

2005

longitudinal surveys
methodology

luzetarako inkesten
metodologia

metodología de las
encuestas
longitudinales

PETER LYNN

45

CONTENTS

1. WHAT IS A LONGITUDINAL SURVEY?.....	3
1.1 Strengths of longitudinal surveys	4
Data collection strengths.....	4
Analysis strengths	6
1.2 Weaknesses of longitudinal surveys	8
Data collection weaknesses.....	8
Analysis weaknesses	9
1.3 Survey error framework.....	10
2. SURVEY DESIGN AND SAMPLE DESIGN FOR LONGITUDINAL SURVEYS.....	13
2.1 Fixed panel.....	13
2.2 Fixed panel plus “births”	13
2.3 Repeated panel.....	15
2.4 Rotating panel.....	16
2.5 Split panel	18
2.6 Differences between designs.....	19
2.7 Representing “births”	20
Coverage of new entrants	21
Coverage of short-term entrants	21
3. NON-RESPONSE AND ATTRITION IN LONGITUDINAL SURVEYS.....	23
3.1 Components of non-response.....	23
Reasons for non-contact.....	24
Reasons for refusal.....	25
3.2 Special features of non-response on longitudinal surveys.....	27
Non-contact	27
Refusals.....	28
3.3 Non-response patterns on longitudinal surveys	29
4. MEASUREMENT ERROR ON LONGITUDINAL SURVEYS.....	34
4.1 Introduction to measurement error.....	34
4.2 Measurement error in estimates of change	35
4.3 Seam effects	37
4.4 Methods for reducing measurement error in longitudinal surveys.....	38
Dependent interviewing.....	38
Calendar and related methods.....	47
5. WEIGHTING AND IMPUTATION IN LONGITUDINAL SURVEYS.....	51
5.1 Introduction to weighting	51
5.2 Longitudinal weights	51
5.3 Which combinations of waves?.....	52
5.4 Which variables?	54
5.5 Introduction to imputation.....	55
5.6 Longitudinal imputation.....	56
5.7 Revision of imputations.....	56
REFERENCES	58

LIST OF TABLES

Table 1: Perceived benefits and drawbacks of survey participation.....	26
Table 2: Measurement Error in a Categorical Variable	35
Table 3: True transition matrix.....	36
Table 4: Observed transition matrix, with measures subject to measurement error	36
Table 5: Income: reporting at wave t conditional upon reporting at wave $t-1$	42
Table 6: Income receipt indicators from administrative and survey data	43

Table 7: Characteristics of independent and reactive reporters of income sources.....44
Table 8: Percentage reporting change in employment characteristics.....45
Table 9: Socio-demographic correlates of the effect of dependent interviewing on estimates of change.....46
Table 10: Study populations and samples52

LIST OF FIGURES

Figure 1: Survey Error Framework11
Figure 2: Errors of Non-Observation12
Figure 3: Fixed Panel Design14
Figure 4: Fixed Panel Design plus "Births"14
Figure 5: Repeated Panel Design16
Figure 6: Rotating Panel Design: 1-1-1 Rotation Pattern17
Figure 7: Rotating Panel Design: 1-1-0-0-1-1 Rotation Pattern18
Figure 8: Split Panel Design.....19
Figure 9: Sampling New Births.....22
Figure 10: Possible wave response patterns under policy 1: all units at every wave.....30
Figure 11: Possible wave response patterns under policy 2: wave 1 respondents at every wave.....31
Figure 12: Possible wave response patterns under policy 3: previous wave respondents only.....31
Figure 13: Wave response patterns on the England and Wales Youth Cohort Study, Cohort 5.....31
Figure 14: An Event History Calendar50

1. WHAT IS A LONGITUDINAL SURVEY?

A *longitudinal survey* is one that collects data from the same sample elements on multiple occasions over time. This is not co-terminous with *longitudinal data*. Longitudinal data are data which *refer* to the same sample elements on multiple occasions over time, but they need not necessarily have been *collected* on multiple occasions. They could have been collected on one single occasion, by retrospective recall or by extracting information from records. The distinguishing feature of longitudinal surveys is therefore the data collection process, though longitudinal surveys certainly produce longitudinal data, so several aspects of such data are important in considering the methodology of longitudinal surveys.

Longitudinal surveys vary greatly in terms of the nature of the information collected, the nature of the population being studied, and the primary objectives (Binder, 1998). Some examples that illustrate this variation are:

- **Surveys of businesses carried out by national or regional statistical offices.** These surveys tend to collect a limited range of information, typically restricted to key economic indicators. Data may be collected at frequent intervals, such as monthly or quarterly and the main objectives are usually to publish regular series of statistics on totals, means and net change between periods, often for cross-cutting domains such as regions and industries;
- **Surveys of school-leavers, graduates or trainees.** Institutions offering education or training, such as Universities, or government agencies responsible for related policy, often wish to assess the outcomes of such education or training at a micro (student) level. These outcomes are often medium-term or long-term and consequently it is necessary to keep in touch with students/trainees for some time after they have completed their study. Longitudinal surveys are often used for this purpose, collecting data from students/trainees on several occasions, perhaps beginning while they are still students/trainees and for up to several years after they have completed the course. The information collected is often quite complex, including perhaps full employment and activity histories between each survey wave and maybe also reasons for changes and decisions made. Sometimes, one or more 'control' groups may be included in the survey in an attempt to assess the impact of the education/training;
- **Household panel surveys.** In several countries, long-term panel surveys of the general household population are carried out. The oldest, the Panel Survey of

Income Dynamics (PSID) in the USA, has been interviewing the same people since 1968. These surveys are multi-topic and general purpose, collecting behavioural, attitudinal and circumstantial data on a range of social and economic issues. The main objective is to provide a rich data resource to be used by a wide range of users for a broad set of purposes. The data structure is complex, involving interviews with each person in the household of each sample member at each wave, in addition to household-level information and, often, additional survey instruments such as self-completion questionnaires or health measurements.

There is also considerable variation between longitudinal surveys in practical constraints as diverse as the level of financial resources available, the means by which the study population can be accessed, and regulations on respondent burden. All of this variation leads to a wide variety of survey design, as discussed in section 2.

1.1 Strengths of longitudinal surveys

The strengths of longitudinal surveys are associated with data collection and analysis possibilities that either cannot be achieved with cross-sectional surveys or cannot be achieved in a satisfactorily accurate or reliable way.

Data collection strengths

- a) It is possible to collect much *longer* continuous histories of events and transitions than could be collected retrospectively in a single interview, simply due to the volume of data involved (and hence the length of the interview or questionnaire);
- b) It is possible to collect more accurate data than would be possible in a single interview with retrospective recall, in which the data might be subject to severe recall error;
- c) It is possible to collect information about expectations and choices, untainted by subsequent events and outcomes, and also about the subsequent events and outcomes for the same sample units.

In particular, the length of the history being collected and the accuracy of the data are typically inter-related. Almost all surveys questions require some degree of recall on the part of the respondent. But the greater the extent to which the circumstances or events that are the subject of the questions are current or recent, the less reliance is made on the respondents' recall. If a survey aims to collect a record of every occurrence of a particular

type of event, the extent of error in the respondent reports will depend on the *recall period* and the *saliency* of the events in question. For a high saliency event, such as giving birth or getting married, most respondents will be able to recall the essential details of the event many years later. So, it is possible to collect lifetime fertility and marriage histories retrospectively in a single interview and obtain reasonably accurate data. But for low saliency events, such as routine purchases of groceries, essential details may be available in respondent memory only for a few days. In a single interview, reasonably accurate data can only be obtained about purchases in the last few days. If questions are asked about routine purchases over a longer period, the responses – if sample members are willing to give them - will be subject to large error. To collect histories of purchasing behaviour over a period of weeks or months, with reasonable accuracy, it is therefore necessary to collect the data from respondents at regular intervals, probably at least once a week.

Many events are of course intermediate in saliency between giving birth and purchasing groceries. For many people, spells of employment, unemployment, education and other activity statuses can be recalled with reasonable accuracy over several months and possibly years. But of course the *difficulty* of the recall task will vary between respondents depending on the number and nature of events that they have experienced. This variation can be very considerable in the case of economic activity histories, causing a dilemma for survey designers. If a survey is aiming to collect complete activity histories over several years for a sample of people who are likely to vary greatly in their experiences, such as a cross-section of the general population, the ideal interval between survey waves will be very different for different sample members. But it is rarely possible to predict this in advance, nor is it often practical to have different between-wave intervals for different individuals. Instead, a standard interval is chosen. Interviews at annual intervals may be inefficient for persons whose circumstances change little (e.g. retired persons or those who remain in the same job for many years). The marginal amount of information collected in each interview, relative to the considerable cost, will be small. But annual interviews might present a considerable recall challenge to persons who experience large numbers of short spells of employment, perhaps interspersed with spells of unemployment or other activities. Thus, to gain maximum benefit from the ability of longitudinal surveys to collect longer and/or more accurate histories of events, the survey designer needs to understand the recall and reporting task being asked of respondents, how it relates to between-wave interval, and how this might vary over sample members.

For many purpose, accurate dating of events is at least as important as accurate recall of the details of the event. But respondents may not be able to recall accurately the date of a specific event, even if they can recall the event itself. Consequently, retrospective

recall questions asked in a single interview may produce biased estimates of frequencies and associated measures. A commonly reported phenomenon is ‘telescoping’, whereby survey respondents report events as having taken place within a reference period when in fact they took place longer ago. Panel surveys have an extra advantage when collecting dates of events. Each interview after the first is ‘bounded’ by the previous interview, so any events reported previously can be discounted from the reports in the current interview in order to avoid telescoping. This of course assumes that it can be unambiguously concluded whether or not reports in two consecutive interviews refer to the *same* event. Sometimes this is difficult, particularly when a respondent has a tendency to experience frequent events of a similar nature, but for many types of survey data it can usually be achieved.

Data about expectations and choices (point c) above) are often desired in order to inform evaluations of outcomes of various kinds and to help understand the processes that lead to those outcomes. It is very difficult for survey respondents to recall their expectations at an earlier point in time or the reasons for which they made certain decisions. Instead, there is a tendency in many situations to “recreate” the reasons in the light of subsequent experiences. If a certain positive outcome arose subsequent to a certain decision made by the respondent, the respondent may rationalize *post hoc* that the reason for the decision must have been in order to achieve that outcome. Questions about expectations and motivations must therefore be asked contemporaneously. Only a longitudinal survey can therefore link the responses to such questions to data about outcomes.

Analysis strengths

It is of course artificial to separate data analysis advantages of longitudinal surveys from analysis advantages, as the reason for collecting a certain type of data is in order to be able to carry out certain types of analyses. The key advantages of longitudinal data (which in most cases can only be accurately collected by longitudinal surveys) are analytical and include the following:

- a) The analysis of gross change;
- b) The analysis of average change at a unit level;
- c) The analysis of stability or instability in characteristics;
- d) The analysis of time-related characteristics of events or circumstances, such as frequency, timing or duration;
- e) Analysis of the ordinal nature of events, which often sheds light on issues of causality.

Analysis of gross change is perhaps one of the most common objectives of longitudinal surveys. Repeated cross-sectional surveys can be used to estimate net change, e.g. the change in employment rate amongst a particular population. But only a longitudinal survey can identify the extent to which this is composed of different elements of gross change. For example, if the same unemployment rate is observed at two time points, is it the same persons who are unemployed on both occasions, or are there equal and opposite flows into and out of unemployment (and, if so, how large are they and what kinds of people are they composed of, etc)? These are the kinds of questions that longitudinal surveys can address.

Individual-level change can also be of interest independently of interest in population-level net change. For example, understanding the nature of, and characteristics associated with, individual-level change in marital status, household composition, income and so on is of great interest to many analysts and can only be achieved with a longitudinal survey. However, often individual level change can only be well interpreted in the context of changes over a considerable period of time. For example, a 2-wave longitudinal survey may be a good vehicle for measuring change in personal income between two points in time. This will allow the analyst to decompose net change in income into its gross change components, i.e. to estimate the distribution of individual-level changes in income. But while the sample distribution of individual-level changes may estimate well the population distribution, each individual observed change may not represent well the 'average' change in income for that individual over a period of time. If we would like to study the associations between personal characteristics and income change, a single measure of change between two points in time may not be a good indicator to use. With a multi-wave panel that collects measures of income at each wave, it may be possible to construct a measure of the 'average' change in income over a relatively long period for each sample member or the variation in change.

Indeed, the extent to which measures such as income are stable or instable over time is of great relevance to policy. Panel surveys with many waves can provide good measures of stability of many characteristics. Poverty analysts, for example, have used household panel data to demonstrate that there is considerable instability over time in the poverty status of many individuals and households in westernized countries. While the proportion of households in poverty may remain relatively stable over time, there may be many entrants to and exits from poverty. A large proportion of households may experience at least one spell of poverty over a long period of time, while very few households may remain continuously in poverty throughout the period. This insight provided by longitudinal surveys may have shifted the policy focus from (stable) characteristics associated with poverty

propensity at a point in time to better understanding poverty dynamics and the factors associated with falling into poverty or persistently failing to exit from poverty.

Understanding the duration of spells in a particular state and the factors associated with exiting from the state (sometimes referred to as 'persistence') are important not only for poverty, but also for many other issues such as unemployment, marital and partnership status, participation in education and training, and company profitability. Hazard modeling and survival analysis are techniques used to better understand the propensity for change (in any status of substantive interest) and the factors associated with such change. These techniques require longitudinal data, which must typically be provided by longitudinal surveys. Researchers would ideally like to identify not only the factors *associated* with change but the factors which *cause* change. Understanding of causal factors can lead directly to policy implications. Longitudinal data can be of great assistance in establishing causality as the chronological ordering of events and changes can be understood. For example, a cross-sectional survey can establish an association between A and B. But a longitudinal survey can establish that for most population units that have experienced both A and B, A happened before B, making it rather more likely that A caused B than that B caused A (though of course a third factor, C, may have caused both A and B and this possibility must always be considered).

1.2 Weaknesses of longitudinal surveys

Longitudinal surveys also have some limitations relative to other surveys. Careful attention must be paid to these at the survey design stage and at the analysis stage.

Data collection weaknesses

There are two aspects of survey data collection that are unique to longitudinal surveys and potentially detrimental:

- a) Panel conditioning.
- b) Panel attrition.

Panel conditioning refers to the possibility that survey responses given by a person who has already taken part in the survey previously may differ from the responses that that person would have given if they were taking part for the first time. In other words, the response may be 'conditioned' by the previous experience of taking part in the survey. This therefore relates to all data collected by longitudinal surveys other than that collected at the

first wave. There are two ways in which conditioning can take place. The way in which respondents *report* events, behaviour or characteristics might change; or the actual *behaviour* might change.

For example, a two-wave survey of unemployed persons might find that more people report a particular type of job search activity at the second wave than at the first wave. This might reflect a genuine increase in the extent to which that activity takes place (independent of taking part in the survey). But it could also be affected by panel conditioning. This could be because the first wave interview made some sample members aware of possible job search activities that they were not currently doing, so the subsequently started doing those things. Thus, there was a genuine increase in the extent of the activity, but only amongst sample members – not amongst the population as a whole. The *behaviour* of the sample members has been conditioned by the experience of the first interview. Alternatively, the experience of the first interviews may have affected the way that some sample members respond to the questions in the second interview, even though their actual job search behaviour may not have changed. Perhaps in the first interview they discovered that reporting no activity of a particular type led to them being asked a series of questions about why they did not participate in that activity. So, to make the second interview shorter, or to avoid embarrassing questions, they now report that they have participated in this particular activity. In this case, the *reporting* of the sample members has been conditioned by the experience of the first interview.

Sample attrition (also referred to as “panel attrition”) refers to the continued loss from the sample due to non-response at each wave of a longitudinal survey. The response rate at any one wave of a longitudinal survey may be just as good as that for any other survey but after, say, five waves the proportion of sample units that have responded at every wave may be quite low. Thus, the effective response rate for longitudinal analysis – for which data from every wave is required – may be lower than the response rates that we are used to having on cross-sectional surveys. The question of how and why sample attrition occurs and what can be done to minimize its detrimental impacts is discussed in section 3 below.

Analysis weaknesses

Longitudinal surveys are often not as good as cross-sectional surveys at providing cross-sectional estimates. This may be perceived as a weakness, but it is simply not something that longitudinal surveys are designed for. Compared with a estimates from a cross-sectional survey, cross-sectional estimates from a longitudinal survey (from wave 2

onwards) may be more likely to suffer from coverage error (because the sample was selected longer ago and may not include recent additions to the population of interest). Also, a longitudinal survey sample may suffer from a lower response rate than a cross-sectional survey (though this does not necessarily imply greater non-response error). The design of a longitudinal survey can often be adjusted to improve the quality of cross-sectional estimates that can be made (see section 2) though this is likely to be resource-intensive and may detract from the central longitudinal aims of the survey.

1.3 Survey error framework

In this section we introduce a general framework for errors in survey-based estimates, based on Groves (1989). This provides the context for the error sources that will be discussed in the remaining four sections of this text. Throughout those sections, we shall refer back to the error framework. A survey error is simply a difference between a survey-based estimate, \hat{Y} , of a population parameter and the true value of the parameter, Y . However, at the survey design stage we are typically concerned with *expected* error of our estimate, so we are interested in the properties of the estimator y rather than a specific realised value of the estimate, \hat{Y} . The *mean square error* of the estimator is the quantity that is often used to measure survey error:

$$MSE(y) = E(y - Y)^2$$

This can be decomposed as follows:

$$\begin{aligned} MSE(y) &= E(y - E[y])^2 + (E[y] - Y)^2 \\ &= Var(y) + Bias^2(y) \end{aligned}$$

So, for any estimator of interest, survey error has both a variance and a bias component. The various potential sources of error are summarised in Figure 1. Each of these sources could contribute either to the variance, or to the bias, or both, of the estimator. This is indicated by the “B” and “V” in the bottom row of the diagram. Error sources can usefully be classified into those which are due to not observing every unit in the study population (“errors of non-observation”) and those which are due to imperfect observation on the units that are studied (“observational errors”).

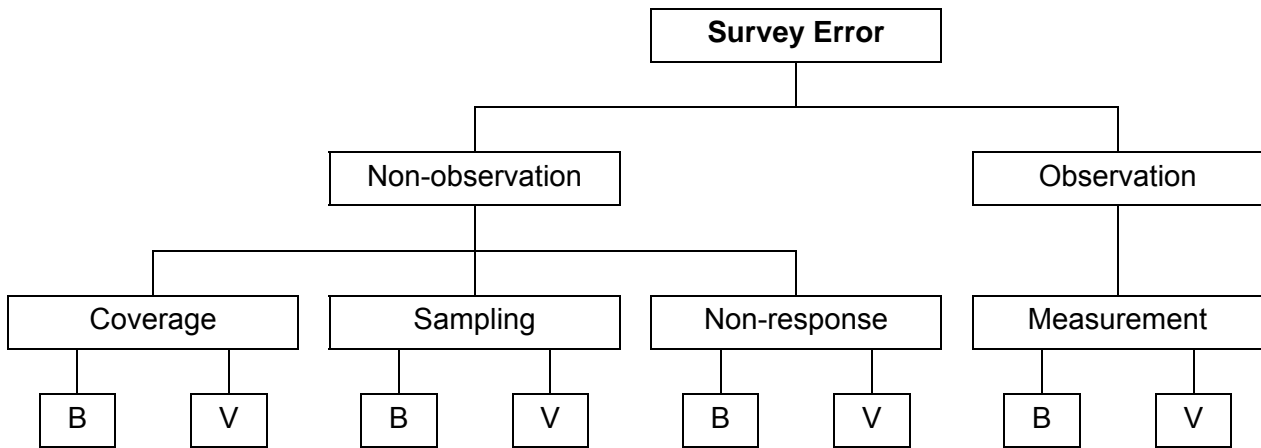


Figure 1: Survey Error Framework

Errors of non-observation can arise at three stages of survey process. First, the sampling frame or sampling method may not give complete coverage of the population. If some population units have a zero chance of being selected, then this will introduce coverage error. Second, we select only a sample of the units on the frame and this introduces sampling error as the sampled units may not have exactly the same characteristics as the complete set of units on the frame. Third, we typically do not succeed in obtaining observations from every sampled unit: we have some non-response, which can introduce non-response error. These three stages of the process of observation and their associated errors are illustrated in Figure 2. Observational errors are essentially co-terminous with measurement error, if one considers the entire process of asking questions, recording answers, coding, data entry and data processing to constitute the measurement process. (Some authors refer to measurement error as coming solely from the process of asking questions and recording answers and refer to errors arising at subsequent stages as *processing errors*. In any case, it can be useful to identify and study the stages at which errors occur.)

This is a general framework that applies to any survey. For longitudinal surveys, the sources of error are just the same but the nature of the errors and the techniques that might be used to reduce them are quite distinct. Those will be the focus of this text.

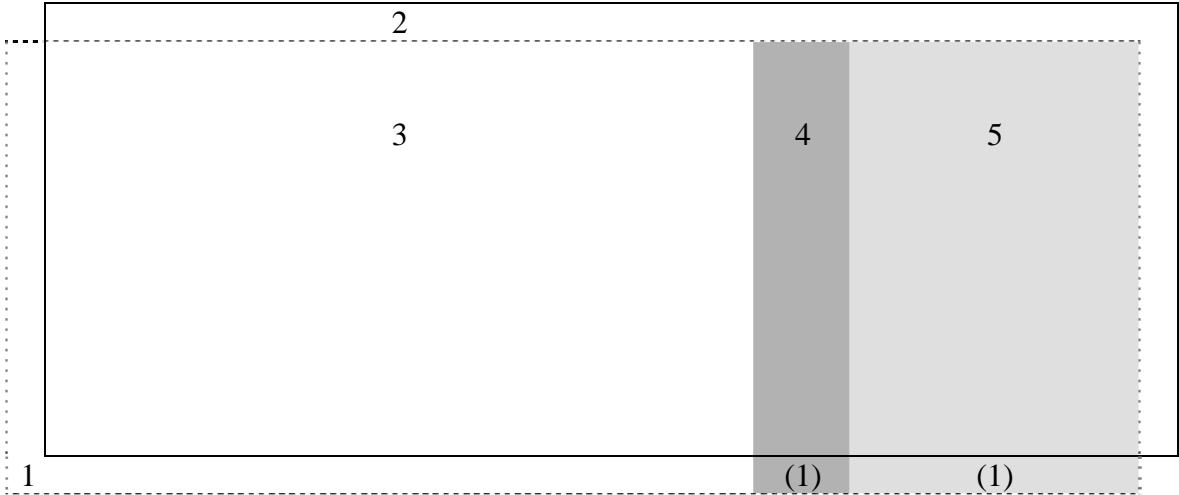
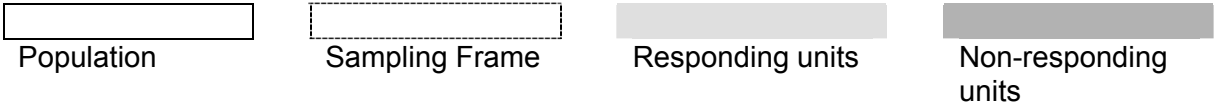


Figure 2: Errors of Non-Observation



- 1: Frame over-coverage (no error, if identified)
- 2: Frame under-coverage (coverage error)
- 3: Non-sampled units (sampling error)
- 4: Non-responding units (non-response error)
- 5: Responding units (observational error)

2. SURVEY DESIGN AND SAMPLE DESIGN FOR LONGITUDINAL SURVEYS

Section 1 showed that longitudinal surveys can be of many kinds. In this section we outline the main types of design and explain why they are used. Longitudinal survey design can be categorized into five broad types: fixed panels, fixed panels with births, repeated panels, rotating panels and split panels. We describe each in turn.

2.1 Fixed panel

This involves attempting to collect survey data from the same units on multiple occasions. After the initial sample selection, no additions to the sample are made. In principle, the only loss to the eligible sample is through “deaths” from the population (e.g. death of a person, closure of a company). This is illustrated in Figure 3 for the example of a 5-wave fixed panel design, ignoring non-response. A sample from the population of interest is selected to take part in wave 1. Some time later, wave 2 takes place. By this time, some sample units have left the population of interests and thereby become ineligible for the survey. These are often referred to as sample “deaths”. In the case of a survey of persons this might include actual death as well as emigration, for example. In the case of a survey of businesses, this might consist of businesses that close. In consequence, the sample eligible for data collection at wave 2 is smaller than that at wave 1. The amount by which it is smaller depends on the time interval between waves and the rate of deaths in the study population.

2.2 Fixed panel plus “births”

This is like a fixed panel, except that regular samples of recent “births” to the population are added. For example, at each wave of data collection a sample of units “born” since the previous wave might be added. This may be preferred to a fixed panel if there are non-trivial numbers of births during the life of a panel and there is a desire to represent the cross-sectional population at the time of each wave as well as the longitudinal population of units surviving since wave 1. Most household panel surveys have this design as a sample of “births” into the eligible age range is added at each wave. Figure 4 illustrates such a design. It can be seen that if the rate of deaths is similar to the rate of births (shown by the vertical scale in Figure 4) then the overall sample size eligible for data collection at each wave remains roughly constant.

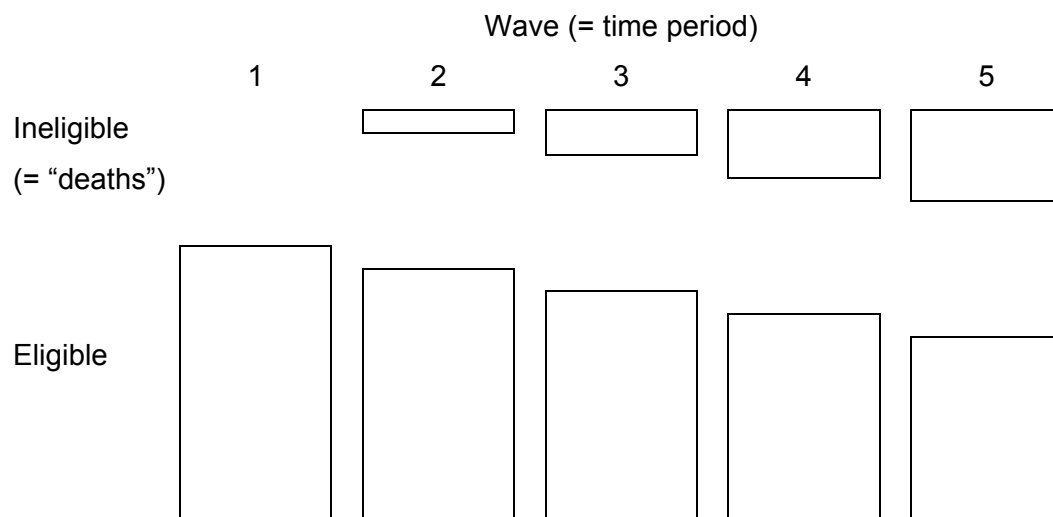


Figure 3: Fixed Panel Design

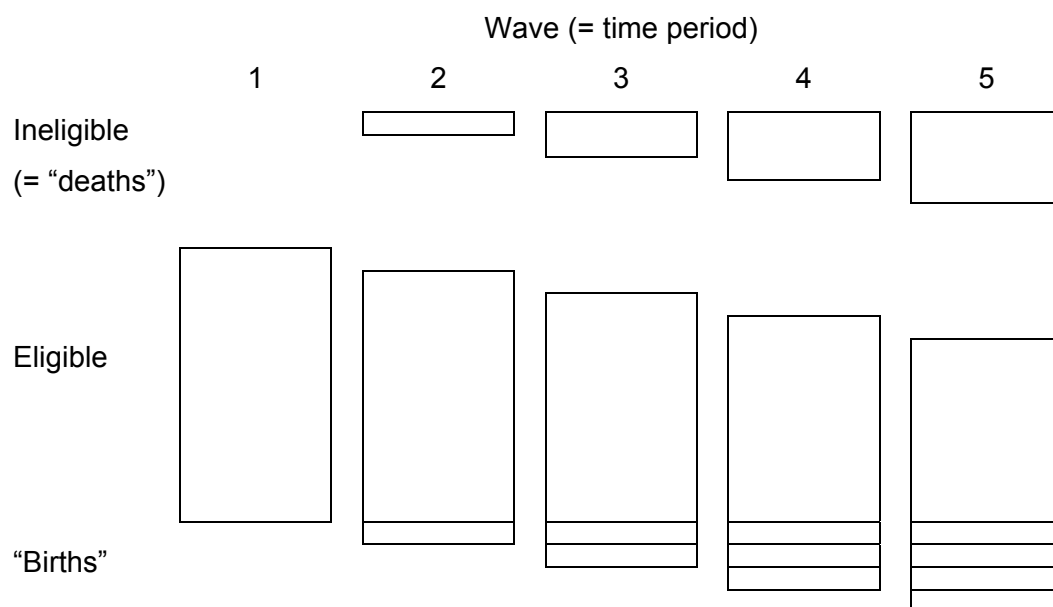


Figure 4: Fixed Panel Design plus "Births"

One important advantage of such a design over the basic fixed panel design is that the sample at each wave is representative of the current cross-sectional population, enabling cross-sectional estimates to be made in parallel with the longitudinal estimates.

For longitudinal analysis, the design also has the advantage of permitting better representation of the population of events such as transitions that take place during a particular period. For example, suppose that the survey is of persons living in a particular city and that survey waves take place at one year intervals. Further suppose that the survey interview asks questions about all periods of employment over the past year and that we are

interested in studying the nature of job changes in the city over a 5-year period in order to better understand the dynamics of the labour market. With a fixed panel design (no additional samples of “births”), at wave 2 the job changes that we observe are restricted to people who have lived in the city at least one year. At wave 3, they are restricted to people who have lived in the city at least two years. And so on. After five years of the survey, the sample of job changes that has accrued will under-represent those experienced by people who have recently moved to the city. Job changes experienced within a year of moving to the city will only be represented in the wave 1 data. Job changes experienced between one and two years of moving to the city will only be represented in the wave 1 and 2 data. And so on. And it is likely that job changes experienced soon after moving to the city will be different in important ways from those experienced after a longer period of living in the city. So, the observed sample of job changes will be biased with respect to the population of all job changes in the city over the 5-year period.

If, at each wave, we add a sample of persons who have moved to the city in the past year, we overcome this problem. The sample of job changes observed over a period of years should be representative of all job changes (though see section 2.7). However, note that the additional samples of “births” cannot be used for all types of longitudinal analysis. If we wish to study micro-level change between years t and $t+4$ (e.g. waves 1 and 5 in Figure 4), for which we need observations from both of those years for each sample unit, then the samples of births added to the survey in each of the intervening years cannot contribute to the analysis.

2.3 Repeated panel

This design involves a series of panel surveys, which may or may not overlap in time. Typically, each panel is designed to represent an *equivalent* population, i.e. the same population definition applied at a different point in time. Surveys of school leavers or university graduates often have this design, each panel consisting of a sample of a particular one-year age cohort selected in different years, each panel involving at least three waves over at least three years.

Figure 5 illustrates a simple example of such a design. In this example, each panel has three waves of data collection and the timing of these waves corresponds with the timing of the start of a new panel. (Note that births and deaths, as shown in Figure 3 and Figure 4, are omitted from Figure 5 for clarity, though in practice these might be a feature of each panel.)

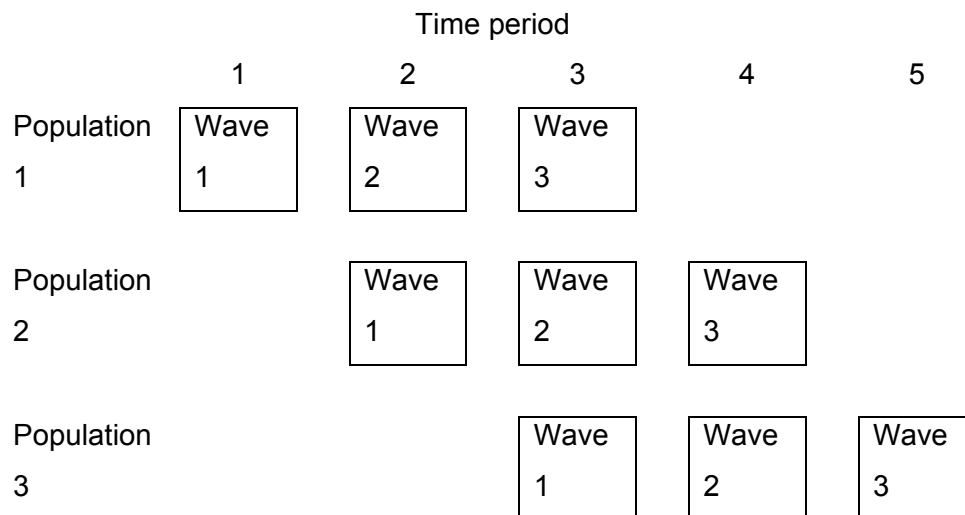


Figure 5: Repeated Panel Design

Population j might consist, for example, of students who graduate in year j . They are sent a questionnaire in each of years j , $j+1$ and $j+2$. But in year $j+1$ the new population of students graduating that year are also sent a questionnaire for the first time, and so on.

Unlike the fixed panel and fixed panel plus births designs, the samples from which data are collected at any particular time period are not collectively representative of a relevant population. In year 3 in Figure 5, wave 1 data are collected from the sample of year 3 graduates, wave 2 data are collected from the sample of year 2 graduates, and wave 3 data are collected from the sample of year 1 graduates. These are three distinct samples representing three different populations. And it is quite possible that different questions are asked at each wave. So these three sets of data would not be added together to form the basis of any analyses. Rather, the main aim of surveys with this design is typically to understand the dynamics of change for each population – i.e. by performing longitudinal analysis on the three waves of data for a particular population – and then to compare these estimate dynamics between populations.

It is of course not necessary that the time interval between each new panel starting is the same as the time interval between each wave of data collection with a panel. For example, a new panel of graduates might begin only once every two years, while data collection with each panel could take place every six months.

2.4 Rotating panel

In a rotating panel design, pre-determined proportions of sample units are replaced at each fieldwork occasion. Typically, each unit will remain in the sample for the same number

of waves. A wide variety of rotation patterns have been used, but we will introduce the idea with a simple pattern in which each unit remains in the sample for three waves, one-third of the sample being replaced each time. This is illustrated in Figure 6. Although this looks very similar to the repeated panel design in Figure 5, it is fundamentally different as each sample is intended to represent the *same* population. In consequence, the three samples interviewed in any one time period can be combined to make cross-sectional estimates. Surveys that use rotating panel designs usually collect the same data from each unit at each wave in order to permit such combination.

In some situations, depending on the interval between waves and the rate of births into the population, it may be necessary to up-weight sampled recent births when making cross sectional estimates. For example, suppose that in each time period, 5% of units in the population “die” and are replaced by a similar number of “births”. Only samples 2 and 3 will contain units born since period 1 and only sample 3 will contain units born since period 2. The total sample included at period 3 will consist of 1.75% units born since period 2, 3.5% born between periods 1 and 2, and 94.7% born before period 1; whereas the population will contain 5% born since period 2, 4.75% born between periods 1 and 2, and 90.25% born before period 1. Clearly, cross-sectional estimates could be biased if this sample composition was not correctly taken into account.

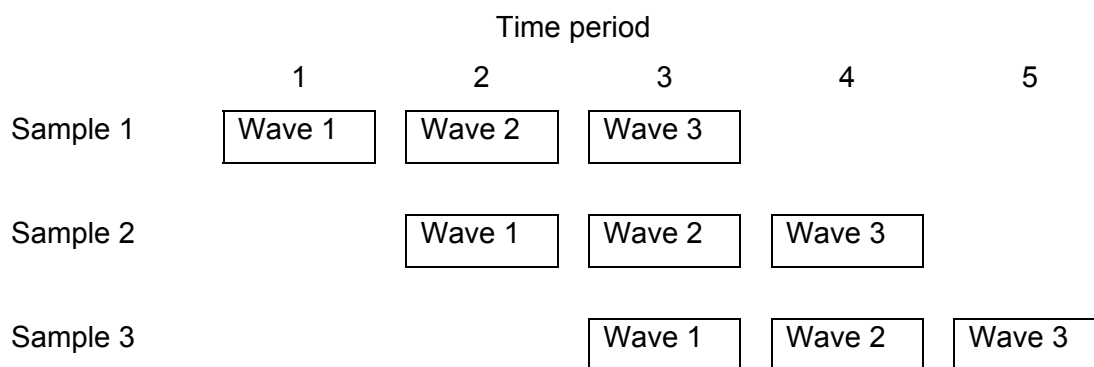


Figure 6: Rotating Panel Design: 1-1-1 Rotation Pattern

As Kalton and Citro (1993) note, a rotating panel is in fact a special case of a repeated panel with overlap. It is special because the overlap pattern is fixed, and is typically balanced, and precisely because each panel that is ‘live’ at one point in time is designed to represent the *same* population, allowing combination of the panels for cross-sectional estimation. Rotating panel designs are often used when the main objectives are cross-sectional estimates and short-term estimates of net and gross change. Labour Force Surveys have a rotating panel design in many countries.

Several common rotating panel designs involve units not being included in the survey at every period between the time they first join the panel and the time they leave. An example is illustrated in Figure 7. The reason for adopting such designs is connected with the trade-off between, on the one hand, respondent burden and representation of recent births, and on the other hand, variance of estimates. Suppose that the time periods in Figure 7 are in fact calendar quarters and that a main objective of the survey is to provide estimates of change from one quarter to the next (“quarterly change”). Further suppose that there is a seasonal component to change, as is often the case with business statistics. Then there will be interest in comparing the quarterly change in any particular quarter with the quarterly change in the *same* quarter of the previous year. For example, the change in period 7 (from period 6) will be compared with the change in period 3 (from period 2). The difference in levels between periods 3 and 7 may also be of direct interest. We can see that with the rotation pattern of Figure 7, half of the sample units in period 3 are also sample units in period 7. And half of the sample units that can be used to estimate change from period 2 to 3 can also be used to estimate change from period 6 to period 7. This high degree of *sample overlap* reduces the variance of estimates of change. A higher degree of overlap would reduce variance further, but this would also increase respondent burden. At the extreme, 100% overlap could be achieved only with the fixed panel design introduced earlier. But that would mean that the sample units would have to remain in the sample indefinitely.

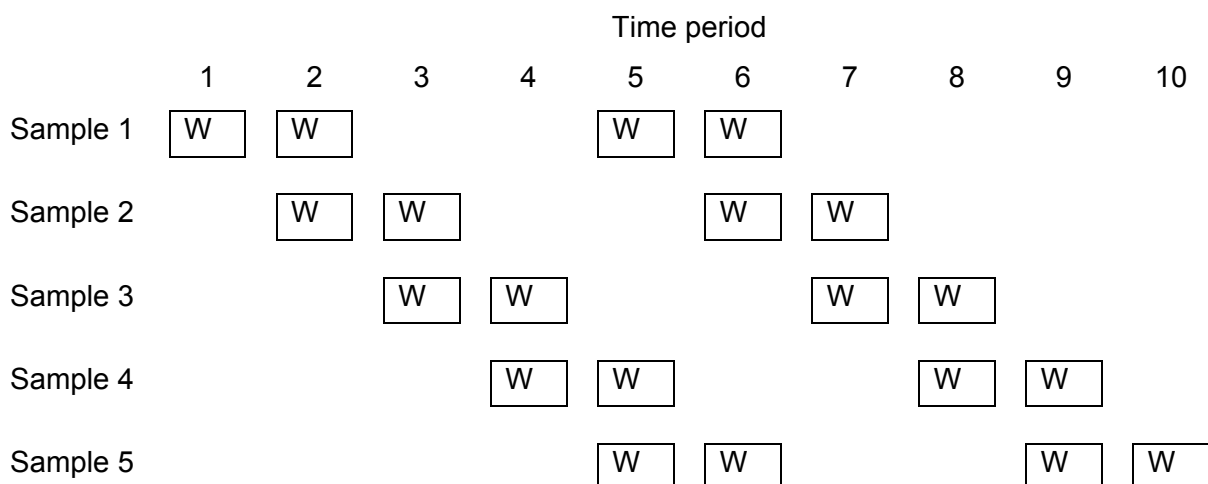


Figure 7: Rotating Panel Design: 1-1-0-0-1-1 Rotation Pattern

2.5 Split panel

This involves a combination of cross-sectional and panel samples at each fieldwork occasion. A common design described by Kish (1987, p.181-183) involves one fixed panel sample from which data are collected on each occasion, plus a supplemental cross-section

sample on each occasion (illustrated in Figure 8). A series of cross-sectional surveys in which a proportion of sample elements are deliberately retained in the sample for consecutive surveys – referred to by Kalton and Citro (1993) as an *overlapping survey* – can also be thought of as a type of split panel. The usual reason for considering such designs is a desire to make both cross-sectional and longitudinal estimates, accompanied with the concerns outlined above about the ability of longitudinal samples to provide adequate cross-sectional estimates.

A variation is a series of repeated cross-sectional surveys, but with a $P=1$ stratum (units in the stratum are included in the survey sample with certainty). With such a design, a subset of the sample units will in practice be included at every occasion. In this case the objective is a series of cross-sectional estimates, but longitudinal estimates are serendipitously possible for the $P=1$ stratum. Several statutory business surveys have this characteristic, with the largest businesses (in terms of, say, turnover or number of employees) forming the $P=1$ stratum.

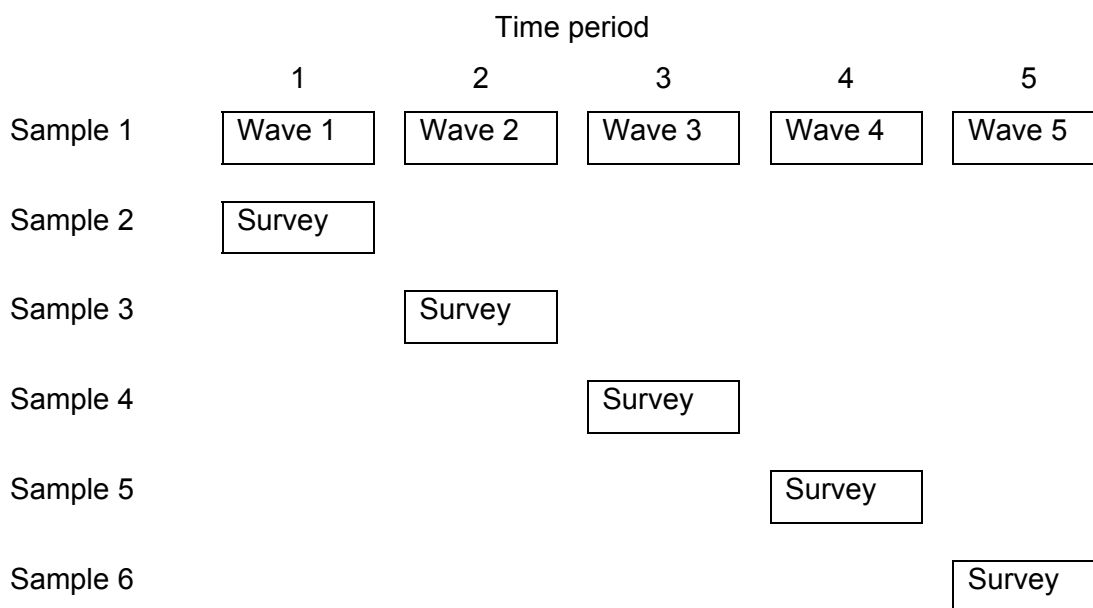


Figure 8: Split Panel Design

2.6 Differences between designs

The five types of longitudinal survey designs outlined in sections 2.1 to 2.5 above are of course only a broad typology. This classification does not describe the full range of possible designs. For example, each panel in a repeated panel design may or may not include additional regular samples of births. And with repeated panels, rotating panels and split panels a wide variety of patterns are possible. The most appropriate variant of a

longitudinal design will depend on the analysis and estimation objectives, the nature of the data to be collected, and practical constraints.

One important consideration is whether the survey operation should provide cross-sectional as well as longitudinal estimates. In principle, a fixed panel plus births can achieve this, but in many situations adding regular samples of births to a panel is a very complex and expensive task. A more efficient way to achieve cross-sectional representativeness can be to select a fresh cross-sectional sample. A split panel design may be used in this case or, if the gross change of interest need be observed only over relatively short periods, a rotating panel design.

A second consideration is the prime unit of analysis. For much longitudinal analysis, the unit of analysis is an *event* of some kind, rather than the person, household or business that experiences the event. For example, in an analysis of labour market transitions the unit of analysis might be a transition between two activity statuses (e.g. unemployed to employed). The sample should therefore be representative of all such events during a specified period. This can only be achieved by the sample being representative of all *persons* who have experienced such events, given that the data are collected from persons. This puts a premium on the inclusion in the sample frame of all persons who may have experienced such events, even if their exposure to the possibility of an event lasts only a short period of time. In other words, it is desirable to continually sample all births into the population of “at-risk” persons. The conclusion would be different if the main analysis aim is to estimate, say, the characteristics of persons who experience particular patterns of events or periods in certain states over a particular time frame. Here the premium might be on making observations on a sample of persons who are “at-risk” over the whole period.

2.7 Representing “births”

An important aim of many longitudinal survey designs is to ensure that “births” to the population are adequately represented in the sample. In this section, we highlight some important considerations in sampling births.

The best way to ensure that all births to a population have a chance of being selected for the survey is to sample from the *flow* of births. This will either mean sampling continuously in time or sampling from a permanent register of new entrants. For example, in the case of (legal) immigrants to a country, we could either continuously sample people entering the country at airports, seaports and other borders, screening them to identify

immigrants, or we could sample from official registers of immigrants. But often it is not possible to sample continuously and no permanent register exists. In that case, the best available approach might be to select a fresh cross-sectional sample at regular intervals, screening it to identify new entrants since the time of the previous sample. With this approach, it is important that each sample provides good coverage of recent entrants. It is also important to be aware of the possibility of new entrants who have exited the population again by the time the next sample is selected. We address each of these two issues in the next two sub-sections.

Coverage of new entrants

Often, a sampling frame that will provide acceptable coverage of the total (stock) population will be inadequate as a frame of new entrant (flow) population. This could be because the mechanism by which new entrants get added to the frame has a time delay, so it is not up to date. Or it could be because new entrants are simply more likely to have the kind of characteristics that are associated with absence from the frame. Thus, simply selecting a new sample from the same frame that was used to select the original wave 1 sample may not be adequate. Instead, this may need to be augmented by special procedures. If coverage of new entrants is poor, this could lead to coverage errors (see section 1.3).

Coverage of short-term entrants

Some units may exit from the study population a relatively short time after entering. Examples include new businesses that either close or are bought up by a larger business, or persons who, soon after attaining the age of 18, emigrate (in the case of a survey of residents aged 18 or over). When sampling recent births by identifying them within a new cross-sectional sample, only those births that have remained in the population until the time when the new sample is selected will have a chance of selection. Others will be omitted, again potentially leading to a coverage error. The extent of under-coverage will of course depend on the prevalence of these “short-term entrants” and on the interval between occasions when a new sample is drawn. This is illustrated in Figure 9.

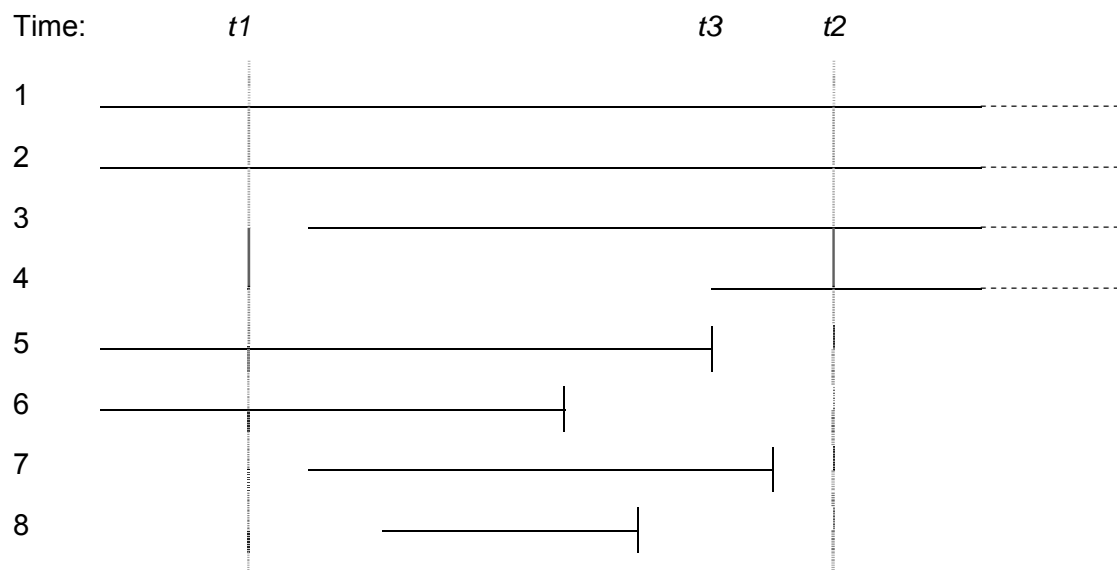


Figure 9: Sampling New Births

Each horizontal line in Figure 9 represents a unit in the population, the extent of the line indicating the period of time for which the unit remains in the population. The dotted vertical lines labelled t_1 and t_2 correspond to the points in time at which the initial sample (t_1) and the first extra sample of new births (t_2) are selected. A dashed line indicates population membership continuing into the future. Units 1, 2, 5 and 6 are all in the t_1 sample frame. Units 1, 2, 3 and 4 are all in the t_2 sample frame though, if selected, units 1 and 2 would be *screened out* as ineligible. Units 3 and 4 have entered the population since t_1 and are thus eligible for inclusion in the sample of recent births to be added to the survey at t_2 . However, units 7 and 8 are not included in the sampling frame at either t_1 or t_2 as they entered the population after t_1 and left it before t_2 . If an analytical aim is to study the characteristics of certain transitions taking place during the period, then if units 7 and 8 experienced any such transitions, the under-coverage may lead to error. One simple way to minimise the likely extent of such under-coverage is to increase the frequency with which new samples are added to the survey. For example, if the first extra sample of new births were selected not at time t_2 , but rather at time t_3 , then unit 7 would be on the sample frame and unit 8 would be the only unit with a zero probability of selection.

3. NON-RESPONSE AND ATTRITION IN LONGITUDINAL SURVEYS

Non-response is the failure to obtain complete, useable, measurements on all sample units in a survey (Groves et al, 2002). *Unit non-response* refers to a complete failure to obtain any measurements for particular sample units; *item non-response* refers to cases where measurements are obtained for some, but not all, items for particular sample units. This section is concerned with unit non-response. Item non-response is addressed in section 5, but is dealt with more fully by Särndal and Lundström (2005).

In the case of a longitudinal survey, non-response can occur at each wave of data collection. Its effects will tend to be cumulative, in the sense that the number of sample units lost to non-response at each wave typically exceeds the number (if any) who return to the survey having been non-respondents previously. Thus, non-response has an effect of continually eroding the sample, making the responding sample smaller and smaller over time. For this reason, this process – unique to longitudinal surveys - is often referred to as *sample attrition*. In this section, we outline what causes non-response and attrition on longitudinal surveys, what effects this can have on survey estimates, and what we can do to reduce the extent and effect of non-response and attrition. Sub-section 3.1 summarises the reasons why non-response occurs in general. Sub-sections 3.2 and 3.3 then discuss issues that are specific to longitudinal surveys.

3.1 Components of non-response

Unit non-response can occur for a number of reasons. These can be classified as follows, corresponding to steps in the process of attempting to gain a response from a sample unit:

- Failure of the data collector to locate/identify the sample unit;
- Failure to make contact with the sample unit;
- Refusal of the sample unit to participate;
- Inability of the sample unit to participate (e.g. ill health, absence, etc);
- Inability of the data collector and sample unit to communicate (e.g. language barriers);
- Accidental loss of the data/ questionnaire.

Often, these are further grouped into three broad categories for the purposes of discussion and presentation of survey outcomes, namely “non-contact” (the first two categories), “refusal” (the third) and “other reasons for non-response” (the fourth, fifth and sixth). The reasons

why each of these types of outcomes might arise are many and varied, depending on the particular characteristics of a survey. Here we briefly summarise the main determinants of the level of each type of non-response. For further discussion, see Lynn (2006).

Reasons for non-contact

Failure to locate or to identify a sample unit is often related to inadequacy of information on the sampling frame such as name, address or telephone number details, which may be out-of-date or otherwise incorrect. Failure to make contact with an identified sample member is an outcome that primarily applies to interview surveys (either face-to-face or by telephone). In the case of self-completion surveys, a failure to receive the questionnaire is usually due to a failure to correctly locate the sample unit, i.e. a paper questionnaire is mailed to the wrong address, or an email request to complete a web questionnaire is sent to the wrong email address.

In interview surveys, failure to make contact is the result of the interaction between:

- a) the timing and number of attempts made by the interviewer to make contact, and
- b) the times at which the sample member is available to be contacted.

In the case of surveys of the general household population, several research studies have examined interviewers' calling patterns and how these relate to survey outcomes or at-home patterns of sample households. This has been done for face-to-face surveys (Swires-Hennessy and Drake, 1992; Campanelli *et al*, 1997; Phillipens and Billiet, 2004) and for telephone surveys (Kulka and Weeks, 1988; Bennett and Steel, 2000). Some general guidance can be summarised as follows.

For face-to-face surveys of households:

- Interviewers should make a minimum of 7 visits to each address before accepting an outcome of "non-contact" if the overall non-contact rate is to be as low as 4% (findings from UK, USA, Netherlands);
- As number of visits increases, the conditional probability of finding someone at home at the next visit decreases;
- Sunday and Monday evenings are the times with the highest probability of finding someone at home, followed by other weekday evenings;
- Probability of contact is higher at weekends than during the daytime on weekdays;

- Contact at a time when it may be inconvenient to interview is still valuable as it can enable a suitable appointment to be made (e.g. contact on Sunday evenings in UK);
- By calling at times when sample members are less likely to be at home, interviewers have to make more visits to achieve a particular response rate.

For telephone surveys:

- A call-scheduling system is needed to ensure that, by the end of the fieldwork period, all sample cases have been attempted a sufficient number of times, with an appropriate spread over times of day and days of the week (and weeks);
- If a number is busy, the best time to try again is 10 - 30 minutes later. If that is not possible, another good time is the same time the next day;
- If there is no reply (business number), the best time to try again is the next day;
- If there is no reply (private number, daytime), the best time to try again is 2 to 6 hours later;
- If there is no reply (private number, evening), the best time to try again is the following evening.

Reasons for refusal

The past decade has seen a considerable volume of research into the reasons why people do, and do not, co-operate with sample surveys. This has led to considerable advances in our understanding of the survey participation process. Understanding refusals is important, as they often constitute a large proportion of survey nonresponse.

A decision about whether or not to co-operate is an outcome of an interaction between interviewer and sample member. The behaviour of both sample member and interviewer during the interaction will be largely influenced by two sets of factors. These can be broadly labelled the “social environment” and the survey design. (Both the interviewer and the sample member will of course also have their own personal characteristics and predispositions upon which these two sets of factors act.) The social environment includes the degree of social cohesion, the legitimacy of institutions, and so on. These influence the degree of social responsibility felt by a sample person and the persuasion strategies and decision making strategies used by interviewers and respondents respectively. Also, the immediate environment in which the survey interview is to take place is likely to affect a

sample member's willingness to be interviewed. Relevant factors include comfort and perceived safety.

Many aspects of survey design affect response rates. Those that are particularly relevant to longitudinal surveys are discussed in section 3.3 below. Other, broad, aspects of survey design can be considered as constraints upon the interaction between sample member and interviewer. Mode of interview is very important. Interviewers are much more limited in the ways they can communicate with a sample member if they are talking on the telephone rather than standing in front of them face-to-face. They cannot show the sample member documents or identity cards, they cannot use body language or gestures, and so on. These limitations may contribute to the lower levels of success that interviewers seem to have in avoiding refusals on telephone surveys. How interviewers introduce the survey is also likely to be influenced by the length and content of the interview. For example, if a sample member seems generally willing but appears not to have much time available currently, then faced with a long interview an interviewer may suggest that she returns at a more convenient time ("retreat and return") rather than asking to start the interview immediately. But if the interview is short, she may be more likely to suggest starting the interview immediately. These tactics may have different implications for the survey outcome.

Ultimately, when faced with a request to take part in a survey a sample member is likely to rapidly (and in most cases, sub-consciously) weigh up the likely benefits (advantages) and drawbacks (disadvantages) of complying (Groves et al, 2000). The challenge for the survey designer is to find ways of emphasising the advantages and de-emphasising the disadvantages. This is complicated by the fact that different sample members will have different views on which potential advantages and disadvantages are relevant to them. The main types of benefits and drawbacks, likely to be relevant to many sample members, are summarised in Table 1.

Benefits	Drawbacks
Answering the questions will be enjoyable (interesting topic)	It will take time that I could usefully spend doing something else
The experience of the interview will be enjoyable (social interaction)	It might make me feel uncomfortable or stressed (sensitive topic)
I will be doing some good (I agree with the stated survey aims and believe that the survey can achieve those aims)	I might be risking my safety (inviting a stranger into my home, giving out personal information)
I will receive some immediate benefit (payment, voucher, gift)	There could be other unwanted consequences (direct marketing, further interviews, law enforcement)

Table 1: Perceived benefits and drawbacks of survey participation

3.2 Special features of non-response on longitudinal surveys

The main types of non-response introduced in section 3.1 apply to all surveys, including longitudinal surveys. However, the reasons why they arise, the impact they might have, and the techniques that can be used to tackle them, can all be rather distinct in the case of longitudinal surveys. In this sub-section we emphasise those unique aspects of non-response in the context of longitudinal surveys.

Non-contact

Subsequent to the first wave of data collection, a major component of non-contact on longitudinal surveys is caused by geographical mobility of sample members. Between waves, a proportion of sample members will move home, change employment, change telephone number or change email address. If the survey organisation is reliant upon any or all of these details to contact the sample member, then they will need to take extra measures in order to be able to make contact at the subsequent wave. They will also need to have interviewers available to attempt contact with sample members in all the places to which sample members could move. (This is not an issue for a telephone, postal or web survey, but is important for a face-to-face interview survey.) For example, surveys of households in the UK with data collection at one-year intervals find that around 10% of persons on average change address between each consecutive pair of survey waves (Laurie et al, 1999).

To overcome this potential addition loss due to non-contact there are several things that the survey organisation can do:

- Collect a range of contact information at the first wave and at each subsequent wave. This could include telephone number (including mobile); name, address, and telephone number of friends or relatives; contact details at place of work; email addresses). This information can then be used when needed to help establish contact at subsequent waves;
- Use a range of ways to keep in touch with sample members between data collection waves. These should include mechanisms by which the sample members can inform the survey organization of address changes. For example, mailings (such as newsletters and birthday cards) can include prepaid return post cards; emails can include a hotlink to a survey website with a simple reply form, and so on;

- Employ special tracing efforts to locate the contact details of sample members who have moved and whose address is not currently known. The nature of these will depend on the circumstances. There may be particular types of official records that could be consulted, for example.

Refusals

In addition to the universal features of refusals described above in section 3.1, there are two special features of importance with longitudinal surveys. The first is that participation requires a considerable commitment on the part of sample members – not just a single interview, but several, over a period of time. This is sometimes referred to as high *respondent burden*. In consequence, special incentives or motivation may be needed to compensate. Typically (but not always), longitudinal surveys offer sample members a small payment for each interview, or some other form of gift, as well as putting particular effort into making the sample member feel like an important, irreplaceable, component of the study and persuading them that the study itself is valuable.

The second special feature of longitudinal surveys relevant to refusals is that, after the first wave, sample members have already experienced the survey interview and therefore have a very good idea of exactly what it consists of, what kinds of questions will be asked, how difficult they find it, and so on. This is very different from a typical survey situation, where a sample member will have only a rather vague and general impression of what they are being asked to do at the time when they are being asked to co-operate. Consequently, on a longitudinal survey it is very important to try to make the interview experience as pleasant as possible for the respondent. If a respondent finds the interview difficult, frustrating, embarrassing, uninteresting or simply too long, they will be less likely to be willing to take part again at the next wave.

One important design question for longitudinal interview surveys is whether it is advantageous, where possible, to assign the same interviewer to a respondent at each wave. In terms of the effect on response rate, evidence is rather equivocal. 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 use data from an experimental design that interpenetrated interviewers and areas on the British Household Panel Survey. 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, then, there is no direct evidence that maintaining interviewer continuity improves response rates. It may be helpful for some sample members and not for others.

Another tactic that can be used to reduce refusal rates is to attempt to “convert” a refusal by a follow-up approach. One way to do this is for the follow-up to be made by telephone directly from the survey office. This can be quite successful (Burton et al, 2006) as the person making the follow-up call is able to ascertain why the sample member refused and can attempt to address those concerns. Often, refusals are prompted by some temporary circumstance (e.g. a busy, stressful, or even traumatic period in the sample member’s life) which may no longer apply if the refusal conversion attempt is made a few weeks later. Or the sample member may simply not have felt comfortable with the interviewer who visited them – for reasons of personality, appearance, gender, etc – which is a good reason for the refusal conversion attempt to be undertaken by a different person. If the sample member can be persuaded to allow an interviewer to visit (a different interviewer if necessary), then an appointment can be made.

A general tactic that can be used on longitudinal surveys to reduce the risk of refusals is to “tailor” the approach to each sample member to suit their particular circumstances and preferences. From wave 2 onwards, a lot of information is known about each sample member prior to each approach for interviewer. This includes both the survey responses and the process data from earlier waves (for example, at what times of day contact attempts were successful or unsuccessful; whether a refusal conversion attempt was needed and what were the reasons for the initial refusal). This information can be used to identify a promising approach at the next wave. Aspects of the approach that could depend upon this prior information include the timing of contact attempts, the wording of any advance letter or other written communication, the nature of any incentive offered, the messages that are emphasised by the interviewer when introducing the survey, and even the mode of data collection.

3.3 Non-response patterns on longitudinal surveys

On a longitudinal survey, unit non-response could potentially occur at each wave of the survey. When this happens for a sample member who has previously responded at at least one wave, this is often referred to as *wave non-response*, to distinguish it from the case where no data at all is obtained for a sample member. Over several waves of a survey, there

are therefore many patterns of wave response that could arise. The number of possible patterns depends on the data collection policy of the survey. Below, we illustrate the patterns that could arise under each of three common policies:

1. Attempt to collect data from all eligible units at every wave;
2. Attempt to collect data at each wave from all eligible units that responded at wave 1. This policy is often adopted when a main aim of the survey is to understand change from the situation at the time of wave 1. The data from subsequent waves are therefore not very valuable if the wave 1 data are missing;
3. Attempt to collect data at each wave from all eligible units that responded at the previous wave. This policy may be adopted where the central aims of the survey require complete data from every wave.

For illustration, assume a 4-wave survey. Then, under policy 1 there are 16 possible response patterns as shown in Figure 10. The shaded boxes represent a response and the blank boxes a non-response. Under policy 2 there are 9 possible response patterns (Figure 11) and under policy 3 there are 5 possible patterns (Figure 12). These response patterns have important implications for how the data can be used in analysis and for non-response adjustment techniques such as weighting, which will be discussed in section 5.

Wave:	1	2	3	4
1	Shaded	Shaded	Shaded	Shaded
2	Shaded	Shaded	Shaded	Blank
3	Shaded	Shaded	Blank	Shaded
4	Shaded	Blank	Shaded	Shaded
5	Blank	Shaded	Shaded	Shaded
6	Shaded	Shaded	Blank	Blank
7	Shaded	Blank	Shaded	Blank
8	Blank	Shaded	Shaded	Blank
9	Shaded	Blank	Blank	Shaded
10	Blank	Shaded	Blank	Shaded
11	Blank	Blank	Shaded	Shaded
12	Shaded	Blank	Blank	Blank
13	Blank	Shaded	Blank	Blank
14	Blank	Blank	Shaded	Blank
15	Blank	Blank	Blank	Shaded
16	Blank	Blank	Blank	Blank

Figure 10: Possible wave response patterns under policy 1: all units at every wave

Wave:	1	2	3	4
1	■	■	■	■
2	■	■	■	□
3	■	■	□	■
4	■	□	■	■
5	■	■	□	□
6	■	□	■	□
7	■	□	□	■
8	■	□	□	□
9	□	□	□	□

Figure 11: Possible wave response patterns under policy 2: wave 1 respondents at every wave

Wave:	1	2	3	4
1	■	■	■	■
2	■	■	■	□
3	■	■	□	□
4	■	□	□	□
5	□	□	□	□

Figure 12: Possible wave response patterns under policy 3: previous wave respondents only

An example of the patterns of response observed over the waves of a longitudinal survey is presented below in Figure 13. This example uses data from cohort 5 of the England and Wales Youth Cohort Study, a postal self-completion survey of persons aged 16-19 in England and Wales (Lynn et al, 1994). Sample members were sent questionnaires at one-year intervals, the first wave being in the autumn of the year that they had completed compulsory schooling. The data collection policy for this survey was to attempt to collect data at each wave from all eligible sample members (policy 1) and the survey had three waves of data collection.

Wave:	1	2	3	No. of cases	% of cases
1	■	■	■	8,396	44.9
2	■	■	□	2,555	13.7
3	■	□	□	2,660	14.2
4	■	□	■	900	4.8
5	□	■	■	386	2.1
6	□	■	□	575	3.1
7	□	□	■	352	1.9
8	□	□	□	2,855	15.3

Figure 13: Wave response patterns on the England and Wales Youth Cohort Study, Cohort 5

Figure 13 illustrates a number of common features of wave response patterns on longitudinal surveys. First, the most prevalent pattern is complete response. In this case, 45% of sample members have responded at all three waves. However, although this is the most prevalent response pattern it does not constitute a majority of the sample. In other contexts, a response rate of 45% might be considered disappointing, even worrying. The second most prevalent response pattern is complete non-response: 15% of sample members do not respond at any wave. This suggests that there is some consistency in the response behaviour of sample members: a majority of sample members either respond at *every* wave or do not respond at *any* wave.

The next most prevalent patterns are the attrition patterns: 14% of sample members respond at each of the first two waves but not the third wave and a further 14% respond only at the first wave. It is fairly common for sample members to respond up until a certain point in the life of a longitudinal survey and then to stop responding.

Response patterns in which sample members respond at a wave having previously not responded on at least one occasion (patterns 4, 5, 6 and 7 in Figure 13) are rather less common. As already suggested, the use that can be made of the data from such respondents is typically limited. These respondents cannot contribute to estimates of change since age 16 (wave 1). In the case of patterns 6 and 7 they cannot contribute to any kind of estimates of change, as they were only observed on one occasion. And it is relatively expensive to collect data for these response patterns. Bear in mind that at each wave three reminder mailings are sent to each sample member who has not yet responded. By definition, this means that all three reminders are sent to each person who ultimately does not respond. Assuming an average of 1.5 reminder mailings per respondent, over three waves a total of 169,463 questionnaires were mailed in order to obtain 36,457 completed questionnaires – an average of 4.65 mailings per completed questionnaire. If the survey had instead adopted the policy of only mailing wave 1 respondents, the total questionnaires mailed would have dropped to 138,667 and the total completed questionnaires would have been 34,758 – an average of 3.99 mailings per questionnaire. And if the survey had adopted the policy of mailing only respondents to the previous wave, 33,858 completed questionnaires would have been received at the cost of only 3.22 mailings each. Relative to this policy, then, mailing all sample members at each wave can be seen to be an expensive way to obtain a relatively small amount of additional data that are in any case of limited use.

It is perhaps also worth considering the reasons for the wave response patterns observed on this survey. In the UK, a sizeable minority of young people leave the parental

home soon after completing compulsory education. It is possible that for many in this subgroup the address obtained from the sampling frame (home address during final year at school) was already out of date by the time of wave 1 of the survey. The questionnaire may therefore not have reached these sample members. More generally, young people are very mobile between ages 16 and 19, often moving to different places to work or study. In particular, most of those who went on to university will have moved to a different place (very few university students continue living with their parents while at university in the UK) between waves 2 and 3. At wave 1 (and wave 2) the questionnaire included a space for recording the contact details (name, address and phone number) of relatives or other people who might know the sample members whereabouts in the future in case they move. These were used at the subsequent wave in cases where the mailing was returned by the Post Office or where no reply was received after the first reminder.

Consequently, it is likely that there were many sample members who did not receive the wave 1 questionnaire (“non-contacts”). If this happened, then it is likely that they would not have received the questionnaire at any of the subsequent waves either. On the other hand, amongst sample members who responded at wave 1, there was a high probability of contact at subsequent waves due to the extra contact details collected on the wave 1 questionnaire (though not all respondents supplied these details), so it is likely that much of the non-response at waves 2 and 3 amongst wave 1 respondents (response patterns 2, 3 and 4) was due to refusals rather than non-contacts. The questionnaire was quite similar each year, so it may have seemed repetitive and unnecessary to some sample members. It is difficult to persuade sample members otherwise in the absence of an interviewer.

4. MEASUREMENT ERROR ON LONGITUDINAL SURVEYS

4.1 Introduction to measurement error

A *measurement error* occurs when the value of an item available for analysis (i.e. in the data set) for a particular respondent does not correspond with the ideal value that would have been recorded if the underlying concept of interest had been measured perfectly. We will refer to these two values as the *observation* and the *true value* respectively (while recognizing that often it is quite difficult to determine exactly what the true value is). There are many reasons why the observation could differ from the true value:

- The respondent may not comprehend the question as intended (particularly likely if the question is not well designed);
- The respondent may fail to retrieve from memory (all) the necessary information to answer the question;
- The respondent may make errors of judgment or estimation when converting the retrieved information into an answer;
- The respondent may report the answer inaccurately – either by accident or by deliberate choice;
- In an interview survey, the interviewer may record the answer inaccurately;
- In a pencil-and-paper survey (self-completion or interview), the answer may be recorded illegibly;
- If an answer requires post-interview coding, the wrong code may be applied;
- If the data require post-interview data entry, a keying error may occur;
- Other errors during data editing and processing may affect the datum.

Our interest here is not to discuss why these various forms of error might occur and what can be done to reduce them, in general. Other texts can be consulted on this matter (good starting points would be Groves et al, 2004, Biemer et al, 1991 and Biemer and Lyberg, 2004). Rather, our interest is in how these errors affect survey estimates, particularly in the context of a longitudinal survey, and what specific techniques can be used on longitudinal surveys to minimize the detrimental impacts of measurement errors.

The net effect of all the individual errors in the data is to induce measurement error in survey estimates (see section 1.3). Consider a simple example of estimation of the mean of a continuous variable, such as income. If the measurement errors are random, i.e. they themselves have a mean of zero, then the mean of the observations will equal the mean of the

true values. In other words, measurement error has not introduced any *bias* to the survey estimate – but it will have introduced extra *variance* (see Figure 1). On the other hand, if (some of) the measurement errors are systematic, then the mean of the observations *will* differ from the mean of the true values, so bias has been introduced. This might happen if, for example, persons with very low incomes tend to over-report, perhaps for reasons of social desirability, without a counterbalancing under-report by persons with high incomes. In general, it makes a difference whether measurement errors are *random* (mean zero) or *systematic* (mean not zero).

Now suppose we want to estimate the association between income and occupation. Even if the measurement errors in observed income are random, they will cause the *association* to appear weaker. So random measurement errors in an individual item can cause a systematic error (bias) in a survey estimate. The same can be seen if we consider estimation of the proportion of units in a particular category of a categorical variable. Consider the example of Table 2, where a sample of 1000 persons is divided into three categories of the variable “economic activity status”. Suppose that the category is correctly observed 90% of the time, but that the other 10% of observations are misclassified, being equally likely to be classified as either of the two incorrect categories. Again, the apparently random measurement error has distorted the proportions in each category. We observe 52.6% employed, when the true proportion amongst the sample is 56.0%, etc. A survey-based estimate of the proportion employed would be downwardly biased.

		Observed categories			Total
		Unemployed	Employed	Other	
True categories	Unemployed	144	8	8	160
	Employed	28	504	28	560
	Other	14	14	252	280
Total		186	526	288	1000

Table 2: Measurement Error in a Categorical Variable

Now we will extend this discussion of the impact of measurement error on survey estimates to the kinds of estimates in which we are typically interested in the case of longitudinal surveys.

4.2 Measurement error in estimates of change

Suppose that we want to estimate the proportion of people who change activity status between one wave of a survey and the next and that the observations in Table 2 represent

the first wave. Suppose that Table 3 represents the relationship between the true values at each wave. In other words, this is the *true* wave-to-wave activity status transition matrix.

Now further suppose that the observations at wave 2 are subject to exactly the same kind of measurement error as the observations at wave 1. That is, 90% of values are observed correctly and the other 10% are randomly misclassified to one of the other two categories. Then, the *observed* transition matrix will be as in Table 4.

		Wave 2			Total
		Unemployed	Employed	Other	
Wave 1	Unemployed	60	50	50	160
	Employed	60	400	100	560
	Other	40	40	200	280
Total		160	490	350	1000

Table 3: True transition matrix

		Wave 2			Total
		Unemployed	Employed	Other	
Wave 1	Unemployed	59	67	60	186
	Employed	77	336	113	526
	Other	50	64	174	288
Total		186	467	347	1000

Table 4: Observed transition matrix, with measures subject to measurement error

It can be seen that the observed transition matrix is really quite different from the true one, despite only a relatively modest level of apparently random measurement error. For example, the true proportion who did not change activity status between the two waves is 66.0%, whereas the observed proportion is 56.9%. The true proportion of persons unemployed at wave 1 who were in employment at wave 2 is 31.2%, whereas the observed proportion is 36.0%. And the true proportion of persons classified as “other” at wave 1 who were in employment at wave 2 is 14.3%, whereas the observed proportion is 22.2%. Measurement error has biased estimates of any of these parameters of change.

Often, we are interested not (just) in estimating levels of change but also in estimating the association of other variables with change (or stability). For example, we might like to know the *characteristics* of persons who move from unemployment to employment, and perhaps also to compare them to the characteristics of persons who remain unemployed. In

our example above, there are 50 sample members who truly move from unemployment to employment, but we observe 67. In fact, only 41 of these have truly moved from unemployment to employment ($50 * 0.9 * 0.9$), while the other 26 have been misclassified in the transition matrix. So, we would base our estimates of the characteristics of persons who have made this transition on 67 persons, only 41 of whom should have been included in the analysis, and we will also be omitting nine persons who should have been included. It is fairly unlikely that the net effect of misclassification would be to have no effect on our estimates. It is therefore very important to try to classify sample units correctly in terms of whether or not they have experienced changes or made transitions.

4.3 Seam effects

Often, longitudinal surveys attempt to collect continuous histories – for example of income receipt, economic activity status, marriage and fertility, etc. This is done by collecting data at each wave about the period of time since the previous wave and then putting together all these short-term histories to create one long history. Such data often suffer from inconsistencies at the “seam” between two interviews. For example, a respondent may report at one wave that they are currently unemployed, but at the next wave six months later they report that six months previously they were in a job. This may lead the analyst to infer that they must have started the job very soon after the previous interview. Alternatively, the respondent’s description at each wave of their job is very different, perhaps leading the analyst to infer that there must have been a change of job very soon after the earlier interview. Consequently, 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).

Seam effects therefore arise from combining data from multiple waves when the data are subject to measurement error. The errors are not specific to longitudinal surveys – although their visibility as seam effects is. Consider 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 $2m$, he is asked about his situation in months $m+1$, $m+2$, ..., $2m$. 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 (in month m) for month m does not match the retrospective report from the second interview (in month $2m$) for month $m+1$. Such mismatches may occur for several reasons (see Martini 1989; Young 1989):

1. *omission or under-reporting*. Suppose that each interview asks about receipt of a particular income source and that the respondent has received this income source continuously throughout the whole reference period. They may correctly report receipt in, say, all but one interview, resulting in an apparent movement off the source in month $m+1$, followed by a move back in month $2m+1$;
2. *misclassification or re-definition* of past information. For example, a respondent may 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 a respondent experienced a change since the previous interview but mistakenly recalled that it had taken place longer ago, they might correctly report on their current situation in $2m$, but then report the same situation for all previous months of the recall period until $m+1$, resulting in change being inferred to have taken place at the seam (instead of more recently);
4. *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, or because of ambiguous descriptions or coding errors.

A change at the seam may be *imputed* if the date of a status change is missing. Dates are often imputed as 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 in time to the actual events and therefore likely to be more reliable (Halpin 1998).

4.4 Methods for reducing measurement error in longitudinal surveys

Dependent interviewing

The term *dependent interviewing* refers to structured interviews where the *choice of questions* and/or the *wording of questions* varies across sample members, depending on responses given by the sample member at a previous interview. For example, a longitudinal survey may attempt to update information collected at a previous wave by presenting the

sample member with that information and asking them to confirm whether or not their circumstances have changed (dependent interviewing), rather than simply asking them to state their current circumstances (independent interviewing).

In some respects, this is similar to techniques used commonly in cross-sectional surveys. First, with “routing” or “skipping” (Oppenheim, 1992, ch.6), the *choice of question* to ask next depends upon the answer(s) given to one or more previous questions. Second, the precise *wording of a question* may be adapted depending on the answers to previous questions (e.g., “... your current job ...” for a respondent who has just answered that they are currently in employment, and “... your most recent job ...” for a respondent who answered that they are not currently in employment but have been employed previously). The difference is simply that the information used to determine which question to ask, or the wording of the question, comes from within the same survey interview, whereas in the case of dependent interviewing, the information is known prior to the commencement of the interview. This creates extra challenges for importing the information in appropriate form into the current interview, but it also brings extra opportunities, as the researcher can interrogate the information prior to designing the survey instrument. Also, it is possible to “clean” textual data that is to be used in question wording, to fit the proposed structure of the question better. The process of extracting the data that will be needed during the dependent interview, cleaning or amending them, and providing them to interviewers in an appropriate form, is often referred to as “feeding forward” survey data (Corti and Campanelli, 1992; Jabine, 1990).

With pencil-and-paper interviewing, feeding forward survey data is laborious and prone to error. Consequently, few surveys used dependent interviewing prior to the advent of computer-assisted interviewing (CAI) methods. Dependent interviewing was previously only used when there was very strong evidence that the quality of the resultant data would be significantly improved (Neter and Waksberg, 1964) or the nature of the data to be fed-forward, and the way that it should be used by interviewers, was simple (Holt, 1979). The advent of CAI (both computer-assisted personal interviewing (CAPI) and computer-assisted telephone interviewing (CATI)) greatly facilitated the use of dependent interviewing, as the need for manual transcription was removed, as was the burden on the interviewer to look up the relevant information and take responsibility for amending the question wording appropriately.

There are many possible ways to word and to structure dependent questions, but a key distinction is between *proactive* and *reactive* questioning methods. *Proactive* dependent

interviewing (PDI) is so called because the information from the previous interview is offered proactively as part of the questioning process (Brown *et al*, 1998). For example, on the US Current Population Survey (CPS), respondents are reminded of the company for which they reported working in the previous quarterly interview and asked whether they still work for the same company (Bureau of Labor Statistics and US Census Bureau, 1997). If yes, industry of employment is assumed to be unchanged and the respondent is asked if his or her activities or duties have changed since the previous interview. If the respondent reports no change in activities or duties, then the description of activities and duties given at the previous interview is read out and the respondent is asked to confirm whether this still applies. If yes, occupation is assumed unchanged. Introduction of these dependent questions greatly reduced apparent change (which was believed to have been largely spurious) and also addressed respondent complaints about repetitiveness (Cantor, 1991; Norwood and Tanur, 1994; Polivka and Rothgeb, 1993). Studies on both the US Survey of Income and Program Participation (SIPP) (Hill, 1994) and the British Household Panel Survey (BHPS) (Sala and Lynn, 2004) drew similar conclusions. Both surveys subsequently introduced PDI for questions about occupation and industry. Aside from occupation and industry questions, household composition details are amongst the question types for which PDI is most commonly used (Mathiowetz and McGonagle, 2000).

With *reactive* dependent interviewing (RDI), the information from the previous interview is offered only in reaction to certain responses. For example, RDI is used on the Canadian Survey of Labour and Income Dynamics (SLID) for wage data. If the respondent reports an amount that is either less than the amount reported in the previous interview one year ago, or more than 10% higher, then a box appears on the CAPI screen showing both amounts and instructing the interviewer to query and enter the reason for the difference. This information is used in subsequent data editing (Hale and Michaud, 1995). Other examples of RDI occur on US Government Agricultural Surveys, where farmers are queried about reported changes in crop acreage (Pafford, 1988) and ranchers are queried similarly about changes in number of cattle (Stanley and Safer, 1997).

The main reason for preferring dependent to independent interviewing is to reduce measurement error – particularly where spurious change is believed to be common. There is evidence (Hill, 1994; Lynn *et al*, 2004a; Rips, 2000; Webber, 1994) to support the view that independent questioning will tend to result in over-estimation of change, particularly where response categories involve long lists of similar items or where open-ended answers require subsequent coding to complex frames. On the other hand, as Bates and Okon (2003) suggest, PDI could invite acquiescence bias, causing spurious change merely to be replaced

by spurious stability – though there is no evidence of this (Lynn et al, 2004b). RDI should avoid the possible acquiescence bias, though it may not be as successful as PDI in reducing spurious change (Lynn et al, 2004a; Sala and Lynn, 2004). Other reasons for preferring dependent interviewing include concerns with respondent and interviewer burden and also costs (Jäckle, 2005). If there is considerable stability in true values, PDI has the potential to reduce – significantly in some cases – the number of questions that need to be asked and the number of open-ended answers that need to be recorded by interviewers and subsequently coded. Weinberg (2002) claims that the introduction of dependent interviewing reduced the interview length for SIPP. Jäckle (2005) also presents evidence that dependent interviewing can reduce interview times.

Dependent interviewing example 1

Lynn et al (2004a) report on an experiment in which a sample of over 1,000 persons, previously interviewed about 16 months earlier, are randomly allocated to three treatment groups: independent interviewing (INDI), proactive dependent interviewing (PDI) and reactive dependent interviewing (RDI).

One set of questions concern sources of income. INDI respondents were asked to look in turn at four cards, each of which displayed a list of possible income sources. The first card listed 6 types of pension, the second listed 10 state benefits related to disability or injury, the third listed 9 other state benefits and the fourth listed 8 other miscellaneous income sources, plus a catch-all category, "any other regular payment". The respondent was asked to say whether they had received any of the types of income or payments shown since the time of the previous interview. PDI respondents were first asked, for each source that had been reported in the previous interview as being received currently, "According to our records, when we last interviewed you, on <date>, you were receiving <source>, either yourself or jointly. For which months since then have you received <source>?" Then, they were shown the four cards in turn and asked whether they have received any of the *other* types of income listed. RDI respondents were first asked the standard INDI question using the four cards. Then, for any source that had been reported at the previous interview but not in the current one, the respondent was asked, "Can I just check, according to our records you have in the past received <source>. Have you received <source> at any time since <date>?"

It was found (Table 5) that the level of reporting of a particular source of income, amongst sample members who had reported that source at the previous wave, was

significantly lower with INDI than with at least one of the dependent methods for 7 out of the 8 income sources with a large enough sample size for meaningful analysis. This suggests under-reporting with INDI.

Income source	INDI	RDI	PDI
NI retirement pension	99	100	100
Ex-employer pension	91	100*	100*
Incapacity benefit	71	96*	85
Income support	82	83	98**
Child benefit	68	86**	93***
Working families tax credit	57	68	87*
Housing benefit	78	94**	94**
Council tax benefit	79	94**	95**

Note: Percentages for RDI and PDI are compared separately with the corresponding percentage for INDI using a Pearson χ^2 test on the relevant 2 x 2 table, with a correction for intra-household correlation, implemented in *Stata* using *svytab*. * indicates $0.01 < P \leq 0.05$, ** $0.001 < P \leq 0.01$, *** $P < 0.001$.

Table 5: Income: reporting at wave *t* conditional upon reporting at wave *t*-1

For 6 of these 8 income sources – all state welfare payments – validation data were obtained from government records and used to estimate directly the levels of measurement error in the survey data. In Table 6 we see that dependent interviewing does indeed seem to have reduced measurement error, though the effects are not uniform. The false negative (under-reporting) rate is significantly lower for DI, compared with INDI, for two income sources, while the false positive rate (over-reporting) is not significantly different for any source. The overall measurement error – the difference between the true value as measured by the administrative data and the value observed in the survey – is smaller with RDI for three income sources and smaller with PDI for one source.

Row proportions		True negative	True positive	False negative	False positive	Admin	Survey	Diff
		P_{00}	P_{11}	P_{10}	P_{01}	$P_{1\bullet}$	$P_{\bullet 1}$	$P_{01} - P_{10}$
Retirement Pension	INDI	.698	.298	-	.004	.298	.302	.004
	PDI	.663	.330	.004	.004	.333	.333	.000
	RDI	.664	.321	.011	.004	.332	.325	-.007
Incapacity Benefit	INDI	.882	.057	.057	.004	.115	.061	-.054
	PDI	.897	.058	.042	.004	.100	.062	-.038
	RDI	.869	.073	.040	.018	.113	.091	-.022*
Income Support	INDI	.790	.179	.023	.008	.202	.187	-.015
	PDI	.785	.180	.035	-	.215	.180	-.034
	RDI	.818	.168	.015	-	.183	.168	-.015
Child Benefit	INDI	.767	.172	.050	.012	.221	.183	-.038
	PDI	.774	.192	.008**	.027	.199	.218	.019*
	RDI	.770	.208	.018*	.004	.226	.212	-.014*
Family Credit	INDI	.901	.057	.023	.019	.080	.076	-.004
	PDI	.877	.092	.004*	.027	.096	.119	.023
	RDI	.894	.077	.026	.004	.102	.080	-.022
Housing Benefit	INDI	.767	.179	.038	.015	.218	.195	-.023
	PDI	.644	.264	.061	.031	.326	.295	-.031
	RDI	.668	.274	.029	.029	.303	.303	.000**

Notes: P_{ab} is the proportion of respondents in group ab, where a = 1 if receipt indicated by the administrative data, 0 if not; b = 1 if receipt is indicated by the survey response, 0 if not. The columns headed "Admin" and "Survey" show prevalence rates for receipt estimated from the administrative and survey data respectively. Bases are 262 INDI cases, 261 PDI and 274 RDI. For further details, see Lynn et al (2004b).

Table 6: Income receipt indicators from administrative and survey data

This study also provides indications that sample members were not all equally likely to under-report income sources with INDI. Certain types of people were more likely than others to only report income receipt in response to the DI question. These included relatively young persons, persons not living with a spouse or partner and disabled persons (Table 7). This suggests that measurement error may be correlated with respondent characteristics, which would certainly introduce bias into certain types of survey estimates.

	Independent reporters	Reactive reporters
	%	%
Male	32	33
Born before 1943	55	25**
In paid work	26	33
Retired	47	22**
NI pension recipient (at wave <i>t</i> -1)	48	22**
Children under 12 in household	21	14
Living with a spouse or partner	58	31**
University-level qualification	4	6
General health "excellent" or "good"	52	50
Registered disabled	23	39*
Has lived in h'hold more than 1 year	94	92
Has regular use of a car	43	44
Has mobile phone	52	61
Likes current neighbourhood	90	83
<i>Base</i>	198	36

Note: The analysis is based on the 234 RDI respondents who reported receipt of at least one of the 8 income sources listed in Table 5. Independent reporters are those who always reported those source(s) in response to the independent question; reactive reporters are those who reported at least one of those sources only in response to the reactive follow-up question, having initially failed to identify the source at the independent question. * indicates $0.01 < P \leq 0.05$, ** $0.001 < P \leq 0.01$, *** $P < 0.001$

Table 7: Characteristics of independent and reactive reporters of income sources

Dependent interviewing example 2

In the same experimental study described in example 1, questions were also asked about employment. Specifically, respondents who were currently employed were asked about three characteristics of their employment (occupation, employed status, and whether or not the respondent has managerial or supervisory responsibilities) and three characteristics of the employing organisation (industry, type of organisation, number of employees). For the INDI group, the questions were identical to those asked in the previous interview. The PDI group were presented with the answer they had given in the previous interview and asked if this still applied. If they replied "no", the standard question was then asked. The RDI group were first asked the standard question, but this was followed up with a check question asking the respondent to confirm whether or not this represented a change since last time. For occupation and employer, the check question was asked of *all* RDI respondents, feeding back the answer given last time. For employee status, managerial status, and number of employees the check question was only asked if the answer given did not correspond with the answer given at wave 8. We focus here on the estimation of change

(from the previous interview) in each of these six characteristics, summarising some findings from Sala and Lynn (2004).

It was found that PDI produces lower levels of observed change for occupation, industry and number of employees (Table 8, first three columns of figures). For example, using the UK Standard Occupational Classification (SOC), with occupations classified in 371 groups, it can be seen that with INDI 53% of in-work respondents appear to have changed occupation between the two interviews, whereas with PDI this falls to 24%. 43% of INDI respondents appear to have changed the industry in which they are employed (using 503 classes of industry), falling to 20% with PDI.

This reduction in observed change appears to represent a reduction in measurement error as the effect of PDI is particularly pronounced amongst respondents who have not reported a change in job between survey waves (Table 8, last three columns of figures). Levels of change in employment characteristics amongst INDI respondents who have not reported a change in job remain implausibly high.

Percentage reporting change	All respondents in work at both waves			All respondents in same occupation at both waves		
	INDI	RDI	PDI	INDI	RDI	PDI
<u>Occupation</u>						
Managerial duties (2)	14.7	9.2	8.9	16.1	1.3***	2.2***
Employee/ self-employed	2.7	2.9	1.3	0	2.2	0.9
SOC major group (9)	31.0	28.3	20.1*	23.7	15.2	0***
SOC minor group (76)	46.5	35.5	20.8***	33.0	20.7+	0***
SOC unit group (371)	52.8	41.3+	23.5***	37.1	25.0	0***
<u>Employing Organisation</u>						
Type of organisation (8)	10.1	9.2	6.7	8.0	7.7	4.3
Number of employees (9)	49.6	42.4	32.8**	38.4	37.7	18.9**
SIC sections (17)	26.6	22.0	16.4*	14.7	11.4	0***
SIC divisions (60)	30.9	28.0	17.9*	18.9	15.9	0***
SIC groups (222)	38.8	40.9	20.1***	25.3	29.5	0***
SIC classes (503)	43.2	43.2	20.1***	30.5	33.0	0***
<i>Base</i>	<i>146</i>	<i>138</i>	<i>150</i>	<i>97</i>	<i>92</i>	<i>106</i>

Note: bases for some estimates are slightly smaller than indicated due to item non-response. RDI and PDI are each compared separately with INDI using a one-sided Pearson χ^2 test on the relevant 2x2 table. + indicates $0.06 \geq P > 0.05$; * $0.05 \geq P > 0.01$; ** $0.01 \geq P > 0.001$; *** $0.001 \geq P$.

Table 8: Percentage reporting change in employment characteristics

Sala and Lynn (2004) also examined the demographic characteristics of respondents whose responses were sensitive to the method of questioning. They found that PDI is less likely to make a difference to estimates of change for respondents aged under 36 than for those aged 36 or over. For three levels of Standard Industrial Classification (SIC) – 17, 60 and 222 groups - and for the 76-group level of SOC, a significant reduction in the level of change is observed only for the two older groups. For the 9-group level of SOC, managerial responsibilities, type of organisation and number of employees, a significant reduction is observed for only one of the two older groups. Also, the effect of PDI appears stronger for the most highly qualified respondents. For two variables, managerial responsibilities and 371-group SOC, the same also appears true for RDI. As regards sex, DI seems to be more likely to make a difference for men than for women, though differences are inconsistent. There are three measures for which RDI makes a difference for men only. With PDI, there are two measures for which only men are affected and one for which only women are affected.

		Men	Women	Age			Qualifications		
				18-35	36-50	51+	Low	Medium	High
Managerial responsibilities	PDI				*				
	RDI	*							**
Employee / self-employed	PDI								
	RDI								
SOC unit group (371)	PDI	***	**	*	***	**	*	***	***
	RDI	*			*				**
SOC minor group (76)	PDI	***	**		***	**		**	***
	RDI								
SOC major group (9)	PDI	**			*				*
	RDI								
Type of organisation	PDI				*				
	RDI								
No. of Employees	PDI		**			*	*		*
	RDI								
SIC sections (17)	PDI				*	**			*
	RDI	*							
SIC divisions (60)	PDI	*			**	**	*		*
	RDI			*					
SIC groups (222)	PDI	**	**		**	**	*		*
	RDI								
SIC classes (503)	PDI	***	**		***	***	*	*	**
	RDI								

Notes: for each cell in the table, the proportion of respondents exhibiting a change in the relevant employment characteristic is compared with the INDI group using a one-sided Pearson χ^2 test. 'High' qualifications is defined as at least one 'A' level pass or equivalent; 'Medium' is at least one GCSE, 'O' level, CSE or equivalent pass; 'Low' is no pass at GCSE or equivalent. * indicates $0.05 \geq P > 0.01$; ** $0.01 \geq P > 0.001$; *** $0.001 \geq P$

Table 9: Socio-demographic correlates of the effect of dependent interviewing on estimates of change

In summary, these two examples of the effects of dependent interviewing appear to show that:

- Measurement error (with independent interviewing) is not random and consequently introduces bias to both estimates of levels and estimates of change;
- Dependent interviewing can reduce that bias;
- However, PDI and RDI are not necessarily equally effective (in our examples, only PDI appears effective for change in employment characteristics but both seem equally effective for receipt of income sources);
- The possibility that PDI introduces some (different) errors should not be ignored, though we find no evidence that any such errors are large.

Calendar and related methods

As discussed in section 1.1, one of the strengths of longitudinal surveys is that survey researchers need rely less on the memory and recall capabilities of respondents. Data covering long periods of time can be obtained by asking only about relatively short periods of time in each interview. However, some retrospective recall is still necessary. The interval between survey waves can be anything from one week to five years or more, depending on the nature of the survey. So, even with longitudinal surveys, measurement error that may be caused by respondents' inability to recall events with sufficient accuracy is a concern. Mediating factors that affect respondents' ability to recall events include not only the length of elapsed time since the event, but also the saliency of the event and the occurrence of other events in the interim that may distort the memory of the event in question (Bound et al, 2001; Eisenhower et al, 1991; Mathiowetz and Duncan, 1988; Waksberg and Valliant, 1978).

Researchers have found that respondents' ability to recall events in response to survey questions can be improved with the help of appropriate "cues" or memory aids. A simple form of memory aid is just to allow the respondent time to think about the circumstances surrounding the event in question and other events that may have happened around that time. One way of facilitating these thoughts is to ask some "context" questions. These are survey questions that are asked in order to try to provide some context for the respondent's thought process, even though the answers to the questions might not be needed for analysis purposes. For example, if a survey is to ask questions about all job changes during a particular period of time, these questions could be preceded by some questions on house moves or partnership changes during the same period.

A further refinement of this idea is the use of Event History Calendars (EHCs). (These were previously referred to as Life History Calendars (LHCs).) EHCs are designed to facilitate retrospective recall of factual data by departing from the standard model of asking fixed questions (no wording variation) in a fixed order (no variation in question order) (Belli et al, 2001; Freedman et al, 1988). EHCs are designed to collect time line data for several domains in parallel, using a flexible interviewing approach. This takes advantage of idiosyncratic structures in autobiographical memory. In other words, different people store and recall information from memory in different ways and in consequence a standard interviewing approach will be sub-optimal for some respondents. The EHC is a chart used by the interviewer and respondent together to indicate when events took place. EHCs have also been used successfully with telephone interviewing.

For example, Belli et al (2001) describe an EHC that collects data in seven dimensions: landmark events, residence, household composition, employment, other activity statuses, time away from work, and welfare entitlements. EHCs can be used to collect data about a wide range of dimensions, depending on the topic of the survey.

Figure 14 presents an example EHC for data on five dimensions: location of residence, household composition, education, employment and welfare entitlements. In this case, data are recorded in months and the reference period is three years. The years and months are pre-printed at the top of the page and the months are also repeated two-thirds of the way down the page for ease of reference. Within each dimension, data relate to periods spent in a state, each period being demarcated by an event. For example, dimension 3 is the location of residence. The respondent lived in London SE3 at the start of the reference period and until July 2003. This period ended with the event of moving to Colchester, where the respondent has lived since then. Events (and the start/end of the reference period) are marked by "X" and the intervening periods in a particular state are indicated by lines.

The utility of the EHC design as a memory aid can be seen if we look at dimension 6, employment. The respondent was in full-time employment during the time she lived in London. This ended in the same month that she moved home, July 2003, followed by a month without employment. She then started part-time work and increased her hours until, in January 2004 she was again working full-time. The full-time employment ended in September 2004 and the following month a child appears in the household. What has happened is that the respondent had a child in October 2004. This event and the house move in July 2003 are probably very salient events that the respondent can remember (and date) well. But they are also related to her employment experiences – and welfare receipt.

So, by first indicating location and household composition on the EHC, this serves as a useful framework for recalling the dates of changes in employment and in welfare receipt. It seems quite plausible that the reported dates for transitions in employment and in welfare receipt are likely to be more accurate than they would have been if they had been asked using a standard questionnaire item, without the immediate – and visual – context of the EHC. Indeed, there is evidence that EHCs improve the quality of retrospective reports – both recalling events and correctly dating them (Axinn et al 1997; Belli et al, 2001; Belli et al, 2004; Caspi et al, 1996).

As well as serving as a memory aid and thus improving the quality of information provided by respondents, EHCs can also serve to facilitate in-interview edit checks. The visual display makes it relatively easy for the interviewer to spot inconsistencies in the answers recorded, such as simultaneous recording of states that should be mutually exclusive (like two categories of “employment” in Figure 14), or missing data for a dimension that should be represented in every time period (like dimension 2, residential location, in Figure 14). The interviewer can then query and resolve these inconsistencies with the respondent before moving on the next section of the interview. This further enhances the quality of the resultant data.

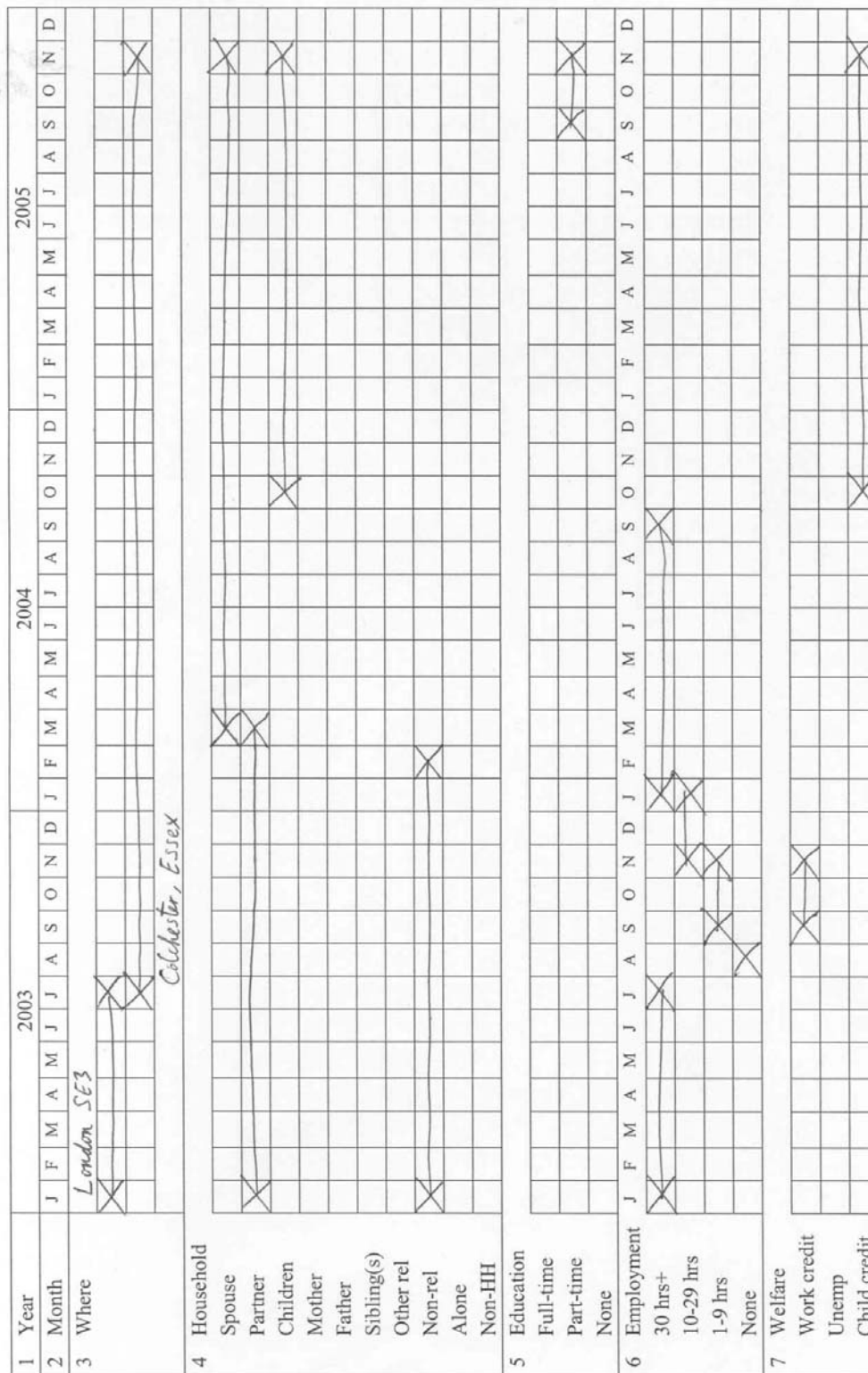


Figure 14: An Event History Calendar

5. WEIGHTING AND IMPUTATION IN LONGITUDINAL SURVEYS

5.1 Introduction to weighting

Weighting involves giving each unit in a survey sample a numeric value (weight), representing the contribution that the unit will make to estimates based upon the survey data (Lynn, 2004). The weights are designed to make the sample representative of the study population. The weight for any particular responding unit can be interpreted as the relative number of population units that it represents. The calculation and application of weights is part of the process of statistical inference, by which conclusions can be drawn about a population of interest based upon knowledge of a sample drawn from that population.

There are four main reasons for weighting: to correct for differences in coverage rates, selection probabilities and (non-)response probabilities, and to correct for the effects of random sampling variance. With respect to the sources of survey error presented in Figure 1, these correspond to the objectives of minimising coverage bias, sampling bias, non-response bias and sampling variance. Whichever is the source of error, the basic objective of weighting is to ensure that groups of sample units are represented in the same proportions in which they appear in the population. These considerations apply to any survey. The features of weighting that are distinct in the case of longitudinal surveys primarily relate to definition of the study population and dealing with complex patterns of non-response. For some types of longitudinal survey, special methods may also be needed to estimate selection probabilities. In this section, we focus on these distinct features of longitudinal surveys.

5.2 Longitudinal weights

Many longitudinal surveys provide data users with two types of weights – “cross-sectional weights” and “longitudinal weights”. The distinction reflects a difference in the population to be represented, which in turn is related to different estimation objectives. For illustration, consider a simple 2-wave survey with data collected at time points $t1$ and $t2$. Analysts could use the data in one of three ways:

1. To make longitudinal estimates, using the data from both $t1$ and $t2$;
2. To make cross-sectional estimates, using only the data from $t1$;
3. To make cross-sectional estimates, using only the data from $t2$.

These three types of analysis are distinguished *both* in terms of which respondent units contribute to the analysis *and* in terms of the population to which the estimates refer as summarised in Table 10. For each type of analysis, weights can be constructed in any of the usual ways, which typically involve dividing the population in groups and calculating the weight for each group as the ratio of number of population units to number of sample units. Thus, for each type of analysis the numerators (population sizes or proportions) and the denominators (responding sample sizes or proportions) of the weights will be different, so three different sets of weights are needed. The weights for type 1 analysis are referred to as longitudinal weights, while those for type 2 and type 3 analysis are both types of cross-sectional weights.

Type of analysis	Study population	Respondent units
1	In population at both $t1$ and $t2$ (N_{12})	Responding at both $t1$ and $t2$ (n_{12})
2	In population at $t1$ (N_1)	Responding at $t1$ (n_1)
3	In population at $t2$ (N_2)	Responding at $t2$ (n_2)

Table 10: Study populations and samples

Note that the size of the population for type 1 analysis (N_{12}) is necessarily no larger than either N_1 or N_2 . If there are any births and deaths in the population between $t1$ and $t2$, which is usually the case, then $N_{12} < N_1$ and $N_{12} < N_2$. If births exceed deaths, then $N_1 < N_2$ and *vice versa*. Equivalent relationships hold for the sample sizes. A common problem for longitudinal surveys is to find a good way to estimate the population distribution, across relevant variables, for each of the study populations. Cross-sectional population estimates may well be available from external sources, but that is rarely the case for longitudinal populations. Typically, a model-based approach is used to estimate the characteristics of the longitudinal populations, taking the wave 1 cross-sectional population as the starting point and then using sample-based estimates of births and deaths to model the changes in the population structure. This is tricky as the sample-based estimates are likely to be subject to non-response error, which is one of the error sources for which we would like the weights to correct.

5.3 Which combinations of waves?

In the simple case of Table 10, it would be possible to create three sets of weights and make them available to data users. But in general, there are $2^t - 1$ possible populations

that can be represented by a t -wave longitudinal survey, of which t are cross-sectional populations and $2^t - (t + 1)$ are longitudinal populations. Consequently, there are potentially t sets of cross-sectional weights and $2^t - (t + 1)$ sets of longitudinal weights that could be created. For surveys with more than 2 or 3 waves, it would not be feasible to create all these sets of weights. For example, if $t=10$, then $2^t - 1 = 1,023$. It could also be confusing to users to have so many sets of weights available. And it is probably not necessary anyway, as many of the sets of weights would be so similar to one another that the choice between them would make no practical difference to any estimates.

One solution to this problem would be to provide data users with all the data necessary to calculate weights for any combination of waves – and also some guidance or even a program that will calculate the weights. Then, each user can specify the set of waves relevant to his or her analysis and produce tailor-made weights. However, this is rarely done, either because some of the necessary data can not be released at the unit level or because users much prefer to be provided with ready-to-use weights.

A practical alternative is for the data provider to produce weights for a limited subset of the possible combinations of waves. This should be accompanied by guidance to users on what to do if the combination in which they are interested is not one of those for which weights are provided. The choice of wave combinations should be guided by the (likely) main uses of the data. For example, if the main objective of the survey is to permit analysis of change relative to baseline data that were collected at wave 1, then there is very little point in producing weights for combinations of waves that do not include wave 1. If a module of questions on a particular topic is included only at waves 1, 4, 7 and 10, then that particular combination should be a strong candidate for weighting. For almost all longitudinal surveys, the complete set of waves should be one of the combinations for which weights are produced. The only exception would be if, by design, there are no units that were eligible for data collection at every wave.

However, it is important to be aware that carrying out analysis based on respondents to a particular set of waves using weights designed for a different set of waves is sub-optimal. Consider a 3-wave survey as in Figure 13 (section 3.3) and suppose that only one set of longitudinal weights is provided, designed to make the set of persons who responded to all three waves representative of the 3-wave longitudinal population. Suppose we want to estimate some parameter of change between wave 1 and wave 3, for which we only need to use data collected at waves 1 and 3. We could use all units with response patterns 1 (XXX)

or 4 (X0X). But the longitudinal weights will be set to 0 for units with response pattern 4. No longitudinal weight is defined for these units. In consequence, 10% of available cases (900 out of 9,296) will be dropped from the analysis because of the unavailability of an appropriate weight. For this estimation, it would have been better to produce a set of weights to represent the X?X population (units that are in the population at the times of both waves 1 and 3 regardless of whether or not they are also in the population at the time of wave 2). These would be non-zero for all units in the XXX and X0X samples.

Another important consideration is that sets of weights are usually produced at several different points in time during the life of a longitudinal survey. Often, this is done after each new wave of data is available as analysts will want to analyse the latest data without waiting until the next wave is completed. Thus, as a minimum, at each wave a set of weights will be produced representing the longitudinal population at all waves to date. This means that ultimately weights will be available for every “attrition sample”. For example, after five waves there will be weights for the X0000, XX000, XXX00, XXXX0 and XXXXX samples. If the survey has a policy of attempting to collect data only from previous wave respondents (policy 3 in section 3.3) this will be all the weights that are needed. Otherwise, the task is to identify which other combinations of waves are sufficiently important to warrant the calculation of weights.

5.4 Which variables?

Once the researcher has identified relevant longitudinal populations for which weights are to be produced, it remains to identify a *method* of calculating weights and a set of *auxiliary variables* that will define the weighting classes and weights. The criteria for both the method and the variables are no different from those for any other kind of survey. Essentially (Lynn, 2004), the objective is to choose a method and a set of variables such that when the method is used to create a set of classes defined by the variables, the resulting classes have the following properties:

- Inclusion propensities (coverage rates, selection probabilities, response probabilities) vary over the classes;
- Values of key sample statistics (e.g. means, proportions, regression coefficients, etc) vary over the classes;
- Values of key sample statistics are similar for included and excluded units (sampled and not sampled, respondents and non-respondents) within each class.

It is often appropriate to calculate weights for each source of survey error separately (e.g. design weights for selection probabilities, post-stratification weights for sampling variance and non-response weights for response propensity). The criteria apply to each stage.

On a longitudinal survey, we should bear in mind that the key sample statistics will tend to be measures of change and measures of association of other variables with measures of change. This is likely to have important implications for the creation of weighting classes. The auxiliary variables that correlate most strongly with these measures of change may well be survey variables from previous waves – and particularly (often) measures of change in previous periods. For this reason, non-response weighting for longitudinal surveys is often done sequentially. For non-response at wave 1, it is necessary to use data external to the survey as auxiliary data. But from that point on (unless the survey includes responses at later waves from units that did not respond at wave 1) response propensity at subsequent waves can be estimated conditional upon response at wave 1 (or other previous waves). In its simplest form, non-response (NR) weights for the attrition samples could be calculated as follows:

- Weights for wave 1 NR (w_1) use auxiliary data external to survey;
- Weights for wave 2 NR conditional upon wave 1 response ($w_{2|1}$) use wave 1 data as auxiliary data. The weight for wave 2 NR is $w_2 = w_1 \times w_{2|1}$;
- Weights for wave 3 NR conditional upon wave 2 response ($w_{3|2}$) use wave 1 and 2 data as auxiliary data (perhaps including measures of change between waves 1 and 2). The weight for wave 3 NR is $w_3 = w_2 \times w_{3|2}$;
- Etc...

This simple form would obviously have to be amended if units with wave non-response patterns were also to be included in the analyses.

5.5 Introduction to imputation

Analysts of any kind of data are typically faced with item non-response and must find ways to deal with it. They may choose complete case analysis or, better, available case analysis, but these are inefficient and wasteful of data. If variables are categorical then item non-response may be treated as a separate substantive category in the analysis. Weighting

could be used to adjust for the non-response, as discussed above, but the available sample of respondent units would be different for each estimate, requiring weights to be calculated each time. This would be very time-consuming (one of the attractions of weighting as an adjustment method is that once the weights are calculated they can be used for all analyses).

A popular method for dealing with item missing data is *imputation*. This involves assigning (imputing) a value wherever an item is missing in the data set. The appeal of imputation is that it results in a completed data set, so standard analysis methods can be used and the same set of units will contribute to any estimate for the same study population. (With available case analysis, sample bases can change from one estimate to the next creating inconsistencies between estimates.) In practice, the standard methods should be adjusted to allow for the fact that some of the data values are not in fact observed values but rather imputed values.

There are many ways of choosing the value to impute. Classes of imputation methods include deductive and rule-based methods, imputation of the mean or mode, imputation of the mean or mode within a class, random imputation, hot deck imputation, distance function matching and regression imputation. As with weighting, the choice of imputation method is only one part of the story. The other important choice is the choice of auxiliary data and how they should be used.

5.6 Longitudinal imputation

In a longitudinal survey context, a notable feature of imputation is often that the auxiliary data can include – or even be restricted to – previous period values of the variable for which an imputed value is being sought. In a cross-sectional survey, income may have to be imputed on the basis of variables such as economic activity status, occupation if employed, sex, age and household composition. But in a longitudinal survey, income can be imputed on the basis of income at each previous wave. Often, income at the previous wave will be a much better predictor of income at the current wave than any set of current wave measures. As with weighting, then, auxiliary data for imputation can include survey measures from previous waves and, indeed, this is often advantageous.

5.7 Revision of imputations

Having just completed, say, wave 2 of a longitudinal survey, income reported at wave 1 may well be the best available predictor of income at wave 2 and therefore the best choice

of auxiliary variable for income imputation. But later, when wave 3 data are available, it may turn out that income is reported at both waves 1 and 3 for some respondents for whom income was missing (and therefore imputed) at wave 2. For these respondents, use of the wave 1 and 3 data in combination may lead to a better (and different) imputation for wave 2 than use of wave 1 data alone. But by this time analysts will have already been using the wave 2 data with the imputations provided at the time of wave 2. The data provider is faced with a dilemma. The options are:

- Provide revised (hopefully better) imputations to replace the ones provided previously, on the basis that the best possible data should always be provided even if this means that there will be inconsistencies between analyses carried out at different points in time;
- Do not revise any imputations already provided in order to avoid inconsistencies between analyses. If this strategy is adopted, the advantages of using subsequent wave data as auxiliary variables for imputation may lead to the conclusion that imputations should not be made until the subsequent wave: e.g. imputations for wave 2 missing values will not be made available until the release of the wave 3 data. However, this can still lead to inconsistent estimates.
- Provide revised imputations and also continue to provide the original imputations, in order that analysts can check the sensitivity of their results to the imputation procedures.

The issue of whether or not to revise imputations is an important one for longitudinal surveys. Different surveys have adopted different policies and it is difficult to be prescriptive as the best policy will depend on the characteristics of the particular survey in question. However, it is generally felt to be important that the policy should be decided in consultation with users and at the earliest stage possible.

REFERENCES

- Axinn, W.G., Barber, J.S. & Ghimire, D.J. (1997) The neighbourhood history calendar: a data collection method designed for dynamic multilevel modelling, *Sociological Methodology* **27**, 355-392.
- Belli, R.F., Shay, W.L. and Stafford, F.P. (2001) Event History Calendars and Question List Surveys, *Public Opinion Quarterly*, Vol. 65, 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, 185-218.
- Bennett, D.J. & Steel, D. (2000) An evaluation of a large-scale CATI household survey using random digit dialling, *Australian and New Zealand Journal of Statistics* **42**, 255-270.
- Biemer, P.P. & Lyberg, L.E. (2003) *Introduction to Survey Quality*, New York: John Wiley & Sons.
- Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. & Sudman, S. (ed.s) (1991) *Measurement Errors in Surveys*, New York: John Wiley & Sons.
- Binder, D. (1998) Longitudinal Surveys: why are these surveys different from all other surveys? *Survey Methodology*, **24:2**, 101-108
- Bound, J., and Krueger, A.B., "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" in *Journal of Labor Economics*, Vol. 9(1), 1991, pp. 1-24.
- Brown, A., Hale, A. and Michaud, S. (1998) 'Use of Computer Assisted Interviewing in Longitudinal Surveys', in M. P. Couper, *et al.* (eds) *Computer Assisted Survey Information Collection*, New York: John Wiley and Sons.
- Bureau of Labor Statistics and Bureau of the Census (1997). *CPS Questionnaire*. Available at <http://www.bls.census.gov/cps/bqgestair.htm>.
- Burkhead, D. and Coder, J. (1985) 'Gross Changes in Income Reciprocity from the Survey of Income and Program Participation', *Proceedings of the Social Statistics Section*, Washington DC, American Statistical Association, 351-356.
- Burton, J., Laurie, H. & Lynn, P. (2006, forthcoming) The long-term effectiveness of refusal conversion procedures on longitudinal surveys, *Journal of the Royal Statistical Society (Series A): Statistics in Society*, **169:3**
- 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
- Cantor, D. (1991). Draft recommendations on dependent interviewing, *unpublished manuscript*, Rockville MD, Westat.
- Caspi, A., Moffitt, T.E., Thornton, A., Freedman, D., Amell, J.W., Harrington, H., Smeijers, J. & Silva, P.A. (1996) The life history calendar: a research and clinical assessment method for collecting retrospective event-history data, *International Journal of Methods in Psychiatric Research* **6**, 101-114
- Corti, L. and Campanelli, P. (1992). The utility of feeding forward earlier wave data for panel studies, pp. 109-118 in Westlake, A. et al (ed.s) *Survey and Statistical Computing*, North Holland, Elsevier.
- Dillman, D.A. (2000) *Mail and internet surveys: the tailored design method (2nd ed.)*. New York: Wiley.
- Dillman, D.A., Eltinge, J.L., Groves, R.M. & Little, R.J.A. (2002) Survey nonresponse in design, data collection, and analysis. In: R.M. Groves, D.A. Dillman, J.L.Eltinge & R.J.A.Little (ed.s), *Survey nonresponse*. New York: Wiley.
- Eisenhower, D., Mathiowetz, N.A., Morganstein, D. (1991) Recall Error: Sources and Bias Reduction Techniques, in Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (eds.), *Measurement Errors in Surveys*, New York: Wiley.
- Freedman, D., Thornton, A., Camburn, D., Alwin, D. & Young-DeMarco, L. (1988) The life history calendar: a technique for collecting retrospective data, *Sociological Methodology* **18**, 37-68.
- Groves, R. M. (1989) *Survey Errors and Survey Costs*, New York: John Wiley & Sons.
- Groves, R.M. & Couper, M.P. (1998) *Nonresponse in Household Interview Surveys*. New York: John Wiley.
- Groves, R.M., Dillman, D.A., Eltinge, J.L. & Little, R.J.A. (2002) *Survey Nonresponse*, New York: John Wiley & Sons.
- Groves, R.M., Singer, E. and Corning, A. (2000) Leverage-saliency theory of survey participation: description and an illustration, *Public Opinion Quarterly* **64**, 299-308
- Groves, R.M., Fowler, F.J.Jr., Couper, M.P., Lepkowski, J.M., Singer, E. & Tourangeau, R. (2004) *Survey Methodology*, New York: John Wiley & Sons.
- Halpin, B. (1998) 'Unified BHPS Work-Life Histories: Combining Multiple Sources Into a User-Friendly Format', *Bulletin de Méthodologie Sociologique*, 60(Oct.): 34-79.
- Hill, D. H. (1987) 'Response Errors Around the Seam: Analysis of Change in a Panel with Overlapping Reference Periods', *Proceedings of the Survey Research Methods Section*, Washington DC, American Statistical Association.

Holt, M. (1979). The use of summaries of previously reported interview data in the National medical Care Expenditure Survey: a comparison of questionnaire and summary data for medical provider visits, in Sudman, S. (ed.) *Health Survey Research Methods*, Washington DC, US Department of Health and Human Services.

Jabine, T.B. (1990). *SIPP Quality Profile*, Washington DC, US Department of Commerce, US Bureau of the Census.

Jäckle, A. (2005) Does Dependent Interviewing Really Increase Efficiency and Reduce Respondent Burden? *Working Paper 2005-11 of the Institute for Social and Economic Research*. Colchester: University of Essex. <http://www.iser.essex.ac.uk/pubs/workpapers/pdf/2005-11.pdf>

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

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.

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.

Kasprzyk, D., Duncan, G., Kalton, G. & Singh, M.P. (ed.s) (1989) *Panel Surveys*, New York: John Wiley & Sons.

Kish, L. (1987) *Statistical Design for Research*, New York: John Wiley & Sons.

Kulka, R.A. & Weeks, M.F. (1988) Toward the development of optimal calling protocols for telephone surveys: a conditional probabilities approach, *Journal of Official Statistics* **4**, 319-3

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

Lemaitre, G. (1992) 'Dealing with the Seam Problem for the Survey of Labour and Income Dynamics', *SLID Research Paper Series, No. 92-05*, Ottawa: Statistics Canada.

Lynn, P. (2004) Weighting. In: Kempf-Leonard, K. (ed.), *Encyclopedia of Social Measurement*. London: Academic Press.

Lynn, P. (2005) *Outcome Categories and definitions of response rates for panel surveys and other surveys involving multiple data collection events from the same units*. Colchester: ISER, University of Essex.

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

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

- Lynn, P., Jäckle, A., Jenkins, S. and Sala, E. (2004a) The effects of dependent interviewing on responses to questions on income sources, *Working Paper 2004-16 of the Institute for Social and Economic Research*. Colchester: University of Essex.
<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-16.pdf> (also forthcoming in *Journal of Official Statistics*)
- Lynn, P., Jäckle, A., Jenkins, S. and Sala, E. (2004b) The impact of interviewing method on measurement error in panel survey measures of benefit receipt: evidence from a validation study, *Working Paper 2004-28 of the Institute for Social and Economic Research*. Colchester: University of Essex.
<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-28.pdf>
- Lynn, P. J., Nicolaas, G., Beerten, R., Laiho, J., and Martin, J. (2005) *Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys (2nd edition)*. Colchester: ISER, University of Essex.
- Mangione, T.W. (1995) *Mail surveys: improving the quality*. Thousand Oaks, California: Sage.
- Martini, A. (1989) Seam effect, recall bias, and the estimation of labor force transition rates from SIPP, *Proceedings of the Section on Survey Research Methods 1989*, American Statistical Association, 387-392.
- Mathiowetz, N., Duncan, G. (1988) Out of work, out of mind: response errors in retrospective reports of unemployment, *Journal of Business and Economic Statistics*, Vol. 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.
- Moore, J. and Kasprzyk, D. (1984) 'Month-to-Month Reciprocity Turnover in the ISDP', *Proceedings of the Survey Research Methods Section*, Washington DC, American Statistical Association, 726-731.
- Morton-Williams, J. (1993) *Interviewer approaches*. Aldershot: Dartmouth.
- Neter, J. and Waksberg, J. (1964) A study of response error in expenditures data from household interviews, *Journal of the American Statistical Association*, **59**, 18-55.
- Norwood, J.L. and Tanur, J.M. (1994) Measuring unemployment in the nineties, *Public Opinion Quarterly*, **58:2**, 277-294.
- Oppenheim, A.N. (1992) *Questionnaire Design, Interviewing and Attitude Measurement, New Edition*, London, Pinter Publishers.
- Pascale, J. and Mayer, T.S. (2004) Exploring confidentiality issues related to dependent interviewing: preliminary findings, *Journal of Official Statistics*, **20:2**, 357-377.
- Pafford, B. (1988) *The influence of using previous survey data in the 1986 April ISO Grain Stock Survey*, Washington DC, National Agricultural Statistical Service.
- Philippens, M. & Billiet, J. (2004) Monitoring and evaluating nonresponse issues and fieldwork efforts in the European Social Survey. *Proceedings of the European Conference on*

Quality and Methodology in Official Statistics (Q2004), Mainz, May 2004 (CD-ROM). Wiesbaden: Federal Statistical Office Germany.

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

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

Rips, L. (2000) Unraveling the seam effect, *Proceedings of the Section on Survey Research Methods 2000*, American Statistical Association, 465-470.

Rope, D. (1993) Preliminary longitudinal nonresponse research with the CPS and CE. *4th international workshop on household survey nonresponse*, Bath, UK

Sala, E. and Lynn, P. (2004) Measuring change in employment characteristics: the effects of dependent interviewing, *Working Paper 2004-26 of the Institute for Social and Economic Research*. Colchester: University of Essex.
<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-26.pdf>

Särndal, C.-E. & Lundström, S. (2005) *Estimation in surveys with nonresponse*. Chichester: Wiley.

Singer, E., Van Hoewyk, J. & Gebler, N. (1999) The effects of incentives on response rates in interviewer-mediated surveys, *Journal of Official Statistics*, **15**, 217-230.

Swires-Hennessy, E. & Drake, M. (1992) The optimum time at which to conduct interviews, *J. Market Research Society* **34**, 61-72.

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*, University of Strathclyde

Trivellato, U. (1999) Issues in the design and analysis of panel studies: A cursory review, *Quality & Quantity*, **33:3**, 339-352

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

Waterton, J. and Lievesley, D. (1987) Attrition in a panel survey of attitudes, *Journal of Official Statistics* **3:2**, 267-282

Webber, M. (1994) The Survey of Labor and Income Dynamics: Lessons Learned in Testing, *SLID Research Paper 94-07*, Statistics Canada, Ottawa.

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

Young, N. (1989) 'Wave-Seam Effects in the SIPP', *ASA Survey Research Methods Section*, 393-398.