Panel Surveys: Adding the Fourth Dimension

GRAHAM KALTON and CONSTANCE F. CITRO[1]

ABSTRACT

Surveys across time can serve many objectives. The first half of the paper reviews the abilities of alternative survey designs across time – repeated surveys, panel surveys, rotating panel surveys and split panel surveys – to meet these objectives. The second half concentrates on panel surveys. It discusses the decisions that need to be made in designing a panel survey, the problems of wave nonresponse, time-in-sample bias and the seam effect, and some methods for the longitudinal analysis of panel survey data.

KEY WORDS: Panel surveys; Rotating panel surveys; Repeated surveys; Panel attrition; Time-in-sample bias; Seam effect; Longitudinal analysis.

1. INTRODUCTION

Survey populations are constantly changing over time, both in composition and in the characteristics of their members. Changes in composition occur when members enter the survey population through birth (or reaching adulthood), immigration, or leaving an institution (for a noninstitutional population) or leave through death, emigration, or entering an institution. Changes in characteristics include, for example, a change from married to divorced, or from a monthly income of $2,000 to one of $2,500. These population changes give rise to a range of objectives for the analysis of survey data across time. This paper reviews survey designs that produce the data needed to satisfy these various objectives.

The paper is divided into two parts. The first part contains a review of the general issues involved in conducting surveys across time, including the objectives of such surveys and the types of survey design that may be employed. This part is to be found in Section 2. The second, and main, part of the paper discusses one particular survey design, a panel survey that follows the same sample of units through time. The considerations involved in designing, conducting, and analyzing a panel survey are reviewed in Section 3. Section 4 provides some concluding remarks.

2. SURVEYS ACROSS TIME

This section presents an overview of analytic objectives across time, of designs for surveys across time, and of the extent to which different designs can satisfy the various objectives. The discussion relies heavily on Duncan and Kalton (1987), which contains a more detailed treatment of these issues.

Changes in population characteristics and composition over time lead to a variety of objectives for surveys across time. These objectives include the following:

(a) The estimation of population parameters (e.g., the proportion of the population in poverty) at distinct time points.

(b) The estimation of average values of population parameters across time (e.g., the daily intake of iron averaged across a year).

(c) The estimation of net changes, that is changes at the aggregate level (e.g., the change in the proportion of unemployed from one month to the next).

(d) The estimation of gross changes and other components of individual change (e.g., the proportion of persons who were in poverty one year and were not in poverty in the following year).

(e) The aggregation of data for individuals over time (e.g., the summation of twelve monthly incomes to give annual income).

(f) The collection of data on events occurring in a specified time period (e.g., becoming unemployed), and on their characteristics (e.g., duration of spells of unemployment).

(g) The cumulation of samples over time, especially samples of rare populations (e.g., women who become widowed).

(h) The maintenance of a sample of members of a rare population that was identified at one point of time (e.g., scientists and engineers identified from a large-scale survey at one point of time).

[1] Graham Kalton, Westat, 1650 Research Blvd., Rockville, Maryland, U.S.A. 20850; Constance F. Citro, National Research Council, 2101 Constitution Ave. N.W., Washington, D.C., U.S.A., 20418.

A number of survey designs have been developed to provide the data needed to address these objectives. These designs are:

- *Repeated survey.* A repeated survey is a series of separate cross-sectional surveys conducted at different time points. No attempt is made to ensure that any of the same elements are sampled for the individual surveys. The elements are sampled from a population defined in the same manner for each individual survey (*e.g.*, the same geographical boundaries and age-limits) and many of the same questions are asked in each survey.

- *Panel survey.* A panel survey collects the survey data for the same sample elements at different points of time.

- *Repeated panel survey.* A repeated panel survey is made up of a series of panel surveys each of a fixed duration. There may be no overlap in the time period covered by the individual panels, for instance one panel may start only as (or after) the previous one ends, or there may be an overlap, with two or more panels covering part of the same time period.

- *Rotating panel survey.* Strictly, a rotating panel survey is equivalent to a repeated panel survey with overlap. Both limit the length of a panel, and have two or more panels in the field at the same time. However, it seems useful to distinguish between the two designs because they have different objectives. Rotating panel surveys are widely used to provide a series of cross-sectional estimates and estimates of net change (*e.g.*, of unemployment rates and changes in such rates), whereas repeated panel surveys with overlaps also have a major focus on longitudinal measures (*e.g.*, durations of spells of unemployment). In consequence, repeated panel surveys tend to have longer durations and have fewer panels in operation at any given time than rotating panel surveys.

- *Overlapping survey.* Like a repeated survey, an overlapping survey is a series of cross-sectional surveys conducted at different time points. However, whereas the repeated survey does not attempt to secure any sample overlap from the survey at one time point to the next, an overlapping survey is designed to provide such overlap. The aim may be to maximize the degree of sample overlap while taking into account both the changes desired in selection probabilities for sample elements that remain in the survey population and also changes in population composition over time.

- *Split panel survey.* A split panel survey is a combination of a panel survey and a repeated survey or rotating panel survey.

The choice of design in a particular case depends on the objectives to be satisfied. Some designs are better than others for some objectives but poorer for other objectives. Some designs cannot satisfy certain objectives at all. For a detailed discussion, see Duncan and Kalton (1987).

The strength of a repeated survey is that it selects a new sample at each time point, so that each cross-sectional survey is based on a probability sample of the population existing at that time. A panel survey is based on a sample drawn from the population existing at the start of the panel. Although attempts are sometimes made to add samples of new entrants to a panel at later time points, such updating is generally difficult to do and is done imperfectly. Moreover, nonresponse losses from a panel as it ages heighten concerns about nonresponse bias when the panel sample is used to estimate cross-sectional parameters for later time points. For these reasons, repeated surveys are stronger than panel surveys for producing cross-sectional and average cross-sectional estimates (objectives (a) and (b)). With average cross-sectional estimates, another factor to be considered is the correlation between the values of the survey variables for the same individual at different time points. When this correlation is positive, as it generally is, it increases the standard errors of the average cross-sectional estimates from a panel survey. This factor thus also favours repeated surveys over panel surveys for average cross-sectional estimates.

The superior representation of the samples for a repeated survey at later time points also argues in favour of a repeated survey over a panel survey for estimating net change (assuming that the interest in net change relates to changes in both population composition and characteristics). However, in this case the positive correlations of the values of the survey variables for the same individuals across time decreases the standard errors of estimates of net change from a panel survey. Hence the presence of this correlation operates in favour of the panel design for measuring net change.

The key advantages of the panel design are its abilities to measure gross change, and also to aggregate data for individuals over time (objectives (d) and (e)). Repeated surveys are incapable of satisfying these objectives. The great analytic potential provided by the measurement of individual changes is the major reason for using a panel design.

Repeated surveys can collect data on events occurring in a specified period and on durations of events (*e.g.*, spells of sickness) by retrospective questioning. However, retrospective questioning often introduces a serious problem of response error in recalling dates, and the risk of telescoping bias. A panel survey that uses a reference period for the event that corresponds to the interval between waves of data collection can eliminate the telescoping problem by using the previous interview to bound the recall (*i.e.*, an illness reported at the current interview can be discarded if it had already been reported at the previous one). Similarly, a panel survey can determine the duration of an event from successive waves of data collection, limiting the length of recall to the interval between waves.

Repeated data collections over time can provide a vehicle for accumulating a sample of members of a rare population, such as persons with a rare chronic disease or persons who have recently experienced a bereavement. Repeated surveys can be used in this manner to generate a sample of any form of rare population. Panel surveys, however, can be used to accumulate only samples of new rare events (such as bereavements) not of stable rare characteristics (such as having a chronic disease). If a sample of members with a rare stable characteristic (*e.g.*, persons with doctoral degrees) has already been identified, a panel survey can be useful for maintaining the sample over time, with suitable supplementation for new entrants at later waves (for an example, see Citro and Kalton 1989).

Rotating panel surveys are primarily concerned with estimating current levels and net change (objectives (a) and (c)). As such, elements are usually retained in the panel for only short periods. For instance, sample members remain in the monthly Canadian Labour Force Survey for only six months. The extent to which individual changes can be charted and aggregation over time can be performed is thus limited by the short panel duration. A special feature of rotating panel surveys is the potential to use composite estimation to improve the precision of both cross-sectional estimates and estimates of net change (see Binder and Hidiroglou 1988; Cantwell and Ernst 1993). See also Fuller *et al.* (1993) for an alternative method of using past information in forming estimates from a rotating panel design.

Like rotating panel surveys, overlapping surveys are primarily concerned with estimating current levels and net change. They can also provide some limited information on gross change and aggregations over time. Overlapping survey designs are applicable in situations where some sample overlap is required and where the desired element selection probabilities vary over time. This situation arises in particular in establishment surveys, where the desired selection probability for an establishment may vary from one cross-sectional survey to the next to reflect its change in size and type of activity. In such circumstances, a Keyfitz-type procedure can be applied to maximize the retention of elements from the previous survey while taking account of changes in selection probabilities and population composition (see, for example, Keyfitz 1951; Kish and Scott 1971; Sunter 1986). The U.S. Internal Revenue Service Statistics of Income Division's corporate sample provides an example of an overlapping survey design (Hinkins *et al.* 1988).

By combining a panel survey with a repeated survey or a rotating panel survey, a split panel survey can provide the advantages of each. However, given a constraint on total resources, the sample size for each component is necessarily smaller than if only one component had been used. In particular, estimates of gross change and other measures of individual change from a split panel survey will be based on a smaller sample than would have been the case if all the resources had been devoted to the panel component.

In comparing alternative designs for surveys across time, the costs of the designs need to be considered. For instance, panel surveys avoid the costs of repeated sample selections incurred with repeated surveys, but they face costs of tracking and tracing mobile sample members and sometimes costs of incentives to encourage panel members to continue to cooperate in the panel (see Section 3). If two designs can each satisfy the survey objectives, the relative costs for given levels of precision for the survey estimates need to be examined.

3. PANEL SURVEYS

The repeated measures over time on the same sampled elements that are obtained in panel surveys provide such surveys with a key analytic advantage over repeated surveys. The measurements of gross change and other components of individual change that are possible with panel survey data form the basis of a much greater understanding of social processes than can be obtained from a series of independent cross-sectional snapshots. The power of longitudinal data derived from panel surveys has long been recognized (see, for instance, Lazarsfeld and Fiske 1938; Lazarsfeld 1948), and panel surveys have been carried out in many fields for many years. Subjects of panel surveys have included, for example, human growth and development, juvenile delinquency, drug use, victimizations from crime, voting behaviour, marketing studies of consumer expenditures, education and career choices, retirement, health, and medical care expenditures. (See Wall and Williams (1970) for a review of early panel studies on human growth and development, Boruch and Pearson (1988) for descriptions of some U.S. panel surveys, and the Subcommittee on Federal Longitudinal Surveys (1986) for descriptions of U.S. federal panel surveys.) In recent years, there has been a major upsurge in interest in panel surveys in many subject-matter areas, and especially in household economics. The ongoing U.S. Panel Study of Income Dynamics began in 1968 (see Hill 1992 for a description of the PSID) and similar long-term panel studies have been started in the past decade in many European countries. The U.S. Bureau of the Census started to conduct the Survey of Income and Program Participation (SIPP) in 1983 (Nelson *et al.* 1985; Kasprzyk 1988; Jabine *et al.* 1990), and Statistics Canada introduced the Survey of Labour and Income Dynamics (SLID) in 1993. The growth in interest in panel surveys has also given rise to an increase in literature about the methodology of such surveys, including such recent texts as Kasprzyk *et al.* (1989), Magnusson and Bergman (1990), and Van de Pol (1989).

This section reviews the major issues involved in the design and analysis of panel surveys. The treatment is geared towards repeated panel surveys of fixed duration like the SIPP and SLID, but most of the discussion applies more generally to all forms of panel survey.

3.1 Design Decisions for a Panel Survey

The time dimension adds an extra dimension of complexity to a panel survey as compared with a cross-sectional survey. In addition to all the decisions that need to be made about the design features of a cross-sectional survey, a wide range of extra decisions needs to be reached for a panel survey. Major design decisions include:

- *Length of the panel.* The longer the panel lasts, the greater is the wealth of data obtained for longitudinal analysis. For instance, the longer the panel, the greater the number of spells of unemployment starting during the life of the panel that will be completed before the end of the panel, and hence the greater the precision in estimating the survival function for such spells. On the other hand, the longer the panel, the greater the problems of maintaining a representative cross-sectional sample at later waves, because of both sample attrition and difficulties in updating the sample for new entrants to the population.

It can sometimes be beneficial to vary the length of the panel between different types of panel members. Thus, for instance, when the analytic objectives call for it, panel members with certain characteristics (*e.g.*, members of a minority population) or who experience certain events during the course of the regular panel (*e.g.*, a divorce) can be retained in the panel for extended periods of observation.

- *Length of the reference period.* The frequency of data collection depends on the ability of respondents to recall the information collected in the survey over time. Thus, the PSID, with annual waves of data collection, requires recall of events occurring in the previous calendar year, whereas SIPP, with four-monthly waves of data collection, requires recall for the preceding four months. The longer the reference period, the greater the risk of recall error.

- *Number of waves.* In most cases the number of waves of data collection is determined by a combination of the length of the panel and the length of the reference period. The greater the number of waves, the greater the risk of panel attrition and time-in-sample effects, and the greater the degree of respondent burden.

- *Overlapping or non-overlapping panels.* With a repeated panel survey of fixed duration, a decision needs to be made as to whether the panels should overlap across time. Consider, for instance, the proposal of a National Research Council study panel that the SIPP should be a four-year panel (Citro and Kalton 1993). One possibility

is to run each panel for four years, starting a new panel when the previous one finishes. Another possibility is run each panel for four years, but starting a new panel every two years. Yet another possibility is to run each panel for four years, starting a new panel every year.

The design of nonoverlapping panels has the benefit of simplicity, since only one panel is in the field at any one time. It also produces a large sample for longitudinal analysis; for instance, the panels with the nonoverlapping design can be roughly twice the size of those with the design that has two overlapping panels at any one time. However, this increase in sample size for nonoverlapping panels does not apply for cross-sectional estimates, since the data from the panels covering a given time point can be combined for cross-sectional estimation. Also, the cross-sectional estimates for a time period near the end of a panel with the nonoverlapping design are at greater risk of bias from attrition, time-in-sample bias, and failure to update the sample fully for new population entrants than is the case with an overlapping design, in which one panel is of more recent origin. Moreover, the overlapping design permits the examination of such biases through a comparison of the results for the two panels for a given time period, whereas no such examination is possible with a nonoverlapping design. Another limitation of the nonoverlapping design is that it may not be well positioned to measure the effect of such events as a change in legislation. For instance, if legislation takes effect in the final year of a nonoverlapping panel, there will be little opportunity to evaluate its effect by comparing the situations of the same individuals before and for some period after the legislation is enacted. With overlapping panels, one of the panels will provide a wider window of observation.

- *Panel sample size.* For a given amount of annual resources, the sample size for each panel is determined by the preceding factors. A larger panel for longitudinal analysis can be achieved by lengthening the reference period and by employing a nonoverlapping design. The sample size for cross-sectional estimates can be increased by lengthening the reference period, but not by using a nonoverlapping design.

The above list determines the major parameters of a panel survey design, but there still remain a number of other factors that need to be considered:

- *Mode of data collection.* As with any survey, a decision needs to be made as to whether the survey data are to be collected by face-to-face interviewing, by telephone, or by self-completion questionnaire, and whether computer assisted interviewing (CAPI or CATI) is to be used. With a panel survey, this decision needs to be made for each wave of data collection, with the possibility of different modes for different waves (for instance, face-to-face

interviewing at the first wave to make contact and establish rapport, with telephone interviewing or mail questionnaires at some of the later waves). When modes may be changed between waves, consideration needs to be given to the comparability of the data across waves. Sometimes a change in mode may involve a change in interviewer, as for instance would occur with a change from face-to-face interviewing to a centralized CATI operation. Then the effects of a change of interviewer between waves on the respondent's willingness to continue in the panel and on the comparability of responses across waves also need to be carefully considered.

- *Dependent interviewing*. With panel surveys there is the possibility of feeding back to respondents their responses at earlier waves of data collection. This dependent interviewing procedure can secure more consistent responses across waves, but risks generating an undue level of consistency. The ease of application of dependent interviewing depends on the length of the interval between waves and the mode of data collection. Processing the responses from one wave to feed back in the next is easier to accomplish if the interval between waves is a long one and if computer assisted interviewing is employed. Edwards *et al.* (1993) describe the use of dependent interviewing with CAPI in the Medical Care Beneficiary Survey, a survey which involves three interviews per year with each respondent.

- *Incentives*. Monetary or other incentives (*e.g.*, coffee mugs, calculators, lunch bags) may be offered to sampled persons to encourage their participation in a survey. With a panel survey, incentives may be used not only to secure initial participation but also to maintain cooperation throughout the duration of the panel. There is an issue of when are the best times to provide incentives in a panel survey (*e.g.*, at the first wave, at an intermediate wave, or at the last wave of the panel). Panel survey researchers often send respondents a survey newsletter, frequently giving some recent highlights from the survey findings, at regular intervals, both to generate goodwill for the survey and to maintain contact with respondents (see below). Birthday cards sent at the time of the respondents' birthdays are also often used for these purposes.

- *Respondent rules*. Survey data are often collected from proxy informants when respondents are unavailable for interview. With a panel survey, this gives rise to the possibility that the data may be collected from different individuals at different waves, thus jeopardizing the comparability of the data across waves. The respondent rules for a panel survey need to take this factor into account.

- *Sample design*. The longitudinal nature of a panel survey needs to be considered in constructing the sample design for the initial wave. Clustered samples are commonly employed for cross-sectional surveys with face-to-face

interviewing in order to reduce fieldwork travel costs and to enable frame construction of housing unit listings to be performed only for selected segments. These benefits are bought at the price of the increase in the variance of survey estimates arising from the clustering. The optimum extent of clustering depends on the various cost factors involved and the homogeneity of the survey variables in the clusters (see, for instance, Kish 1965). With a panel survey, the use and extent of any clustering should be determined in relation to the overall panel with all its waves of data collection. In particular, the benefit of reduced fieldwork costs disappears for waves of data collection that are conducted by telephone interviewing or mail questionnaire. Also the migration of panel members to locations outside the original clusters reduces the benefit of the initial clustering for fieldwork costs at later waves. (However, some benefits of the initial clustering still operate for the large proportion of mobile persons who move within their own neighbourhoods.)

Oversampling of certain population subgroups is widely used in cross-sectional surveys to provide sufficient numbers of subgroup members for separate analysis. Such subgroups may, for instance, comprise persons with low incomes, minority populations, persons in a specified age-group, or persons living in certain geographical areas. Such oversampling can also be useful in panel surveys, but caution is needed in its application. With long-term panels, one reason for caution is that the objectives of the survey may change over time. Oversampling to meet an objective identified at the start of a panel may prove harmful to objectives that emerge later. Another reason for caution is that many of the subgroups of interest are transient in nature (*e.g.*, low income persons, persons living in a given geographical area). Oversampling persons in such subgroups at the outset of the panel may be of limited value for later waves: some of those oversampled will leave the subgroup while others not oversampled will join it. Thirdly, the definition of the desired subgroup for longitudinal analysis needs to be considered. For instance, SIPP data are used to estimate durations of spells on various welfare programs. Since such estimates are usually based on new spells starting during the life of the panel, it may not be useful to oversample persons already enrolled on welfare programs. See Citro and Kalton (1993) for a discussion of oversampling for the SIPP.

When oversampling of a certain subgroup of the population (*e.g.*, a minority population) is desired for a panel survey, the oversampling may require a large screening operation. The assessment of the cost of such screening should be made in the context of the full panel with all its waves of data collection. An expensive screening operation at the first wave may well be justifiable in this context.

• *Updating the sample.* When the sole objective of a panel survey is longitudinal analysis, it may be sufficient to adopt a cohort approach that simply follows the initial sample selected for the first wave. However, when cross-sectional estimates are also of interest, it may be necessary to update the sample at each wave to represent new entrants to the population. Updating for all types of new entrants is often difficult, but it is sometimes possible to develop fairly simple procedures to account for certain types of new entrants. For instance, in a panel of persons of all ages, babies born to women panel members after the start of the panel can be included as panel members. The SIPP population of inference comprises persons aged 15 and over. By identifying in initial sampled households persons who are under 15 years old but who will attain that age before the end of the panel, by following them during the panel, and by interviewing them after they reach 15 years of age, a SIPP panel can be updated for this class of new entrants (Kalton and Lepkowski 1985).

Attention also needs to be paid to panel members who leave the survey population. For some the departure is clearly permanent (*e.g.*, deaths), but for others it may be only temporary (*e.g.*, going abroad or entering an institution). If efforts are made to keep track of temporary leavers, they can be readmitted to the panel if they return to the survey's population of inference.

Panel surveys such as SIPP and PSID collect data not only for persons in original sampled households, but also for other persons – nonsampled persons – with whom they are living at later waves. The prime purpose of collecting survey data for nonsampled persons is to be able to describe the economic and social circumstances of sampled persons. The issue arises as to whether any or all nonsampled persons should remain in the panel after they stop living with sampled persons. For some kinds of analysis it is useful to follow them. However, to follow them would eat significantly into the survey's resources.

When data are collected for nonsample members, these data may be used simply to describe the circumstances of sample members, in which case analyses are restricted to sample members, with nonsample members being assigned weights of zero. Alternatively, nonsample members can be included in cross-sectional analyses. In this case appropriate weights for sample and nonsample persons need to be developed to reflect the multiple ways in which individuals may appear in the dataset. Huang (1984), Ernst (1989) and Lavallée and Hunter (1993) describe the fair share weighting approach that may be used for this purpose.

• *Tracking and tracing.* Most panel surveys encounter the problem that some panel members have moved since the last wave and cannot be located. There are two ways to try to handle this problem. First, attempts can be made

to avoid the problem by implementing procedures for tracking panel members between waves. One widely-used procedure when there is a long interval between waves is to send mailings, such as birthday cards and survey newsletters, to respondents between waves, requesting the post office to provide notification of change of address if applicable. Another tracking device is to ask respondents for the names and addresses or telephone numbers of persons close to them (*e.g.*, parents) who are unlikely to move and who will be able to provide locating information for them if they move.

The second way to deal with lost panel members is to institute various tracing methods to try to locate them. With effort and ingenuity, high success rates can be achieved. Some methods of tracing may be specific for the particular population of interest (*e.g.*, professional societies for persons with professional qualifications) while others may be more general, such as telephone directories, computerized telephone number look-ups, reverse telephone directories for telephone numbers of neighbours, mail forwarding, marriage licence registers, motor vehicle registrations, employers, and credit bureaus. It can be useful to search death records for lost panel members, particularly for long-term panel surveys. Panel members found to have died can then be correctly classified, rather than being viewed as non-respondents. Methods of tracing are discussed by Burgess (1989), Clarridge *et al.* (1978), Crider *et al.* (1971) and Eckland (1968).

3.2 Problems of Panel Surveys

Panel surveys share with all surveys a wide range of sources of nonsampling error. This section does not review all these sources, but rather concentrates on three sources that are unique to panel surveys, namely wave nonresponse, time-in-sample bias and the seam effect.

3.2.1 Wave nonresponse

The nonresponse experienced by panel surveys at the first wave of data collection corresponds to that experienced by cross-sectional surveys. The distinctive feature of panel surveys is that they encounter further nonresponse at subsequent waves. Some panel members who become non-respondents at a particular wave do not respond at any subsequent wave while others respond at some or all subsequent waves. The former are often termed attrition cases and the latter non-attrition cases. The overall wave nonresponse rates in panel surveys increase with later waves, but with well-managed surveys the rate of increase usually declines appreciably over time. For example, with the 1987 SIPP panel, the sample loss was 6.7% at wave 1, 12.6% at wave 2, and it then increased slowly to 19.0% at wave 7 (Jabine *et al.* 1990). The tendency for the nonresponse rate to flatten off at later waves is comforting,

but nevertheless the accumulation of nonresponse over many waves produces high nonresponse rates at later waves of a long-term panel. For instance, in 1988, after 21 annual rounds of data collection, the PSID non-response rate for individuals who lived in 1968 sampled households had risen to 43.9% (Hill 1992).

The choice between the two standard general-purpose methods for handling missing survey data – weighting adjustments and imputation – is not straightforward for wave nonresponse in panel surveys. For longitudinal analysis, the weighting approach drops all records with one or more missing waves from the data file and attempts to compensate for them by weighting adjustments applied to the remaining records. This approach can lead to the loss of a substantial amount of data when the data file covers several waves. On the other hand, the imputation approach retains all the reported data, but requires conducting wholesale imputations for missing waves. A compromise approach uses imputation for some patterns of wave nonresponse (*e.g.*, those with only one missing wave, where data are available from both adjacent waves), and weighting for others (see, for example, Singh *et al.* 1990). For cross-sectional analysis, separate data files may be created for each wave. These files can comprise all the respondents for that wave, with either weighting adjustments or imputations for the wave nonrespondents. Kalton (1986) and Lepkowski (1989) discuss general methods for handling wave nonresponse, Lepkowski *et al.* (1993) discuss imputations for wave nonresponse in the SIPP, and Michaud and Hunter (1993) describe plans for handling wave nonresponse in the SLID.

With wave nonresponse there is the possibility of collecting some or all of the data for the missing wave at a subsequent interview. However, the quality of the retro-spective data collected in this way needs to be carefully assessed. An experiment was conducted to examine the utility of this approach with the 1984 SIPP panel, using a missing wave form to collect responses for a skeleton set of core questions for the missing wave (Huggins 1987; Singh 1993). The analyses showed substantially fewer transitions in receipt of income, assets, and government assistance from the missing wave form than from bench-mark data. In consequence the use of the missing wave form was discontinued. Administrative records may some-times provide another possible source of skeletal data for missing waves.

3.2.2 Time-in-sample bias

Time-in-sample bias, or panel conditioning, refers to the effect that panel members' responses at a given wave of data collection are affected by their participation in previous waves. The effect may reflect simply a change in reporting behaviour. For example, a respondent may recognize from previous interviews that a "Yes" response

to a question leads to follow-up questions, whereas a "No" answer does not. The respondent may therefore give a "No" answer to avoid the burden of the extra questions. Alternatively, a respondent may learn from previous inter-views that detailed information on income is needed, and may therefore prepare for later interviews by collecting the necessary data. The time-in-sample effect may also reflect a change in actual behaviour. For example, a respondent may enroll in the food stamp program as a result of learning of its existence from the questions asked about it at earlier waves of data collection.

A recent experimental study of panel conditioning in a four-year panel study of newlyweds found some evidence that participation in the study did affect marital well-being (Veroff *et al.* 1992). However, that study used in-depth interviewing techniques that are more intrusive than those used in most surveys. A number of studies of panel conditioning that have been conducted in more standard survey settings have found that conditioning effects do sometimes occur, but they are not pervasive (Traugott and Katosh 1979; Ferber 1964; Mooney 1962; Waterton and Lievesley 1989).

A benefit of rotating and overlapping panel surveys is that they enable estimates for the same time period obtained from different panels to be compared. Such comparisons have clearly identified the presence of what is termed "rotation group bias" in the U.S. and Canadian Labour Force Surveys (*e.g.* Bailar 1975, 1989, and U.S. Bureau of the Census 1978, for the U.S. Current Population Survey; Ghangurde 1982, for the Canadian Labour Force Survey). Rotation group bias may reflect nonresponse bias and conditioning effects. In analyses comparing the overlapping 1985, 1986 and 1987 SIPP panels, Pennell and Lepkowski (1992) found few differences in the results from the different panels.

3.2.3 Seam effect

Many panel surveys collect data for subperiods within the reference period from the last wave of data collection. The SIPP, for instance, collects data on a monthly basis within the four-month reference period between waves. The seam effect refers to the common finding with this form of data collection that the levels of reported changes between adjacent subperiods (*e.g.*, going on or off of a welfare program from one month to the next) are much greater when the data for the pair of subperiods are collected in different waves than when they are collected in the same wave. The seam effect has been found to be pervasive in SIPP, and to relate to both recipiency status and amounts received (see, for example, Jabine *et al.* 1990; Kalton and Miller 1991). It has also been found in PSID (Hill 1987). Murray *et al.* (1991) describe approaches used to reduce the seam effect in the Canadian Labour Market Activity Survey.

3.3 Longitudinal Analysis

There is a substantial and rapidly expanding literature on the analysis of longitudinal data, including a number of texts on the subject (*e.g.* Goldstein 1979; Hsiao 1986; Kessler and Greenberg 1981; Markus 1979). This treatment cannot be comprehensive, but rather identifies a few general themes.

- *Measurement of gross change.* As has already been noted, a key analytic advantage of a panel survey over a repeated survey is the ability to measure gross change, that is, change at the individual level. The basic approach to measuring gross change is the turnover table that tabulates responses at one wave against the responses to the same question at another wave. The severe limitation to this form of analysis is that changes in measurement errors across waves can lead to serious bias in the estimation of the gross change (for further discussion, see Kalton *et al.* 1989; Rodgers 1989; Abowd and Zellner 1985; Chua and Fuller 1987; Fuller 1990; and Skinner 1993).

- *Relationship between variables across time.* Panel surveys collect the data necessary to study the relationships between variables measured at different times. For instance, based on the data collected in the 1946 British birth cohort, the National Survey of Health and Development, Douglas (1975) found that children who were hospitalized for more than a week or who had repeated hospitalizations between the ages of 6 months and 3½ years exhibited more troublesome behaviour in school and lower reading scores at age 15. In principle, cross-section surveys may employ retrospective questions to collect the data needed to perform this type of analysis. However, the responses to such questions are often subject to serious memory error, and potentially to systematic distortions that affect the relationships investigated.

- *Regression with change scores.* Regression with change scores can be used to avoid a certain type of model misspecification. Suppose that the correct regression model for individual i at time t is

$$Y_{it} = \alpha + \beta x_{it} + \gamma z_{it} + \epsilon_{it},$$

where x_{it} is an explanatory variable that changes value over time and z_{it} is an explanatory variable that is constant over time (*e.g.*, gender, race). Suppose further that z_{it} is unobserved; it may well be unknown. Then β can still be estimated from the regression on the change scores:

$$Y_{i(t+1)} - Y_{(t)} = \beta(x_{i(t+1)} - x_{it}) + \epsilon_{i(t+1)} - \epsilon_{it},$$

(Rodgers 1989; Duncan and Kalton 1987).

- *Estimation of spell durations.* The data collected in panel surveys may be used to estimate the distribution of lengths of spells of such events as being on a welfare program. In panel surveys like the SIPP, some individuals have a spell in progress at the start of the panel (initial-censored spells), some start a spell during the panel, and some spells continue beyond the end of the panel (right-censored spells). Thus, not all spells are observed in their entirety. The distribution of spell durations may be estimated by applying survival analysis methods, such as the Kaplan-Meier product-limit estimation procedure to all new spells (including right-censored new spells) starting during the life of the panel (*e.g.* Ruggles and Williams 1989).

- *Structural equation models with measurement errors.* The sequence of data collection in a panel survey provides a clear ordering of the survey variables that fits well with the use of structural equation modelling for their analysis. This form of analysis can make allowance for measurement errors, and with several repeated measures can handle correlated error structures (*e.g.* Jöreskog and Sörbom 1979).

4. CONCLUDING REMARKS

The data sets generated from panel surveys are usually extremely rich in analytic potential. They contain repeated measures for some variables that are collected on several occasions, and also measures for other variables that are asked on a single wave. Repeated interviewing of the same sample provides the opportunity to collect data on new variables at each wave, thus yielding data on an extensive range of variables over a number of waves. A panel data set may be analyzed both longitudinally and cross-sectionally. Repeated measures may be used to examine individual response patterns over time, and they may also be related to other variables. Variables measured at a single wave may be analyzed both in relation to other variables measured at that wave and to variables measured at other waves.

The richness of panel data is of value only to the extent that the data set is analyzed, and analyzed in a timely manner. Running a panel survey is like being on a treadmill: the operations of questionnaire design, data collection, processing and analysis have to be undertaken repeatedly for each successive wave. There is a real danger that the survey team will become overwhelmed by this process, with the result that the data are not fully analyzed. To avoid this danger, adequate staffing is needed and a well-integrated organization needs to be established.

In addition it is advisable to keep the panel survey design simple. The survey design should be developed to meet clearly-specified objectives. Adding complexities to

the design to enhance the richness of the panel data set for other uses should be critically assessed. Although persuasive arguments can often be made for such additions, they should be rejected if they threaten the orderly conduct of any stage of the survey process.

As noted earlier, measurement errors have particularly harmful effects on the analysis of individual changes from panel survey data. The allocation of part of a panel survey's resources to measure the magnitude of such errors is therefore well warranted (Fuller 1989). Measurement errors may be investigated either by validity studies (comparing survey responses with "true" values from an external source) or by reliability studies (*e.g.*, reinterview studies). The results of such studies may be then used in the survey estimation procedures to adjust for the effects of measurement errors.

REFERENCES

ABOWD, H.M., and ZELLNER, A. (1985). Estimating gross flows. *Journal of Business and Economic Statistics*, 3, 254-283.

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

BAILAR, B.A. (1989). Information needs, surveys, and measurement errors. *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 1-24.

BINDER, D.A., and HIDIROGLOU, M.A. (1988). Sampling in time. *Handbook of Statistics*, (Vol. 6), (Eds. P.R. Krishnaiah and C.R. Rao). New York: North Holland, 187-211.

BORUCH, R.F., and PEARSON, R.W. (1988). Assessing the quality of longitudinal surveys. *Evaluation Review*, 12, 3-58.

BURGESS, R.D. (1989). Major issues and implications of tracing survey respondents. *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 52-74.

CANTWELL, P.J., and ERNST, L.R. (1993). New developments in composite estimation for the Current Population Survey. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys, Statistics Canada*, 121-130.

CHUA, T.C., and FULLER, W.A. (1987). A model for multinomial response error applied to labor flows. *Journal of the American Statistical Association*, 82, 46-51.

CITRO, C.F., and KALTON, G. (1989). *Surveying the Nation's Scientists and Engineers*. Washington DC: National Academy Press.

CITRO, C.F., and KALTON, G. (1993). *The Future of the Survey of Income and Program Participation*. Washington DC: National Academy Press.

CLARRIDGE, B.R., SHEEHY, L.L., and HAUSER, T.S. (1978). Tracing members of a panel: a 17-year follow-up. *Sociological Methodology*, (Ed. K.F. Schuessler). San Francisco: Jossey-Bass, 389-437.

CRIDER, D.M., WILLITS, F.K., and BEALER, R.C. (1971). Tracking respondents in longitudinal surveys. *Public Opinion Quarterly*, 35, 613-620.

DOUGLAS, J.W.B. (1975). Early hospital admissions and later disturbances of behaviour and learning. *Developmental Medicine and Child Neurology*, 17, 456-480.

DUNCAN, G.J., and KALTON, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review*, 55, 97-117.

ECKLAND, B.K. (1968). Retrieving mobile cases in longitudinal surveys. *Public Opinion Quarterly*, 32, 51-64.

EDWARDS, W.S., SPERRY, S., and EDWARDS, B. (1993). Using CAPI in a longitudinal survey: a report from the Medicare Current Beneficiary Survey. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys, Statistics Canada*, 21-30.

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

FERBER, R. (1964). Does a panel operation increase the reliability of survey data: the case of consumer savings. *Proceedings of the Social Statistics Section, American Statistical Association*, 210-216.

FULLER, W.A. (1989). Estimation of cross-sectional and change parameters: Discussion. *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 480-485.

FULLER, W.A. (1990). Analysis of repeated surveys. *Survey Methodology*, 16, 167-180.

FULLER, W.A., ADAM, A., and YANSANEH, I.S. (1993). Estimators for longitudinal surveys with application to the U.S. Current Population Survey. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys, Statistics Canada*, 309-324.

GHANGURDE, P.D. (1982). Rotation group bias in the LFS estimates. *Survey Methodology*, 8, 86-101.

GOLDSTEIN, H. (1979). *The Design and Analysis of Longitudinal Studies*. New York: Academic Press.

HILL, D. (1987). Response errors around the seam: analysis of change in a panel with overlapping reference periods. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 210-215.

HILL, M.S. (1992). *The Panel Study of Income Dynamics: A User's Guide*. Newbury Park, CA: Sage Publications.

HINKINS, S., JONES, H., and SCHEUREN, F. (1988). Design modifications for the SOI corporate sample: Balancing multiple objectives. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 216-221.

HSIAO, C. (1986). *Analysis of Panel Data*. New York: Cambridge University Press.

HUANG, H. (1984). Obtaining cross-sectional estimates from a longitudinal survey: Experiences of the Income Survey Development Program. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 670-675.

HUGGINS, V. (1987). Evaluation of missing wage data from the Survey of Income and Program Participation (SIPP). *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 205-209.

JABINE, T.B., KING, K.E., and PETRONI, R.J. (1990). *Survey of Income and Program Participation Quality Profile*. Bureau of the Census, Washington DC: U.S. Department of Commerce.

JÖRESKOG, K.G., and SÖRBOM, D. (1979). *Advances in Factor Analysis and Structural Equation Models*. Lanham MD: University Press of America.

KALTON, G. (1986). Handling wave nonresponse in panel surveys. *Journal of Official Statistics*, 2, 303-314.

KALTON, G., KASPRZYK, D., and McMILLEN, D.B. (1989). Nonsampling errors in panel surveys. *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 249-270.

KALTON G., and LEPKOWSKI, J.M. (1985). Following rules in SIPP. *Journal of Economic and Social Measurement*, 13, 319-329.

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, 235-245.

KASPRZYK, D. (1988). *The Survey of Income and Program Participation: An Overview and Discussion of Research Issues*. SIPP Working Paper No. 8830. Washington DC: U.S. Bureau of the Census.

KASPRZYK, D., DUNCAN G., KALTON, G., and SINGH, M.P. (Eds.) (1989). *Panel Surveys*. New York: John Wiley.

KESSLER, R.C., and GREENBERG, D.F. (1981). *Linear Panel Analysis*. New York: Academic Press.

KEYFITZ, N. (1951). Sampling with probabilities proportional to size: adjustment for changes in the probabilities. *Journal of the American Statistical Association*, 46, 183-201.

KISH, L. (1965). *Survey Sampling*. New York: John Wiley.

KISH, L., and SCOTT, A. (1971). Retaining units after changing strata and probabilities. *Journal of the American Statistical Association*, 66, 461-470.

LAVALLÉE, P., and HUNTER, L. (1993). Weighting for the Survey of Labour and Income Dynamics. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*, Statistics Canada, 65-75.

LAZARSFELD, P.F. (1948). The use of panels in social research. *Proceedings of the American Philosophical Society*, 42, 405-410.

LAZARSFELD, P.F., and FISKE, M. (1938). The panel as a new tool for measuring opinion. *Public Opinion Quarterly*, 2, 596-612.

LEPKOWSKI, J.M. (1989). Treatment of wave nonresponse in panel surveys. *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 348-374.

LEPKOWSKI, J.M., MILLER, D.P., KALTON, G., and SINGH, R. (1993). Imputation for wave nonresponse in the SIPP. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*, Statistics Canada, 99-109.

MAGNUSSON, D., and BERGMAN, L.R. (Eds.) (1990). *Data Quality in Longitudinal Research*. New York: Cambridge University Press.

MARKUS, G.B. (1979). *Analyzing Panel Data*. Beverly Hills, CA: Sage Publications.

MICHAUD, S., and HUNTER, L. (1993). Strategy for minimizing the impact of non-response for the Survey of Labour and Income Dynamics. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*, Statistics Canada, 89-98.

MOONEY, H.W. (1962). *Methodology in Two California Health Surveys*. Public Health Monograph No. 70, Washington DC: U.S. Department of Health, Education, and Welfare.

MURRAY, T.S., MICHAUD, S., EGAN, M., and LEMAÎTRE, G. (1991). Invisible seams? The experience with the Canadian Labour Market Activity Survey. *Proceedings of the 1991 Annual Research Conference*. U.S. Bureau of the Census. Washington DC: U.S. Department of Commerce, 715-730.

NELSON, D., McMILLEN, D., and KASPRZYK, D. (1985). *An Overview of the SIPP, Update 1*. SIPP Working Paper No. 8401. Washington DC: U.S. Bureau of the Census.

PENNELL, S.G., and LEPKOWSKI, J.M. (1992). Panel conditioning effects in the Survey of Income and Program Participation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 566-571.

RODGERS, W.L. (1989). Comparisons of alternative approaches to the estimation of simple causal models from panel data. *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley, 432-456.

SINGH, R.P. (1993). Methodological experiments in the Survey of Income and Program Participation. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*, Statistics Canada, 157-166.

SINGH, R., HUGGINS, V., and KASPRZYK, D. (1990). *Handling Single Wave Nonresponse in Panel Surveys*. SIPP Working Paper No. 9009, Bureau of the Census, Washington DC: U.S. Department of Commerce.

SKINNER, C.J. (1993). Logistic modelling of longitudinal survey data with measurement error. *Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys*, Statistics Canada, 269-276.

SUBCOMMITTEE ON FEDERAL LONGITUDINAL SURVEYS (1986). *Federal Longitudinal Surveys*. Statistical Policy Working Paper 13. Washington DC: Office of Management and Budget.

SUNTER, A.B. (1986). Implicit longitudinal files: A useful technique. *Journal of Official Statistics*, 2, 161-168.

TRAUGOTT, M., and KATOSH, K. (1979). Response validity in surveys of voting behavior. *Public Opinion Quarterly*, 43, 359-377.

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

VAN DE POL, F.J.R. (1989). *Issues of Design and Analysis of Panels*. Amsterdam: Sociometric Research Foundation.

VEROFF, J., HATCHETT, S., and DOUVAN, E. (1992). Consequences of participating in a longitudinal study of marriage. *Public Opinion Quarterly*, 56, 315-327.

WALL, W.D., and WILLIAMS, H.L. (1970). *Longitudinal Studies and the Social Sciences*. London: Heinemann.

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