. Sala and Lynn (2004) reported that among respondents who had not changed their job, 24% appeared to be in a different occupation, 15% in a different industry (even according to the coarsest measures of the Standard Industry and Occupation Classifications) and 16% reported having managerial duties they did not report in the previous year, or vice versa. These spurious changes were significantly reduced with DI. The characteristics associated with reductions in measurement error were associated with an increased propensity to have complex jobs, the characteristics of which are difficult to describe, and therefore more likely to be reported inconsistently across interviews. These findings were supported by Hill (1994) who reported that the majority of occupational and industrial changes observed with independent methods in the Survey of Income and Program Participation (SIPP) were not associated with changes in working hours,

wages or employers. These associations were, however, much stronger with proactive DI, leading him to conclude that "most of the observed 'change' with independent data collection methods is a result of variability in the response/coding process". Similarly, the introduction of proactive dependent questions in the Current Population Survey greatly reduced apparent change in industry and occupation (which were assumed to have been largely spurious) and also addressed respondent complaints about repetitiveness (Norwood and Tanur, 1994; Polivka and Rothgeb, 1993).

Dependent interviewing can also reduce spurious changes in *labour market activity*. Murray *et al* (1991) examined the effects of proactive DI on estimates of transitions out of employment in the Canadian Labour Market Activity Survey. Comparing monthly transition rates with aggregate administrative data they concluded that DI, and the corresponding editing rules, successfully removed the seam effect. Jäckle and Lynn (2004) reported that 50% of observed transitions with independent interviewing occurred at the seam, compared to 20% with proactive DI. This was contrary to expectations that proactive DI would eliminate seam effects (a uniform distribution of transitions over months would yield about 5% of transitions at the seam). DI worked by precluding non-corresponding reports at the seam, but missing transition dates meant that seam effects were not eliminated.

A common concern regarding DI is that respondents may not make the cognitive effort required to answer questions well and may instead simply confirm previous information. So far there is, however, no evidence that DI leads to over-reporting of income sources or to under-reporting of changes which such 'satisficing' behaviour would imply (Hale and Michaud, 1995; Jäckle and Lynn, 2004).

Limitations of dependent interviewing

Although dependent interviewing reduces errors that lead to seam effects, it does not eliminate them:

1. DI does not improve the reporting of transition dates. As long as respondents fail to report dates, seam effects due to imputations are not eliminated (Jäckle and Lynn, 2004).

2. Constant wave reporting remains a problem. Although DI improves the reporting of whether an income source was received (at any time) during the references period, it does not improve the dating of changes in income receipt. In the study by Lynn *et al* (2004), 80-95% of respondents reported receipt for all months in the (on average 18 month) reference period. Marquis *et al* suggested for the SIPP that "many respondents adopt a simple heuristic or rule of thumb to quickly answer questions about recurring events in the four-month reference period (such as monthly income sources and amounts). Respondents use the simple rule as a substitute for detailed, direct recall and as a substitute for checking their personal records. […] Instead of being caused by memory decay, forgetting, telescoping, deliberate lying, and the like, the measurement errors may be caused by trying to reconstruct a complex past using too simple a rule" (Marquis, Moore and Huggins, 1990, p. 568). If this hypothesis is true, the authors argue that measures to reduce recall difficulties, for example by shortening reference periods, providing memory cues or dependent interviewing, will not be successful at preventing seam effects. Instead they suggest informing respondents of the detailed data requirements, teaching them how to fulfil these (for example by keeping financial records) and training interviewers to detect inappropriate response strategies and to steer respondents towards better strategies.

3. For "check all" type questions, some under-reporting remains, because DI can only have an impact on respondents who are actually asked the DI question, that is, those respondents who reported an item in the previous wave. Given that the propensity to under-report is likely to be associated with some fixed characteristics of the survey respondent, those who under-reported at the current wave could be expected to have an increased propensity to have under-reported also at the previous wave (Lynn *et al,* 2004). In this spirit, Marquis and Moore (1989) suggested that dependent interviewing might increase the error covariance.

For industry and occupation, it is debatable whether dependent interviewing actually improves the accuracy with which codes are assigned, although it is clear that it improves the consistency in coding over time within a respondent. In other words, it may not reduce measurement error in estimates of, say, the *distribution* of occupations, but it does reduce measurement error in estimates of levels of *change* in occupation, due to the consistency of errors over time. Finally, DI does not work well for items that are unstable in the respondent's perception.

7.3 CONCLUSION

There is clearly a need for more information about the nature of measurement error in repeated measures, both to gauge the direction and magnitude of bias introduced into models that do not account for measurement error and to develop reliable models that do. Regarding measurement error in continuous histories, there is a need to further develop dependent interviewing techniques. Lynn *et al* (2004), for example, propose extensions to reduce the problem of correlated under-reporting, by using information from multiple previous interviews or by filtering DI questions on predictors of income receipt. Many questions remain open as to how the reporting of dates can be improved, both to reduce misdating and resulting constant wave responses and to reduce item non-response. At the same time, there is a need to develop analysis methods in the presence of seam effects.

8. Panel Conditioning

8.1 WHAT IS PANEL CONDITIONING?

Panel conditioning is a potential source of non-sampling error in panel surveys. Panel conditioning refers to a change in response that occurs because the respondent has had one or more prior interviews. In any survey whose primary purpose is to measure change over time, distinguishing true change (or stability) from reported changes induced by repeated interviews is clearly important. Possible conditioning effects are of particular concern to a panel design where the length of time between interview points may be relatively short (weeks or months). It is generally thought to be less of an issue in longitudinal designs where the length of time between interviews is typically much longer (years) even though no evidence is available to either confirm or refute this.

The literature describes several ways in which conditioning effects may come about and in general, conditioning is seen as having negative rather than positive effects on data quality.

- The respondent may change the way they answer questions because of prior interviews. For example they may be more inclined to give a similar response as at a previous interview. Conversely, they may vary their response in an effort to provide more interesting information or alter their response in some other way.

- The respondent may change their behaviour in response to a previous interview. A prior interview may alert the respondent to some element of their own behaviour or circumstances that prompts them to act in a way they would not have acted otherwise.

- The response given to an interviewer may be due to changes in the interviewer's behaviour at subsequent interviews. This is most likely where the same interviewer returns to interview the same respondents at successive points. The interviewer, on becoming familiar with the respondent's circumstances may make

unwarranted assumptions which lead to questions not being asked fully and correctly. The respondent may also behave differently towards the interviewer whom they have grown to know over a period of time.

While these types of effects are potentially harmful to data quality, some conditioning effects may be positive. Levels of item non-response to sensitive items may reduce over time. The respondent learns that providing such details does not result in any direct problems for them and as they come to trust their interviewer, they may be more forthcoming with details of income or other personal information. Respondents may also learn the rules of the interview and provide only information which is relevant and more accurate at later rounds.

8.2 THE EFFECTS OF PANEL CONDITIONING

Possible conditioning effects have been reported in a number of studies. Bailar (1972), Cohen and Burt (1985) and Corder and Horvitz (1989) analyse a US survey of health care utilisation and have reported the tendency of respondents to gradually decrease the number of positive responses to questions over time, resulting in a decline in incidence and prevalence measures on particular items. Cohen and Burt refer to this as the 'data collection frequency' effect. In a validation study comparing survey reports with administrative medical records they found that individuals with four rounds of data collection had a significantly higher level of agreement with administrative records than those with five rounds of data collection. Neter and Waksberg (1964) also found that house owners reported fewer jobs and less expenditure on housing repairs and alterations on the second and third time in a panel. While these differences were apparent in the data it is not clear to what extent they were due to recall error or to the confounding effects of differential panel attrition. Possible changes in respondent behaviour have been observed in another validation study by Traugott and Katosh (1979) who reported increased levels of voting following a pre-election survey. They hypothesised that the first interview had highlighted the importance of the election for respondents who then altered their behaviour and voted. Ferber (1964) found that the quality of savings data reported increased over time but concluded that differential attrition, where poorer members

53

of the panel dropped out, was partly the cause combined with an increase in accuracy for those that remained in the panel. Silberstein and Jacobs (1989) in a study of the US Consumer Expenditure Survey identified an under-reporting of certain types of expenditures. They concluded that conditioning effects due to the length of time in the sample were moderate or minimal and that the main effects were due to recall error.

With all of these studies it is clear that panel attrition and recall error may have a more significant effect on the overall distributions than panel conditioning. Waterton and Lievesley (1989) attempted to go some way towards disentangling the effects of attrition and conditioning using a split-panel design on the British Social Attitudes Panel. This is a survey of attitudes asking questions about a range of social and political issues. The aim was to try and separate out real change from changes in reporting due to previous interviews. Their analysis found that there was some suggestion that respondents became politicized by the interview, that they reported more honestly, and that they became less likely to answer 'don't know' when compared to the control sample. Waterton and Lievesley concluded that conditioning was not a major problem and there were greater dangers for representativeness in the effects of differential attrition from the sample between interviews. These findings were, in their own words, tentative and it was not clear to what extent they would be generalisable to other surveys with differing content and design. The length of time between interviews is also likely to be a factor. In this case the interview gap was one year but surveys repeated at closer intervals may find greater conditioning effects. Cantor (1989) also points out that interview spacing may affect conditioning even though there is little evidence to either support or deny this. Shorter gaps between interviews may increase respondent burden and lead to higher rates of attrition, once again confounding the effects of conditioning.

Kalton *et al* (1989) discuss the use of rotating panel designs in which fresh replicate samples are added to the panel at each round as a means to examine panel conditioning through having a comparison group. While in theory this method should enable the effects to be observed, it relies on holding all other survey conditions constant, something which in practice is very difficult to do. Various US studies have

compared the results from different rotation groups and identified what they term as rotation group bias. This bias is simply the difference between estimates for the incoming group when compared to previously interviewed panel members (Bailar, 1975, 1979; McCarthy, 1980; U.S.Bureau of the Census, 1978). As Kalton *et al* (1989) point out, rotation group bias has been demonstrated as a pervasive effect in rotating panel designs. Despite this, the cause and effect is not known but may include differential attrition effects as well as response effects due to repeated interviews.

8.3 CONCLUSION

One of the most interesting facets of the panel conditioning debate is the lack of evidence showing any clear effects. The majority of work has been done in the US and there is little recent research in this area. O'Muircheartaigh (1989) in a discussion of sources of non-sampling error reviews the evidence on panel conditioning. He takes the view that in order to disentangle the effects of interviewer conditioning, respondent conditioning, and interviewer-respondent interactions, a controlled experiment in which different combinations of these are allocated randomly within the available sample would be required. While conditioning continues to be an issue of concern for longitudinal panel surveys, the expense of carrying out such methodological work using experimental designs to attempt to measure the possible effects would be relatively high. It remains open to question whether such investigations are warranted.

9. Interviewing Strategies

9.1 DEPENDENT INTERVIEWING

The term dependent interviewing (DI) refers to the use of prior respondent-specific information in the course of a survey interview to determine the questions asked, the question wording, in-interview edit checks and so on. The prior information typically consists of answers provided in a previous interview – a situation which is therefore unique to panel surveys – but could also be from the sampling frame or other source. The development of Computer Assisted Interviewing (CAI) in recent decades has dramatically increased the scope for sophisticated use of DI techniques.

Reasons cited for using DI mainly fall into two categories. First, it can save interview time, thereby saving costs and/or making it possible to ask extra questions. Second, it can improve the quality of the survey data. However, there are many ways in which DI can affect survey data and not all are necessarily beneficial.

Variants of DI

There are two main types of DI – proactive and reactive. The former involves explicitly incorporating the prior information into the questioning, which typically becomes confirmatory in nature, for example, "When we interviewed you last, you were receiving Income Support. Is that still the case?" Reactive DI involves asking an independent question (e.g. "Are you currently receiving any income or payments from any of these sources?") and then having an automatic check determine whether one or more follow-up questions are needed. For example, if Income Support is not mentioned, "Can I just check, when we interviewed you last you were receiving Income Support. Are you still receiving it now?"

Some surveys and survey organisations seem to have adopted conventions about when to use proactive DI and when to use reactive DI. These seem to be based largely on prior expectations or general theory regarding respondent behaviour. There is also some distinction between new and long-standing surveys. The latter

tend to be concerned with the preservation of time series (keeping systematic errors as constant as possible) while the former are more concerned with maximising data quality (minimising systematic errors).

There has only been one experimental comparison of proactive and reactive DI, resulting in a recent series of papers (Lynn *et al*, 2006; Lynn *et al*, 2004; Sala and Lynn, 2004; Jäckle and Lynn, 2004; Jäckle, 2005). Previous studies had compared either proactive DI to independent interviewing (Cantor, 1991; Hill, 1994) or reactive DI to independent interviewing (Dibbs *et al*, 1995).

Response accuracy and seam effects

A prime concern of researchers has been to reduce errors in the estimation of change over time caused by spurious measures of change between interviews. If each interview attempts to collect a continuous history over the period since the previous interview, then such errors will appear as 'seam effects' (see section 7.2 above). If a single measure is obtained at each interview, equivalent effects will result, but they may not be so obviously apparent as *every* transition in the data will be at a seam (see section 7.1).

The modest volume of research into the effects of DI on seam effects (Cantor 1991, Doyle *et al* 2000, Hale and Michaud 1995, Hiemestra *et al* 1993, Hill 1994, Webber 1994, Jäckle and Lynn, 2004) has mainly concentrated on the level of reduction in estimates of change, without explicitly estimating the accuracy of the estimates. This is based on an assumption that less change corresponds with better data; in other words that all or most of the errors in independently collected data are in the direction of spurious change. Brown *et al* (1998) suggest that this assumption requires validation.

Regarding estimates of change with repeated measures data, Cantor (1991), Hill (1994) and Sala & Lynn (2004) found, on the CPS, SIPP and UK-ECHP respectively, that proactive DI for industry and occupation questions dramatically reduced the over-estimation of wave-on-wave change, though Hill also found that DI introduced a

failure to identify some real change. Reactive DI offers the opportunity to reduce both forms of error. The findings of Sala & Lynn (2004) are consistent with this possibility, though direct evidence of measurement error is not presented.

Reactive DI has been compared with independent interviewing for the case of household income sources on by Dibbs *et al* (1995) and Lynn *et al* (2006), with the conclusion that it greatly reduced omissions. The latter study also made a comparison made with proactive DI and concluded that the reactive variant was almost as successful at reducing omissions.

Phenomena relevant to measurement error in the context of DI include:

- acquiescence bias (which tends to cause under-reporting of change and could be accentuated by DI);

- memory effects (which can cause either over-reporting or under-reporting depending on the context, but could potentially be reduced by DI);

- social desirability bias (which tends to cause under-reporting of sensitive information and could be either reduced or heightened by DI, depending on the context);

- coding error (which will tend to cause over-estimation of change and may also introduce systematic errors into other sorts of estimates, depending on the context. DI has the potential to reduce the impact of coding error by reducing required levels of dual-coding).

The net effect of DI on measurement error in any particular context will result from the interaction of these phenomena. Though some of the DI studies investigate net effects in particular contexts, none have attempted to separate out any of the component phenomena (Mathiowetz and McGonagle 2000). This makes it difficult to generalise to survey items or situations other than those specifically investigated.

Interview time

A number of surveys cite the saving of interviewer time and/or reduction of respondent burden as a realised advantage of their DI implementation (e.g. CPS, SIPP, ELSA), though the only direct evidence of this appears in Jäckle (2005) (aside from some very limited information in Weinberg, 2002). (Dibbs *et al* (1995) and Pascale and Meyer (2004) both report a favourable reaction to DI from respondents, who expect the interviewer to have information from the previous interview.) It is clear that the potential time savings are greater if the researchers are prepared to make greater assumptions. For example, in the UK Labour Force Survey, respondents are asked if they are still in the same job as at the time of the previous interview (3 months earlier). If they say "yes," further details about the job (industry, hours, employment status, etc) are not collected but are assumed to be the same as reported in the previous interview (Knight 2001). This saves a lot of interview time, but may result in a failure to identify some changes of details (Stanley and Safer 1997). In general, the impacts of the assumptions necessary to save interview time have not been assessed.

Types of questions

The limited literature existing on DI has addressed specific survey items, mostly in the areas of occupation and income. The only attempt to try to generalise some of the issues is Mathiowetz and McGonagle (2000), but this is severely hampered by a lack of evidence. It seems likely that there are many aspects of DI that can be generalised, though some will be question- or topic-specific. There is a need to identify the generalisable aspects and their nature. In order to do this, there is also a need to experiment with DI in other subject areas. Candidates would include questions on high-frequency and low-frequency behaviours, attitudes, skills and qualifications, health, consumption, expenditure and crime victimisation.

Conclusions on DI

Dependent interviewing is an important area in which both developed theory and empirical evidence to provide guidance to survey researchers are limited. A number of areas in which research is needed have been identified in this scoping study.

9.2 INTERVIEWER STABILITY

Response rates

Conventional wisdom holds that face-to-face panel surveys achieve better response rates when the same interviewer returns at successive waves. Theory suggests some reasons why that might be the case (Groves *et al* 1992, Morton-Williams 1993), but there is very little empirical evidence. Most studies that purport to demonstrate an effect of this sort are non-experimental (Waterton and Lievesley 1987, Rendtel 1990, Rope 1993, Taylor *et al* 1996) and, in consequence, confound interviewer stability with area effects.

Three related studies provide an exception to this, all being based upon a design that interpenetrated interviewers and areas on the BHPS. Campanelli and O'Muircheartaigh (1999) found no effect of continuity at wave 2. Laurie *et al* (1999) extended the analysis to waves 3 and 4 and found significant differences. Campanelli and O'Muircheartaigh (2002) re-analysed the same data and concluded that the apparent differences could be accounted for by non-random interviewer attrition.

In conclusion, there remains little or no evidence that interviewer stability affects response rates. However, no studies have attempted to examine specific sample subgroups or survey circumstances in which the theory would suggest that effects might occur. It may simply be the case that these subgroups or circumstances are sufficiently rare that their impact cannot be detected at the whole-sample level.

Some further research would appear to be warranted, as this is an important practical issue with potentially serious cost implications for some surveys.

Data quality

There is considerable conjecturing in the literature about the likely impacts of interviewer stability on responses. Hypotheses include the following:

- Having the same interviewer will improve respondent-interviewer rapport and trust, which should reduce the negative impacts of confidentiality and sensitivity concerns;

- Conversely, improved rapport could exacerbate social desirability biases by reducing the social distance between respondent and interviewer;

- Repeat visits by the same interviewer should shift the nature of the respondent commitment from general to personal, with an attendant likely reduction in satisficing and therefore measurement error;

- Interviewer stability could result in an increase in correlated interviewer error due to an increased tendency for interviewers to make assumptions rather than ask all questions fully and neutrally.

However, we have been unable to locate any empirical studies that address any of these hypotheses.

It may also be noted that some of these effects might contribute to observed "panel conditioning" effects (which are usually ascribed entirely to the respondents rather than the interviewers), as they change in strength over waves.

The BHPS interpenetrated design would appear to offer a good opportunity to study some of these issues.

10. Sampling

There are a number of aspects of sample design that are rather distinctive in the case of longitudinal surveys (Kalton and Citro, 1993). These are primarily to do with ways in which the sample can change in nature over time, for example in terms of geographical location and other non-permanent characteristics. There are also issues relating to the analytical use of longitudinal data, such as modelling techniques that are appropriate for data with a particular sample structure. Research investigating specifically these aspects is limited.

10.1 SAMPLE CLUSTERING AND DESIGN EFFECTS

There are a number of reasons why sample design considerations may be very different for a longitudinal survey compared with a cross-sectional survey of the same population with similar survey instruments. First, the impact of sample clustering on data collection costs will be less, in relative terms, for a longitudinal survey collecting data over a long period of time, due to sample dispersal. In other words, any cost savings at wave 1 from initially selecting a clustered sample will lessen at each successive wave as the sample units move location causing de-clustering of the sample. The extent of this de-clustering and its impact on costs will depend on the study population, the nature of the clustering, and the methods of data collection employed. To make informed decisions about the optimal initial sample design, researchers require information about the timing of all future waves of data collection and the modes of data collection to be used, plus estimates of the degree of mobility likely to be experienced by the sample members during the lifetime of the survey. Such estimates are rarely available, though in principle existing survey data could be used to shed light on these issues. We are not aware of any research that has attempted to quantify the nature of mobility of particular populations over periods of time such that the extent of sample de-clustering could be estimated.

In the same way that cost benefits from sample clustering are likely to diminish over waves so, possibly, are the detrimental design effects due to clustering. Again,

however, this is an issue upon which we have not been able to locate published studies. It would be a relatively simple matter to estimate design effects using existent longitudinal survey data.

10.2 SAMPLE COVERAGE

With longitudinal surveys - much more so than with other surveys - issues of sample design tend to be intimately bound up with issues of definition of the study population (Elliot *et al*, 2007). Longitudinal populations require definition in time as well as the other usual dimensions. An initial sample design in combination with a set of follow-up rules defines the longitudinal population represented by the continuing sample. Some surveys may suffer from a lack of explicit consideration of the longitudinal population of interest and in consequence may use sub-optimal sample designs. In this context, the sample design can be thought of as including the sample follow-up rules and any procedures for adding new sample members at waves subsequent to wave 1 - in other words, the re-design of the sample at each wave.

Longitudinal survey designs are typically not good at capturing the effects of either internal migration processes (due to attrition related to geographical mobility) or external migration and immigration (due to limitations of sample design and following rules). Researchers should have better information about how much this matters and what might be done about it in typical longitudinal survey situations.

10.3 SELECTION PROBABILITIES AND ASSOCIATED WEIGHTING

For non-monotone sample designs, processes of change in unit composition over time (where units may be households, businesses, schools, etc) generate considerable complexity and uncertainty over eligibility and inclusion probabilities. The only method proposed to deal with this situation, to our knowledge, is the "fair shares" method of weighting (Ernst 1989, Lavallée 1995). Limited research has been carried out into the properties of this method in different situations, and possible alternatives. Again, researchers concerned with longitudinal surveys of this sort should have a better understanding of the implications.

10.4 SAMPLE STRUCTURE

For many surveys, a substantive case can be made for over-sampling certain population subgroups. For longitudinal surveys, the implications may be less obvious than for cross-sectional surveys, especially when the potential over-sampling criteria have an element of time-dependency (e.g. Lynn & Plewis, 2003). There is a lack of guidance available on these issues.

It can also be argued that analytical methods require a simple random sample (Davies and Crouchley, 1992), though this argument has had opponents (Hedges, 1992; Rose *et al*, 1991). On the other hand, the use of multi-level modelling approaches requires an explicit multi-level structure in the sample, which can be achieved by a clustered sample. Other than the references cited here, there seems to be little explicit discussion in the literature of the relative merits of alternative sample structures. In particular, survey documentation for major longitudinal surveys rarely provides detailed justification for the design chosen.

Given that longitudinal surveys sample in time as well as space, decisions about intervals between data collection waves and the trade-offs to be made with sample size and the overall length (life) of a longitudinal survey can also be considered as a sample structure issue. Citro and Kalton (1993, chapter 4), in the context of the Survey of Income and Program Participation, discuss alternative survey designs in terms of frequency of selecting a new sample (panel), intervals between waves, and numbers of waves. They identify a number of the key considerations that should drive design choice.

REFERENCES

Allen, K., Simpson, G., Winn, D. and Giovino, G. (1991) Response rates and characteristics of adolescents in the 1989 Teenage Attitudes and Practices Survey, in *1991 Proceedings of Survey Research Methods Section of the American Statistical Association*, 575-580. http://www.amstat.org/sections/srms/proceedings

Apodaca, R., Lea, S. and Edwards, B. (1998) The effect of longitudinal burden on survey participation, *1998 Proceedings of Survey Research Methods Section of the American Statistical Association,* 906-910. http://www.amstat.org/sections/srms/proceedings

Baddeley, A.D., Lewis, V., Nimmo-Smith, I. (1978) When did you last…?, in Gruneberg, M.M., Morris, P.E., Sykes, R.N. (eds.), *Practical Aspects of Memory*, New York: Academic Press, 77-83.

Bailar, B.A (1975) The effects of rotation group bias in estimates from panel surveys, *Journal of the American Statistical Association,* **70**, 23 – 30.

Baldinger, A., O'Neill, H. W. Wilson, W. Adams, C. and Borvers, D. (1995) The CMOR Respondent Co-operation Study, paper presented at the 1995 Market Research Association Conference.

Bassi, F., Hagenaars, J. A., Croon, M. A. and Vermunt, J. K. (2000) Estimating True Changes When Categorical Panel Data are Affected by Uncorrelated and Correlated Classification Errors, *Sociological Methods & Research,* **29(2)**, 230-268.

Battaglia, M., Zell, E.R., Ching, P.L.Y.H. (1996) Can participating in a panel sample introduce bias into trend estimates? *1996 Proceedings of Survey Research Methods Section of the American Statistical Association*, 1010-1013. http://www.amstat.org/sections/srms/proceedings

Baumgartner, R. and Rathbun, P. (1997) Prepaid monetary incentives and mail survey response rates, paper presented at the Annual Conference of the American Association of Public Opinion Research, Norfolk, Virginia, 1997.

Beatty, P. and Herrmann, D. (2002) To answer or not to answer: decision processes related to survey item nonresponse, 71-85 in Groves, R.M., Dillman, D.A., Eltinge, J.L. & Little, R.J.A. (ed.s) *Survey Nonresponse,* John Wiley & Sons, New York.

Behr, A., Bellgardt, E. and Rendtel, U. (2004) The effect of panel attrition on measures of income inequality and change in the European Community Household Panel User Data Base (ECHP UDB), 220-234 in Ehling, M, Rendtel, U. et al (ed.s) *Harmonisation of Panel Surveys and Data Quality*, Wiesbaden, Germany: Statistisches Bundesamt.

Belli, R.F. (1998) The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys, *Memory* **6(4)**, 383-406.

Belli, R.F., Shay, W.L. and Stafford, F.P. (2001) Event History Calendars and Question List Surveys, *Public Opinion Quarterly*, **65(1)**, 45-74.

Belli, R. F., Lee, E. H., Stafford, F. P. and Chou, C-H. (2004). Calendar and question-list survey methods: Association between interviewer behaviors and data quality. *Journal of Official Statistics*, **20(2)**, 185-218. http://www.jos.nu/Articles/abstract.asp?article=202185

Bennett, D.J. and Steel, D. (2000) An evaluation of a large-scale CATI household survey using random digit dialling, *Australian and New Zealand Journal of Statistics* **42(3)**, 255-270. http://www.ingentaconnect.com/content/bpl/anzs/2000/00000042/00000003/art00126

Berk, M.L., Mathiowetz, N.A., Ward, E.P. and White, A.A. (1987) The effect of prepaid and promised incentives: results of a controlled experiment, in *Journal of Official Statistics* **3(4)**, 449-457. http://www.jos.nu/Articles/abstract.asp?article=34449

Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., and Yamamoto, K. (1992) An Experiment in Monetary Incentives, *1992 Proceedings of the Survey Research Methods Section of the American Statistical Association*, 292-298. http://www.amstat.org/sections/srms/proceedings

Blair, E. and Burton, S. (1987) Cognitive processes used by survey respondents to answer behavioural frequency questions, *Journal of Consumer Research*, **14**, 280-288.

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

American Statistical Association, 1020-1025.
http://www.amstat.org/sections/srms/proceedings

Bollinger, C.R., and David, M.H. (2001) Estimation with response error and nonresponse: food-stamp participation in the SIPP, *Journal of Business & Economic Studies*, **19(2)**, 129-141.

Boudreau, C.(2003) *Duration Data Analysis in Longitudinal Surveys.* University of Waterloo: Department of Statistics and Actuarial Science.

Bound, J., and Krueger, A.B. (1991) The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? *Journal of Labor Economics* **9(1)**, 1-24.

Bound, J., Brown, C., Duncan, G.J., and Rodgers, W.L. (1994) Evidence on the Validity of Cross-sectional and Longitudinal Market Data, *Journal of Labor Economics* **12(3)**, 345-368.

Bound, J., Brown, C. and Mathiowetz, N. (2001) Measurement Error in Survey Data, 3705-3843 in Heckman, J.J., Leamer, E. (eds.), *Handbook of Econometrics*, Vol. 5, North Holland: Elsevier Science.

Bradburn, N.M. and Sudman, S. (1979) *Improving Interviewing Methods and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*, San Francisco: Jossey-Bass.

Bradburn, N.M., Sudman, S., Blair, E., and Stocking, C. (1978) Question Threat and Response Bias, *Public Opinion Quarterly* **42(2)**, 221-234.

Bradburn, N.M., Huttenlocher, J., and Hedges, L. (1994) Telescoping and Temporal Memory, ch. 13 in Schwarz, N., Sudman, S. (eds.) *Autobiographical Memory and the Validity of Retrospective Reports*, New York: Springer Verlag.

Branden, L., Gritz, R.M., and Pergamit, M.R. (1995) The Effect of Interview Length on Nonresponse in the National Longitudinal Survey of Youth, *Proceedings of the 1995 Census Bureau Annual Research Conference*, 129-154.

Brandon, P. D. (1995) Vulnerability to Future Dependence among Former AFDC Mothers, *Institute for Research on Poverty Discussion Paper, No.* 1055-95, Madison: University of Wisconsin.

Brehm, J. (1993) *The Phantom Respondents - Opinion Surveys and Political Representation,* Ann Arbor: University of Michigan Press.

Brown, P. R., and G. F. Bishop (1982) Who Refuses and Resists in Telephone Surveys? Some New Evidence, University of Cincinnati. Quoted in Groves (1989).

Brown, A., Hale, A. and Michaud, S. (1998) Use of Computer Assisted Interviewing in Longitudinal Surveys, 185-200 in M. P. Couper*, et al* (eds) *Computer Assisted Survey Information Collection*, New York: John Wiley and Sons.

Burgess, R. D. (1989) Major Issues and Implications of Tracing Survey Respondents, 52-74 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Burkhead , D. and Coder, J. (1985) Gross Changes in Income Recipiency from the Survey of Income and Program Participation, *1985 Proceedings of the Social Statistics Section of the American Statistical Association*, 351-356.

Burton, J., Laurie, H., and Moon, N. (1999) "Don't ask me nothin' about nothin', I might just tell you the truth" The Interaction between unit non-response and item non-response", paper presented at the International Conference on Survey Response, October 1999, Portland, Oregon.

Burton, S. and Blair, E. (1991) Task conditions, response formulation processes, and response accuracy for behavioural frequency questions in surveys, *Public Opinion Quarterly* **55(1)**, 50-79.

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: National Centre for Social Research.

Campanelli, P. and O'Muircheartaigh, C. (1999) Interviewers, Interviewer Continuity, and Panel Survey Nonresponse, *Quality & Quantity* **33(1)**, 59-76.

Campanelli, P. and O'Muircheartaigh, C. (2002) The Importance of Experimental Control in Testing the Impact of Interviewer Continuity on Panel Survey Nonresponse, *Quality & Quantity* **36(2)**, 129-144.

Cannell, C.F. and Henson, R. (1974) Incentives, Motives, and Response Bias, *Annals of Economic and Social Measurement* **3(2)**, 307-317.

Cannell, C.F., Fisher, G. and Bakker, T. (1965) Reporting of hospitalizations in the health interview survey, *Vital and Health Statistics*, Series 2, Number 6, Washington, DC: Public Health Service.

Cannell, C.F., Groves, R.M., Magilavy, L., Mathiowetz, N.A., and Miller, P.V. (1987) An Experimental Comparison of Telephone and Personal Health Surveys, *Technical Series 2*, Paper No. 106. National Center for Health Statistics.

Cannell, C.F., Miller, P.V., and Olsenberg, L. (1981) Research on Interviewing Techniques, *Sociological Methodology* **13**, 389-437.

Cantor, D. (1989) Substantive Implications of Longitudinal Design Features: The National Crime Survey as a Case Study, 25-51 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Carpenter, J.R., Kenward, M.G. and Vansteelandt, S. (2006) A comparison of multiple imputation and doubly-robust estimation for analyses with missing data, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, **169** (forthcoming).

Carton, A., Beerten, R. and Billiet, J. (1996) The Evaluation and Treatment of Nonresponse in Face-to-Face Interviews, paper presented at the 4th International Conference on Social Science Methodology, Essex, 1-5th July 1996.

Chase, D.R. and Harada, M. (1984) Response error in self-reported recreation participation, *Journal of Leisure Research* **16**, 322-329.

Cheesbrough, S. (1993) Characteristics of Non-Responding Households in the Family Expenditure Survey, Survey Methodology Bulletin **33**, 12-18.

Chetty, R. (2004) Optimal Unemployment Insurance When Income Effects are Large, *NBER Working Paper* W10500, Cambridge, Mass: National Bureau of Economic Research. http://www.nber.org/papers/W10500

Church, A.H. (1993) Estimating the Effects of Incentives on Mail Survey Response Rates: A Meta-Analysis, *Public Opinion Quarterly* **57(1)**, 62-79.

Cialdini, R. B., Braver, S.L. and Wolf, W.S. (1993) Predictors of Nonresponse in Government and Commercial Surveys, paper presented at the 4th International Workshop on Household Non-Response, Bath UK.

Citro, C.F. and Kalton, G. (ed.s) (1993) *The Future of the Survey of Income and Program Participation*, Washington D.C.: National Academy Press.

Cohen, S.B. and Burt, V.L (1985) Data collection frequency effect in the National Medical Care Utilisation and Expenditure Survey, *Journal of Economic and Social Measurement* **13**, 125-151.

Cole, D. and Uttinger, J.E.G. (1956) Estimating expenditure saving and income from household budgets, *Journal of the Royal Statistical Society Series A* **199**, 371-392

Corder, L.S. and Horvitz, D.G. (1989) Panel effects in the National Medical Care Utilisation and Expenditure Survey 304-318 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Couper, M.P. (1991) Modelling Survey Participation at the Interviewer Level in 1991 *1991 Proceedings of the Survey Research Methods Section of the American Statistical Association*, 98-107. http://www.amstat.org/sections/srms/proceedings

Crouchley, R., Bradley, S. and Oskrochi, R. (2005) The impact of sample attrition on employment participation and wages: evidence from the BHPS. Submitted for publication. http://www.lancs.ac.uk/staff/ecasb/papers/seltrunc8.pdf

Davies, R. B. and Crouchley, R. (1992) Sampling for a panel study: the consequences for statistical modelling, in A.P.M. Coxon (ed.) Sample design issues in a panel survey: the case of the British Household Panel Study, *Working Papers of the ESRC Centre on Micro-Social Change*. Paper 3. Colchester: University of Essex.

de Leeuw, E. (1999) Item Non-response: Prevention is Better than Cure, *Survey Methods Newsletter* **19(2)**, 4-8.

DeMaio, Theresa J. (1980) Refusals: Who, Where and Why, *Public Opinion Quarterly* **44(2)**, 223-233.

DeMaio, T.J. (1985) Social Desirability and Survey Measurement: A Review, in C. Turner and E. Martin (eds.), *Surveying Subjective Phenomena,* New York: Russell Sage Foundation and Basic Books.

Dennis, J.M., Saulsberry, C. and Battaglia, M.P. (1999) Analysis of call patterns in a large random-digit dialing survey: The National Immunization Survey, paper presented at the International Conference on Survey Nonresponse, Portland, Oregon, October 28-31.

Dex, S. (1995) The Reliability of Recall Data: A Literature Review, *Bulletin de Méthodologie Sociologique 49*, 58-89.

Dibbs, R., Hale, A., Lovelock, R., and Michaud, S. (1995) Some Effects of Computer-Assisted Interviewing on the Quality of the Survey of Labour and Income Dynamics*. Income Research Paper Series (previously SLID Research Paper Series) paper 95–07,* Ottawa: Statistics Canada. http://www.statcan.ca/bsolc/english/bsolc?catno=75F0002M&CHROPG=1

Dijkstra, W. and Smit, J.S. (2002) Persuading Reluctant Recipients in Telephone Surveys, 121-134, in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (ed.s) *Survey Nonresponse,* New York: John Wiley & Sons.

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

Dillman, D.A. (2000) *Mail and Internet Surveys. The Tailored Design Method,* New York: John Wiley & Sons.

Dillman, D.A. and Tarnai, J. (1991) Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys, 73-93 in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (eds.) *Measurement Errors in Surveys*, New York: Wiley.

Dillman, D.A., Eltinge, J.L., Groves, R.M. and Little, R.J.A. (2002) Survey Nonresponse in Design, Data Collection, and Analysis, 3-26, in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (ed.s) *Survey Nonresponse,* New York: John Wiley & Sons.

Dohrenwend, B.S. and Dohrenwend, B.P. (1968) Sources of Refusals in Surveys, *Public Opinion Quarterly 32(1)*, 74-83.

Donald, M. N. (1960) Implications of Nonresponse for the Interpretation of Mail Questionnaire Data, *Public Opinion Quarterly 24*, 99-114.

Doyle, P., Martin, B., and Moore, J. (2000) *Improving Income Measurement: The survey of Income and Program Participation (SIPP) Methods Panel*, at http://www.census.gov/dusd/MAB/wp234.pdf

Duncan, G.J., Mathiowetz, N.A., Cannell, C.F., Hill, D.H., and Ponza, M. (1985) A Validation Study of Economic Survey Data, *Survey Research Center Research Report Series,* Ann Arbor: University of Michigan.

Dunkelberg, W., and Day, G. (1973) Nonresponse Bias and Callbacks in Sample Surveys, *Journal of Marketing Research* **10**, 160-168.

Eisenhower, D., Mathiowetz, N.A. and Morganstein, D. (1991) Recall Error: Sources and Bias Reduction Techniques, 127-144 in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (eds.), *Measurement Errors in Surveys*, New York: Wiley.

Elliot, D., Lynn, P. and Smith, P. (2007, forthcoming) Sample design for longitudinal surveys, in Lynn, P. (ed.) *Methodology of Longitudinal Surveys*, Chichester: Wiley.

Ernst, L.R. (1989) Weighting issues for longitudinal household and family estimates, 139-159 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Ferber, R. (1955) On the reliability of responses secured in sample surveys, *Journal of the American Statistical Association*, **50**, 788-810.

Ferber, R. (1964) Does a Panel Operation Increase the Reliability of Survey Data: The Case of Consumer Savings, *1964 Proceedings of the Social Statistics Section of the American Statistical Association*, 210-216.

Fisher, R.P. and Quigley, K.L. (1990) Applying Cognitive Theory in Public Health Investigations: Enhancing Food Recall with the Cognitive Interview, in Tanur, J. (ed.) *Questions about Questions: Inquiries into the Cognitive Basis of Surveys*, New York: Russel Sage.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. (1998) The Impact of Attrition in the Panel Study of Income Dynamics on Intergenerational Analysis, *Journal of Human Resources* **33(2)**, 300-344.

Foster, K. (1993) The Electoral Register as a Sampling Frame, *Survey Methodology Bulletin* **33**, 1-11.

Foster, K. (1998) Evaluating Nonresponse on Household Surveys, *GSS Methodology Series* No. 8, London: Government Statistical Service.

Foster, K. and Bushnell, D. (1994) Non-Response Bias on Government Surveys in Great Britain, paper presented at the 5th International Workshop on Household Non-Response, Ottawa, September 26-8th.

Foster, K., Jackson, B., Thomas, M., Hunter, P. and Bennett, N. (1993) *General Household Survey 1993*, London: OPCS Social Survey Division.

Frankel, J., and Sharp, L.M. (1981) Measurement of Respondent Burden, *Statistical Reporter* **81(4)**, 105-111.

Freedman, D., Thornton, A., Camburn, D., Alwin, D. and Young-DeMarco, L. (1988) The Life History Calendar: A Technique for Collecting Retrospective Data, *Sociological Methodology* **18**, 37-68.

Frey, J.H. (1986) An Experiment with a Confidentiality Reminder in a Telephone Survey, *Public Opinion Quarterly* **50(2)**, 267-269.

Friedman, W.J. (1993) Memory for the Time of Past Events, *Psychological Bulletin*, **113(1)**, 44-66.

Galvin, L.N., Sae-Ung, S. and King, K. (2000) Effect of Interview Length and Proxy Interviews on Attrition to the Survey of Income and Program Participation, *2000 Proceedings of Survey Research Methods Section of the American Statistical Association,* 636-640. http://www.amstat.org/sections/srms/proceedings

Goudreau, K., Oberheu, H. and Vaughan, D. (1984) An Assessment of the Quality of Survey Reports of Income from the Aid to Families with Dependent Children (AFDC) Program, *Journal of Business and Economic Statistics* **2(2)**, 179-186.

Goyder, J. C. (1982) Further Evidence on Factors Affecting Response Rates to Mailed Questionnaires, *American Sociological Review* **47**, 550-553.

Goyder, J. and McKenzie Leiper, J. (1985) The Decline in Survey Response: A Social Values Interpretation, Sociology **19(1)**, 55-71.

Goyder, J. (1987) *The Silent Minority - Nonrespondents on Sample Surveys*. Cambridge: Polity Press.

Gray, P.G. (1955) The Memory Factor in Social Surveys, *Journal of the American Statistical Association* **50(270)**, 344-363.

Green, K. E. (1996) Sociodemographic Factors and Mail Survey Response, *Psychology and Marketing* **13(2)**, 171-184.

Grogger, J., Haider, S. J. and Klerman, J. (2003) Why Did the Welfare Rolls Fall during the 1990's? The Importance of Entry, *American Economic Review* **93(2)**, 288-292

Groves, R.M. and Couper, M.P. (1998) *Nonresponse in Household Interview Surveys*, New York: John Wiley and Sons.

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

Groves, R.M., Cialdini, R.B. and Couper, M.P. (1992) Understanding the Decision to Participate in a Survey, *Public Opinion Quarterly* **56(4)**, 475-495.

Groves, R.M., Singer, E. and Corning, A. (2000) Leverage-Saliency Theory of Survey Participation: Description and an Illustration, *Public Opinion Quarterly* **64(3)**, 299-308.

Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (2002) *Survey Nonresponse,* New York: John Wiley & Sons.

Hale, A., and Michaud, S. (1995) Dependent interviewing: impact on recall and on labour market transitions, *Income Research Paper Series (previously SLID Research Paper Series) paper 95–06,* Ottawa: Statistics Canada. http://www.statcan.ca/bsolc/english/bsolc?catno=75F0002M&CHROPG=1

Halpin, B. (1998) Unified BHPS Work-Life Histories: Combining Multiple Sources Into a User-Friendly Format, *Bulletin de Méthodologie Sociologique* **60**, 34-79.

Ham, J. C. and Shore-Sheppard, L. (2005) The Effect of Medicaid Expansions for Low-Income Children on Medicaid Participation and Private Insurance Coverage: Evidence from the SIPP, *Journal of Public Economics,* **89**, 57–83

Hawkins, D. F. (1975) Estimation of Nonresponse Bias, *Sociological Methods and Research* **3**, 461-488.

Heberlein, T. A. and Baumgartner (1979) Factors Affecting Response Rates to Mailed Questionnaires: A quantitive analysis of the published literature, *American Sociological Review* **43**, 447-462.

Hedges, B. (1992) Essex panel study: the sample designs, in A.P.M. Coxon (ed.) Sample design issues in a panel survey: the case of the British Household Panel Study, *Working Papers of the ESRC Centre on Micro-Social Change*. Paper 3. Colchester: University of Essex.

Hiemestra, D. Lavigne, M. and Webber, M. (1993), Labour Force Classification in SLID: Evaluation of 3A Test Results, *Income Research Paper Series (previously SLID Research Paper Series) paper 93–14,* Ottawa: Statistics Canada. http://www.statcan.ca/bsolc/english/bsolc?catno=75F0002M&CHROPG=1

Hill, D. H. (1987) Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods, *1987 Proceedings of Survey Research Methods Section of the American Statistical Association,* 210-215. http://www.amstat.org/sections/srms/proceedings

Hill, D. (1994) The relative empirical validity of dependent and independent data collection in a panel survey, *Journal of Official Statistics* **10(4)**, 359-380. http://www.jos.nu/Articles/abstract.asp?article=104359

Hill, D.H. and Willis, R.J. (2001) Reducing Panel Attrition: A search for effective policy instruments, *Journal of Human Resources* **36(3)**, 416-438.

Himmelweit, H.T., Biberian, M.J. and Stockdale, J. (1978) Memory of past vote: implications of a study of bias in recall, *British Journal of Political Science* **8**, 635-675.

Hox, J.J., de Leeuw, E. and Vorst, H. (1995) Survey Participation as Reasoned Action: A Behavioral Paradigm for Survey Response? *Bulletin de Methodologie Sociologique* **48**, 52-67.

Huttenlocher, J., Hedges, L.V. and Bradburn, N.M. (1990) Reports of Elapsed Time: Bounding and Rounding Processes in Estimation, *Journal of Experimental Psychology: Learning, Memory, and Cognition* **16(2)**, 196-213.

Iyer, R. (1984) NCDS Fourth follow-up 1981: Analysis of Response. *NCDS4 Working Paper*, no. 25, London: National Children's Bureau. http://www.cls.ioe.ac.uk/studies.asp?section=00010002000300070006

Jäckle, A. (2005) Does Dependent Interviewing Really Increase Efficiency and Reduce Respondent Burden?, *ISER Working Paper, No.* 2005-11, Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2005-11.pdf

Jäckle, A. and Lynn, P. (2004) Dependent Interviewing and Seam Effects in Work History Data, *ISER Working Paper, No.* 2004-24, Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-24.pdf

Jaeger, C.M. and Pennock, J.L. (1961) An analysis of consistency of response in household surveys, *Journal of the American Statistical Association*, **56**, 320-327.

James, T.L. (1997) Results of the Wave 1 Incentive Experiment in the 1996 Survey of Income and Program Participation, *1997 Proceedings of Survey Research Methods Section of the American Statistical Association,* 834-839. http://www.amstat.org/sections/srms/proceedings

James, J.M., and Bolstein, R. (1990) The Effect of Monetary Incentives and Follow-up Mailings on the Response Rate and Response Quality in Mail Surveys, *Public Opinion Quarterly* **54(3)**, 346-361.

Jones, A.M., Koolman, X. and Rice, N. (2006) Health-related non-response in the BHPS and ECHP: using inverse probability weighted estimators in nonlinear models, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, **169** (forthcoming).

Jordan, L.A., Marcus, A.C., and Reeder, L.G. (1980) Response Styles in Telephone and Household Interviewing: A Field Experiment, *Public Opinion Quarterly* **44(2)**, 210-222.

Kalton, G., Kasprzyk, D. and McMillen, D.B. (1989) Nonsampling Errors in Panel Surveys, 249-270 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Kalton, G., Lepkowski, J.M., Montanari, G.E., and Maligalig, D. (1990) Characteristics of Second Wave Nonrespondents in a Panel Survey, *1990 Proceedings of Survey Research Methods Section of the American Statistical Association,* 462-467. http://www.amstat.org/sections/srms/proceedings

Kalton, G., Miller, D. P. and Lepkowski, J. (1992) Analyzing Spells of Program Participation in the SIPP, *Technical Report, Survey Research Centre*, Ann Arbor: University of Michigan.

Kalton, G. and Miller, M. E. (1991) The Seam Effect with Social Security Income in the Survey of Income and Program Participation, *Journal of Official Statistics* **7(2)**, 235-245. http://www.jos.nu/Articles/abstract.asp?article=72235

Kalton, G. and Citro, C.F. (1993) Panel surveys: adding the fourth dimension, *Survey Methodology* **19:2**, 205-215.

Kantorowitz, M. (1998) Is it true that Nonresponse Rates in a Panel Survey Increase when Supplement Surveys are Annexed? 121-137 in Koch, A. and Porst, R. (ed.s) *Nonresponse in Survey Research*, ZUMA Nachrichten Spezial No. 4, Mannheim: ZUMA.

Kemsley, W.F.F. (1976) National Food Survey. A Study of Differential Response Based on a Comparison of the 1971 Sample with the Census, *Statistical News* **35**, 18-22.

Kojetin, B.A, Borgida, E. and Snyder, M. (1993) Survey Topic Involvement and Nonresponse Bias, *1993 Proceedings of Survey Research Methods Section of the American Statistical Association,* 838-843.
http://www.amstat.org/sections/srms/proceedings

Kordos, J. (1994) Non-response problems in Polish household surveys, paper presented at the 5th International Workshop on Household Non-Response, Ottawa, September 26-28.

Krosnick, J.A. (1991) Response strategies for coping with the cognitive demands of attitude measure in surveys, *Applied Cognitive Psychology* **5**, 213-236.

Krosnick, J.A. (2002) The causes of no-opinion responses to attitude measures in surveys: they are rarely what they appear to be, 87-101 in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (ed.s) *Survey Nonresponse,* New York: John Wiley & Sons.

Larsen, S.F. and Thompson, C.P. (1995) Reconstructive memory in the dating of personal and public news events, *Memory and Cognition* **23(6)**, 780-790.

Laurie, H., Smith, R. and Scott, L. (1999) Strategies for reducing nonresponse in a longitudinal panel survey, *Journal of Official Statistics* **15(2)**, 269-282. http://www.jos.nu/Articles/abstract.asp?article=152269

Lavallée P (1995) Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method, *Survey Methodology* **21(1)**, 25-32.

Lemaitre, G. (1992) Dealing with the Seam Problem for the Survey of Labour and Income Dynamics, *Income Research Paper Series (previously SLID Research Paper Series) paper 92–05,* Ottawa: Statistics Canada. http://www.statcan.ca/bsolc/english/bsolc?catno=75F0002M&CHROPG=1

Lepkowski, J.M. and Couper, M.P. (2002) Nonresponse in the second wave of longitudinal household surveys, 259-272 in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (ed.s) *Survey Nonresponse,* New York: John Wiley & Sons.

Lievesley, D. (1986) Unit nonresponse in interview surveys. Unpublished manuscript. London: SCPR.

Lillard, L.A. and Panis, C.W.A. (1998) Panel Attrition from the Panel Study of Income dynamics, *Journal of Human Resources* **33(2)**, 437-457.

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* **94(3)**, 489-506.

Linton, M. (1982) Transformations of Memory in Everyday Life, 77-91 in Neisser, U. (ed.) *Memory Observed: Remembering in Natural Contexts*, San Francisco: Freeman,.

Locander, W., and Burton, J.P. (1976) The Effect of Question Forms on Gathering Income Data by Telephone, *Journal of Marketing Research* **13**, 189-192.

Loftus, E.F. and Marburger, W. (1983) Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports with Landmark Events, *Memory and Cognition* **11**, 114-120.

Long, S. K. (1990) *Children and Welfare: Patterns of Multiple Program Participation*, Washington, DC: The Urban Institute.

Loosveldt, G. (1995) The Profile of the Difficult-to-Interview Respondent, Bulletin de Methodologie Sociologique **48**, 68-81.

Loosveldt, G., Pickery, J., and Billiet, J. (1999) Item non-response as a predictor of unit non-response in a panel survey, paper presented at the International Conference on Survey Non-response, Portland, Oregon, USA, October 28-31, 1999. http://www.jpsm.umd.edu/icsn/papers/loosveldt.htm

Luppes, M. and Barnes, B. (1994) On the Use of Incentives - An Overview of Policies in Several Countries, paper presented at the 5th International Workshop on Household Survey Non-Response, Ottawa, September 26-28.

Lyberg, I. and Lyberg, L. (1991) Nonresponse Research at Statistics Sweden, *1991 Proceedings of Survey Research Methods Section of the American Statistical Association,* 78-87. http://www.amstat.org/sections/srms/proceedings

Lyberg, I. and Lyberg, L. (1991) Nonresponse Research at Statistics Sweden, *1991* paper presented at the 2nd International Workshop on Household Survey Non-Response, Washington DC.

Lynn, P. (1993) *Scottish School-Leavers Survey Series: The Survey of 1992 Leavers – Technical Report*, London: Social and Community Planning Research.

Lynn, P. (2001) The Impact of Incentives on Response Rates to Personal Interview Surveys: role and perceptions of interviewers, *International Journal of Public Opinion Research*, **13(3)**, 326-337.

Lynn, P. (2003) Remember, remember the 5th of November (or was it October?): the effects of respondent recall error on survey estimates of crime victimisation, paper presented to Joint Empirical Social Science Seminar Series, University of Essex, March 19.

Lynn, P. (2006) The problem of non-response, in Hox, J., de Leeuw, E. and Dillman, D.A. (ed.s) *The International Handbook of Survey Methodology*, New York: Lawrence Erlbaum Associates.

Lynn, P. and Clarke, P. (2002) Separating refusal bias and non-contact bias: evidence from UK national surveys, *Journal of the Royal Statistical Society Series D (The Statistician)* **51(3)**, 319-333.

Lynn, P., Clarke, P., Martin, J. and Sturgis, P. (2002) The effects of extended interviewer efforts on nonresponse bias, 135-147 in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (ed.s) *Survey Nonresponse,* New York: John Wiley & Sons.

Lynn, P., Jäckle, A., Jenkins, S. P. and Sala, E. (2004) The Impact of Interviewing Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study, *ISER Working Paper, No.* 2004-28, Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-28.pdf

Lynn, P., Jäckle, A., Jenkins, S. P. and Sala, E. (2006) The Effects of Dependent Interviewing on Responses to Questions on Income Sources, *Journal of Official Statistics*: forthcoming.

Lynn, P., Purdon, S., Hedges, B. and McAleese, I. (1994) The Youth Cohort Study: An assessment of alternative weighting strategies and their effects, *Employment Department Research series YCS Report no.30,* Sheffield: Employment Department.

Lynn, P. and Plewis, I. (2003) *Longitudinal Survey of Young People in England (LSYPE) Methodological Development: Report on sub-project 6, sample design*. Report to Department for Education and Skills.

Lynn, P. and Thomas, R. (ed.s) (2006), *Survey Research Practice*, Sage, forthcoming.

Lynn, P., Taylor, B. and Brook, L. (1997) Incentives, Information and Number of Contacts: Testing the effects of these factors on response to a panel survey, *Survey Methods Centre Newsletter* **17(3)**, 7-12.

Mack, S., Huggins, V., Keathley, D., and Sundukchi, M. (1998) Do Monetary Incentives Improve Response Rates in the Survey on Income and Program Participation?, *1998 Proceedings of Survey Research Methods Section of the American Statistical Association,* 529-534.
http://www.amstat.org/sections/srms/proceedings

Marquis, K. H. and Moore, J. (1989) Response Errors in SIPP: Preliminary Results, *Proceedings of the Fifth Annual Research Conference,* Washington, DC, Census Bureau, 515-536.

Marquis, K. H., Moore, J. C. and Huggins, V. J. (1990) Implications of SIPP Record Check Results for Measurement Principles and Practice, *1990 Proceedings of Survey Research Methods Section of the American Statistical Association,* 564-569.
http://www.amstat.org/sections/srms/proceedings

Martini, A. (1989) Seam Effect, Recall Bias, and the Estimation of Labor Force Transition Rates from SIPP, *1989 Proceedings of Survey Research Methods Section of the American Statistical Association,* 387-392.
http://www.amstat.org/sections/srms/proceedings

Marton, A. (1995) Non-Response in the Hungarian Household Surveys, paper presented at the 6th International Workshop on Household Survey Non-Response, Helsinki.

Mason, R., Lesser, V., and Traugott, M.W. (2002) Effect of Item Nonresponse on Nonrespose Error and Inference, 149-161 in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (ed.s) *Survey Nonresponse,* New York: John Wiley & Sons.

Mason, W.M. *et al* (1976) Models of response error in student reports of parental socioeconomic characteristics, 443-484 in Sewell, W.H., Hauser, R.M. and Featherman, D.L. (eds.) *Schooling and Achievement in American Society*, New York: Academic Press.

Massey, J.T., Wolter, C., Wan, S.C. and Liu, K. (1996) Optimum calling patterns for random digit dialed telephone surveys, *1996 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 485-490. http://www.amstat.org/sections/srms/proceedings

Mathiowetz, N. and Duncan, G. (1988) Out of work, out of mind: response errors in retrospective reports of unemployment, *Journal of Business and Economic Statistics* **6**, 221-229.

Mathiowetz, N.A. and McGonagle, K.A. (2000) An assessment of the current state of dependent interviewing in household surveys, *Journal of Official Statistics* **16(4)**, 401- 418. http://www.jos.nu/Articles/abstract.asp?article=164401

McCarthy, P.J. (1980) Some sources of error in labour force estimates from the Current Population Survey' in *Data Collection, Processing, and Presentation: National and Local. Appendix Vol 2: Counting the Labour Force*, National Commission on Employment and Unemployment Statistics, U.S. Government Printing Office, Washington D.C.

McDaniel, S.W., Madden, C.S. and Verille, P. (1987) Do topic differences affect survey non-response? *Journal of the Market Research Society* **29(1)**, 55-66.

McGeoch, J.A. (1932) Forgetting and the Law of Disuse, *Psychology Review* **15**, 352-370.

Means, B., Mingay, D.J., Nigam, A. and Zarrow, M. (1988) A cognitive approach to enhancing health survey reports of medical visits, 537-542 in Gruneberg, M.M., Morris, P.E., Syles, R.N. (eds.), *Practical aspects of memory: current research and issues; Vol. 1 Memory in everyday life.* Chichester UK: Wiley.

Means, B., Nigam, A., Zarrow, M., Loftus, E.F. and Donaldson, M.S. (1989) Autobiographical Memory for Health-Related Events, *Vital and Health Statistics*, Series 6, No. 2, Washington DC: National Center for Health Statistics.

Means, B., Swan, G.E., Jobe, J.B. and Esposito, J.L. (1991) An Alternative Approach to Obtaining Personal History Data, 167-183 in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (eds.), *Measurement Errors in Surveys*, New York: Wiley.

Meier, E. (1991) Response Rate Trends in Britain, *Marketing and Research Today*, **19(2)**, 120-123.

Menard, S. (2002) *Longitudinal Research* (2nd ed.), Beverly Hills, CA: Sage.

Menon, G. (1994) Judgements of behavioural frequencies: memory search and retrieval strategies, 161-172 in Schwarz, N. and Sudman, S. (eds.) *Autobiographical Memory and the Validity of Retrospective Reports*, New York: Springer.

Miller, D. P., Lepkowski, J. M. and Kalton, G. (1992) Estimating Duration of Food Stamp Spells from the SIPP, *1992 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 560-565. http://www.amstat.org/sections/srms/proceedings

Moore, J. and Kasprzyk, D. (1984) Month-to-Month Recipiency Turnover in the ISDP, *1984 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 726-731.

Morton-Williams, J. (1993) *Interviewer Approaches*, Dartmouth Publishing, Aldershot

Murray, T. S., Michaud, S., Egan, M. and Lemaitre, G. (1991) Invisible Seams? The Experience with the Canadian Labour Market Activity Survey, *Proceedings of the US Bureau of the Census Annual Research Conference,* Washington DC, US Census Bureau, 715-730.

Nathan, G. (1999) A Review of Sample Attrition and Representativeness in Three Longitudinal Surveys (The British Household Panel Survey, the 1970 British Cohort Study and The National Child Development Study) *Government Statistical Service, Methodology Series*, No. 13., London: GSS.

Nederhof, A.J. (1987) When neutrality is negative: Pleasantness of most recent survey experience and nonresponse, *Quality and Quantity* **21**, 425-432.

Neter, J. and Waksberg, J. (1964) A study of response errors in expenditure data from household interviews, *Journal of the American Statistical Association* **59**, 18-55.

Neukirch, T. (2002) Nonignorable attrition and selectivity biases in the Finnish subsample of the ECHP: An empirical study using additional register information, Chintex Working paper No.5. http://www.destatis.de/chintex/download/paper5.pdf

Nicolaas, G. and Lynn, P. (2002) Random digit dialling in the UK: viability revisited, *Journal of the Royal Statistical Society Series A (Statistics in Society)* **165(2)**, 297-316.

Nicoletti, C. and Peracchi, F. (2002) A cross-country comparison of survey nonparticipation in the ECHP, *ISER Working Paper, No.* 2002-32, Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2002-32.pdf

Norwood, J. L. and Tanur, J. M. (1994) A Review Measuring Unemployment in the Nineties, *Public Opinion Quarterly* **58(2)**, 277-294.

O'Muircheartaigh, C. (1989) Source of non-sampling error: Discussion, 271-288 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

O'Neil, Michael J. (1979) Estimating the Nonresponse Bias due to Refusals in Telephone Surveys, *Public Opinion Quarterly* **43(2)**, 218-232.

Olson, J.A., and Klein, R.E. (1980) Interviewers perceptions of reasons for participation refusal in a national longitudinal survey, 1979-1980, *1980 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 552-557. http://www.amstat.org/sections/srms/proceedings

Pascale, J. and Mayer, T.S. (2004) Exploring confidentiality issues related to dependent interviewing: preliminary findings, *Journal of Official Statistics* **20(2)**, 357-377. http://www.jos.nu/Articles/abstract.asp?article=202357

Paull, G. (2002) Biases in the Reporting of Labour Market Dynamics, *IFS Working Paper, No.* 02/10, London: Institute for Fiscal Studies. http://www.ifs.org.uk/publications.php?publication_id=2012

Peters, H.E. (1989) Retrospective versus panel data in analysing lifecycle events, *Journal of Human Resources* **23**, 488-513.

Polivka, A. E. and Rothgeb, J. M. (1993) Redesigning the CPS questionnaire, *Monthly Labor Review* **116(9)**, 10-28.

Presser, S. and Traugott, M. (1992) Little White Lies and Social Science Models: Correlated Response Errors in a Panel Study of Voting, *Public Opinion Quarterly,* **56(1)**, 77-86.

Purdon, S., Campanelli, P. and Sturgis, P. (1997) Studying interviewers' calling strategies as a way to improve response rates, in P. Campanelli *et al*, *Can You Hear me Knocking: An Investigation into the Impact of Interviewers on Survey Response Rates*, London, National Centre for Social Research.

Reamer, F.G. (1979) Protecting research subjects and unintended consequences: The effect of guarantees of confidentiality, *Public Opinion Quarterly* **43(4)**, 497-506.

Reiser, B.J., Black, J.B. and Abelson, R.P. (1985) Knowledge Structures in the Organization and Retrieval of Autobiographical Memories, *Cognitive Psychology*, **17**, 89-137.

Rendtel, U. (1990) Teilnahmebereitschaft in Panelstudien: Zwischen Beeinflussung, Vertrauen und Sozialer Selektion, *Kölner Zeitschrift für Soziologie und Sozialpsychologie* **42(2)**, 280-299.

Rendtel, U. (2002) Attrition in Household Panels: A Survey, CHINTEX Working Paper, No. 4. http://www.destatis.de/chintex/download/paper4.pdf

Riphahn, R.T., and Serfling, O. (2002) Item Non-Response on Income and Wealth Questions, *IZA Discussion Paper*, No. 573. http://www.iza.org/

Rips, L. J., Conrad, F. G. and Fricker, S. S. (2003) Straightening the Seam Effect in Panel Surveys, *Public Opinion Quarterly,* **67(4)**, 522-554.

Robins, L. N. (1963) The Reluctant Respondent, *Public Opinion Quarterly* **27(2)**, 276-286.

Robinson, J.A. (1986) Autobiographical memory: a historical prologue, in Rubin, D.C. (ed.), *Autobiographical Memory*, Cambridge: Cambridge University Press.

Rodgers, W. L. (1989) Comparisons of Alternative Approaches to the Estimation of Simple Causal Models from Panel Data, 432-456 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Rodgers, W. L., Brown, C. and Duncan, G. J. (1993) Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages, *Journal of the American Statistical Association,* **88(424)**, 1208-1218.

Romeo, C. J. (2001) Controlling for Seam Problems in Duration Model Estimates, *Journal of Human Resources,* **36(3)**, 467-499.

Rope, D. (1993) Preliminary longitudinal nonresponse research with the CPS and CE. Paper presented at the 4th International Workshop on Household Survey Nonresponse, Bath, UK

Rose, D., Buck, N. and Corti, L. (1991) Design issues in the British Household Panel Study, *Working Papers of the ESRC Centre on Micro-Social Change.* Paper 2. Colchester: University of Essex.

Ross, C.E., and Reynolds, J.R. (1996) The Effects of Power, Knowledge, and Trust on Income Disclosure in Surveys, *Social Science Quarterly* **77(4)**, 899-911.

Rubin, D.C. (1982) On the retention function for autobiographical memory, *Journal of Verbal Learning and Verbal Behaviour* **21**, 21-38.

Rubin, D.C., Baddeley, A.D. (1989) Telescoping is not time compression: a model of the dating of autobiographical events, *Memory and Cognition* **17**, 653-661.

Sala, E. and Lynn, P. (2004) The Effect of Dependent Interviewing on Measures of Employment Characteristics, *ISER Working Paper, No.* 2004-26, Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-26.pdf

Schaeffer, N.C. (1994) Errors of experience: response error in reports about child support and their implications for questionnaire design, in Schwartz, N., Sudman, S. (eds.) *Autobiographical Memory and the Validity of Retrospective Reports*, New York: Springer.

Schleifer, S. (1986) Trends in Attitudes Toward and Participation in Survey Research, *Public Opinion Quarterly* **50(1)**, 17-26.

Schoua-Glusberg, A. and Hunt. E. (1992) Locating NLSY Respondents and maintaining a 90% completion rate in 13 annual rounds, *Symposium 92: Design and Analysis of Longitudinal Surveys, Proceedings* (Statistics Canada), 41-44.

Schräpler, J-P. (2001) Respondent Behavior in Panel Studies. A Case Study of the German Socio-Economic Panel (GSOEP), DIW Discussion Paper No. 244. http://www.diw.de/deutsch/produkte/publikationen/diskussionspapiere/docs/papers/dp244.pdf

Schräpler, J.P. (2003) Respondent Behavior in Panel Studies – A Case Study for Income-Nonresponse by means of the British Household Panel Study (BHPS), *ISER Working Paper, No.* 2003-08, Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2003-08.pdf

Sehulster, J.R. (1989) Content and temporal structure of autobiographical knowledge: remembering twenty-five seasons at the Metropolitan Opera, *Memory and Cognition* **17(5)**, 590-606.

Sharpe, L.M., and Frankel, J. (1983) Respondent Burden: A Test of Some Common Assumptions, *Public Opinion Quarterly* **47(1)**, 36-53.

Shettle, C. and Mooney, G. (1999) Monetary Incentives in Government Surveys, *Journal of Official Statistics* **15**, 231-250. http://www.jos.nu/Articles/abstract.asp?article=152231

Shoemaker, P.J., Eichholz, M., and Skewes, E.A. (2002) Item Nonresponse: Distinguishing Between Don't Know and Refuse, *International Journal of Public Opinion Research* **14(2)**, 193-201.

Silberstein, A.R. and Jacobs, C.A. (1989) Symptoms of repeated interview effects in the Consumer Expenditure Interview Survey, 289-303 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Singer, E. (1984) Public reactions to some ethical issues of social research: Attitudes and behavior, *Journal of Consumer Research* **11**, 501-509.

Singer, E. (2002) The Use of Incentives to Reduce Nonresponse in Household Surveys, 163-177 in Groves, R.M., Dillman, D.A., Eltinge, J.L. and Little, R.J.A. (ed.s) *Survey Nonresponse,* New York: John Wiley & Sons..

Singer, E., Hippler, H.-J. and Schwarz, N. (1992) Confidentiality Assurances in Surveys: Reassurance or Threat? *International Journal of Public Opinion Research* **4(3)**, 256-268.

Singer, E., Mathiowetz, N., and Couper, M. (1993) The Impact of Privacy and Confidentiality Concerns on Survey Participation: The case of the 1990 US Census, *Public Opinion Quarterly* **57(4)**, 465-482.

Singer, E., Van Hoewyk, J., and Gebler, N. (1999) The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys, *Journal of Official Statistics* **15(2)**, 217-230. http://www.jos.nu/Articles/abstract.asp?article=152217

Singer, E., Von Thurn, D.R. and Miller, E.R. (1995) Confidentiality Assurances and Response - A Quantitative Review of the Experimental Literature, *Public Opinion Quarterly* **59(1)**, 66-77.

Singer, E., Van Hoewyk, J. and Maher, M.P. (2000) Experiments with Incentives in Telephone Surveys, *Public Opinion Quarterly* **64(2)**, 171-188.

Sisto, J. and Rendtel, U. (2004) Nonresponse and attrition effect on design-based estimates of household income, 210-219 in Ehling, M, Rendtel, U. et al (ed.s) *Harmonisation of Panel Surveys and Data Quality*, Wiesbaden, Germany: Statistisches Bundesamt.

Skowronski, J.J. *et al* (1991) Social memory in everyday life: recall of self-events and other-events, *Journal of Personality and Social Psychology* **60(6)**, 831-843.

Smith, T.W. (1995) Trends in Nonresponse Rates, *International Journal of Public Opinion Research* **7(2)**, 157-171.

Smith, T.W. (1991) An Analysis of Missing Income Information on the General Social Surveys, *GSS Methodological Report*, No. 71, Chicago: National Opinion Research Center. http://webapp.icpsr.umich.edu/GSS/

Solon, G. (1989) The Value of Panel Data in Economic Research, 486-496 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Sousa-Poza, A., and Henneberger, F. (2000) Wage data collected by telephone interviews: An empirical analysis of the item nonresponse problem and its implications for the estimation of wage functions, *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* **136**, 79-98. Quoted in Riphahn and Serfling (2002).

Sparks, R.F. (1981) Surveys of Victimization: An Optimistic Assessment, *Crime and Justice* **3**, 1-60.

Stanley, J. and Safer, M. (1997) Last time you had 78, how many do you have now? The effects of providing previous reports on current reports of cattle inventories, *1997 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 875 – 879. http://www.amstat.org/sections/srms/proceedings

Steeh, C.G. (1981) Trends in Nonresponse Rates, 1952-1979, *Public Opinion Quarterly* **45(1)**, 40-57.

Sudman, S.and Bradburn, N.M. (1973) Effects of Time and Memory Factors on Response in Surveys, *Journal of the American Statistical Association* **68(344)**, 805-815.

Sudman, S. and Bradburn, N.M. (1974) *Response Effects in Surveys: A Review and Synthesis*, Chicago: Aldine.

Sudman, S., Bradburn, N., Schwarz, N. (1996) *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*, San Fransisco: Jossey-Bass.

Swires-Hennessy, E. and Drake, M. (1992) The optimum time at which to conduct interviews, *Journal of the Market Research Society* **34(1)**, 61-72

Taylor, S. and Lynn, P. (1996) England and Wales Youth Cohort Study (YCS): The effect of time between contacts, questionnaire length, personalisation and other factors on response to the YCS, *Department for Education and Employment Research Paper no.8*, Sheffield: DfEE.

Taylor, B., Heath, A. and Lynn, P. (1996) The British Election Panel Study 1992-95: response characteristics and attrition, *Centre for Research into Elections and Social Trends Working Paper no.40*, Glasgow: University of Strathclyde

Thompson, C.P. (1985) Memory for unique personal events: Effects of pleasantness, *Motivation and Emotion* **9**, 277-289.

Thompson, C.P., Skowronski, J.J. and Lee, D.J. (1987) Reconstructing the Date of a Personal Event, paper presented at the Second Conference on Practical Aspects of Memory, Swansea UK.

Thompson, C.P., Skowronski, J.J., Lee, D.J. (1988) Telescoping in dating naturally occurring events, *Memory and Cognition* **16(5)**, 461-468.

Thompson, C.P., Skowronski, J.J., Larsen, S.F. and Betz, A.L. (1996) *Autobiographical Memory*, Mahwah, New York: Erlbaum.

Tourangeau, R. (1984) Cognitive Sciences and Survey Methods, in Jabine, T., Loftus, E., Straf, M., Tanur, J., Tourangeau, R. (eds.), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines*, Washington, DC: National Academy of Science.

Tourangeau, R., Rips, L.J. and Rasinski, K. (2000) *The Psychology of Survey Response,* Cambridge: Cambridge University Press.

Traugott, M.W. and Katosh, J.P. (1979) Response validity in surveys of voting behaviour, *Public Opinion Quarterly*, **43(3)**, 359 – 377

Triplett, T., Blair, J., Hamilton, T. and Kang, Y.C. (1996) Initial Cooperators Vs. Converted Refusers: Are There Response Behavior Differences, *1994 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 1038–1041. http://www.amstat.org/sections/srms/proceedings

Turner, A.G. (1982) What Subjects of Survey Research Believe About Confidentiality, in J.E. Sieber (ed) *The Ethics of Social Research: Surveys and Experiments*, New York: Springer.

Tzamourani, P. and Lynn, P. (2000) Do respondent incentives affect data quality? Evidence from an experiment, *Survey Methods Newsletter* **20(2)**, 3-7.

U.S.Bureau of the Census (1978) The Current Population Survey: Design and Methodology, *U.S Bureau of the Census Technical Paper No.40,* U.S Government Printing Office, Washington D.C.

US National Research Council (1979) *Privacy and Confidentiality as Factors in Survey Response.* Panel on Privacy and Confidentiality as Factors in Survey Response, Committee on National Statistics, Committee on Behavioral and Social Sciences and Education, Washington, DC: National Academy Press.

Van der Zouwen, J., Dijkstra, W. and van der Vaart, W. (1993) Effects of Measures Aimed at Increasing the Quality of Recall Data, *Bulletin de Méthodologie Sociologique* **39**, 3-19.

Wagenaar, W.A. (1986) My Memory: A Study of Autobiographical Memory over Six Years, *Cognitive Psychology* **18**, 225-252.

Wagner, M., and Motel, A. (2002) Die Qualitaet der Einkommensmessung bei alten Menschen, *Koelner Zeitschrift fuer Soziologie und Sozialpsychologie* **48**, 493-512.

Waksberg, J. and Valliant, R. (1978) *Final Report on the Evaluation and Calibration of NEISS: Report for Consumer Products Safety Commission*, Rockville, Maryland: Westat, Inc..

Wang, K., Cantor, D. and Safir, A. (2000) Panel conditioning in a random digit dial survey, *2000 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 822-827. http://www.amstat.org/sections/srms/proceedings

Waterton, J. and Lievesley, D. (1987) Attrition in a panel survey of attitudes, *Journal of Official Statistics* **3(3)**, 267-282. http://www.jos.nu/Articles/abstract.asp?article=33267

Waterton, J. and Lievesley, D. (1989) Evidence of conditioning effects in the British Social Attitudes Panel, 319-339 in Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. (ed.s) *Panel Surveys*, New York: John Wiley and Sons.

Webber, M. (1994) The Survey of Labour and Income Dynamics: Lessons learned in testing*, Income Research Paper Series (previously SLID Research Paper Series) paper 94–07,* Ottawa: Statistics Canada. http://www.statcan.ca/bsolc/english/bsolc?catno=75F0002M&CHROPG=1

Weeks, M.F., Kulka, R.A. and Pierson, S.A. (1987) Optimal call scheduling for a telephone survey, *Public Opinion Quarterly* **44(1)**, 101-114

Weinberg, D., H. (2002) Survey of Income and Program Participation, Recent History and Future Developments, *Working Paper 232,* US Department of Commerce, Washington DC: Bureau of the Census

Wilcox, J. B. (1977) The Interaction of Refusal and Not-at-home Sources of Nonresponse Bias, *Journal of Marketing Research* **14**, 592-597.

Young, N. (1989) Wave-Seam Effects in the SIPP, *1989 Proceedings of the Survey Research Methods Section of the American Statistical Association,* 393-398. http://www.amstat.org/sections/srms/proceedings.

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, US Census Bureau.