

CHOOSING A LONGITUDINAL SURVEY DESIGN: THE ISSUES

by Nick Buck, John F. Ermisch, and Stephen P. Jenkins

(ESRC Research Centre on Micro-Social Change, University of Essex)

September 1995

ACKNOWLEDGEMENTS

This paper is a revised version of a report commissioned by New Zealand Treasury. The research forms part of the scientific programme of the ESRC Research Centre on Micro-Social Change, supported by the Economic and Social Research Council and the University of Essex. We wish to thank a number of people who gave us advice and assistance of various kinds: George Barker, Richard Burkhauser, Eileen Clucas, John Coder, Brian Easton, Judi Egerton, J Gershuny, Markus Jäntti, Ian Pool, Deidre Ross, Helen Stott, Terry Tostevin, Maryann Webber, and NZT's reviewer. New Zealand Treasury bears no responsibility for the opinions expressed or for any errors or omissions; responsibility lies entirely with the authors.

TABLE OF CONTENTS

0. EXECUTIVE SUMMARY	i
1. INTRODUCTION	1
2. TYPES OF LONGITUDINAL SURVEY INSTRUMENT	2
2.1 The unit of analysis in a longitudinal survey	2
2.2 Retrospective surveys	3
2.3 Panel surveys: overview	5
2.4 Single indefinite life panels: (a) cohort panels	5
2.5 Single indefinite life panel surveys: (b) household panels	6
2.6 Multiple overlapping fixed life panel surveys (rotating panels)	12
2.7 Record linkages	17
2.8 Repeated cross-section surveys	19
3. COMPARISON OF INFORMATION PROVIDED	21
3.1 Analyses of income dynamics	21
3.2 Analyses of transitions between states	21
3.3 Analyses of impacts of policies	26
3.4 Longitudinal data and policy analysis: examples	28
3.5 Conclusions from sections 1-3	32
4. SURVEY IMPLEMENTATION AND RESOURCE REQUIREMENTS	33
4.1 Sampling design	33
4.2 Sample size	34
4.3 Oversampling	35
4.4 Following rules	37
4.5 Instrument design	38
4.6 Issues of timing	40
4.7 Maintaining response rates	41
4.8 Interview mode	42
4.9 Panel conditioning	43
4.10 Measurement error	44
4.11 Data processing and data management issues	44
4.12 Weighting	46
4.13 Implications for survey organisation	47
4.14 Costs relative to cross-sectional surveys	48
4.15 Comparison between single panel and multiple rotating panels	50
5. SUMMARY	52
6. REFERENCES	53

0. EXECUTIVE SUMMARY

The aim of this paper is to analyse the contributions which alternative longitudinal survey designs can make to information about the dynamics of family income, labour market and demographic behaviour, and the related policy issues.

In section 2 we identify three main types of longitudinal survey, and discuss seven real-world examples in detail in order to illustrate their principal features. See Table 1 for a summary:

Table 1. Types of longitudinal survey

TYPE	EXAMPLES DISCUSSED IN SECTION 2
<i>Retrospective</i>	Women and Employment Survey (U.K.) Barriers to Leaving Income Support Survey (U.K.)
<i>Panel</i>	
- Single indefinite life—cohort panel	National Child Development Survey (U.K.)
- Single indefinite life—household panel	Panel Study of Income Dynamics (U.S.) British Household Panel Survey (U.K.)
- Multiple overlapping fixed life ('rotating panel')	Survey of Income and Program Participation (U.S.) Survey of Labour and Income Dynamics (Canada)
<i>Record linkage</i>	Longitudinal Census File (Finland)

In section 3, we examine the information provided by the three longitudinal survey types and compare it with that provided by repeated cross-section surveys. Because the usefulness of data can only be assessed with reference to the analyses for which they are to be used, we focus on three generic types of analysis: income dynamics, transitions between states, and impacts of policies. We also discuss a number of previous studies to illustrate how longitudinal data can be used to analyse topics of policy interest: the persistence of poverty, the effects of welfare benefits on employment choices and on the duration of lone motherhood, and the impact of training programmes on subsequent earnings.

The reviews of Sections 1 and 2 emphasise that the choice of a longitudinal survey design depends on the priorities and goals of potential data users. If information about events occurring with relatively high frequency and short durations, or about individual earnings and income mobility, have the greatest priority, then a panel survey is the best option. However, if the priority is information about lower

frequency events, then it would be sufficient to use a retrospective survey design. There are also circumstances in which greater use of the repeated cross-section surveys would suffice.

If a panel survey is adjudged the best option, there remains an important choice to be made between different sorts of panel study design, in particular between a single indefinite life panel, or a set of multiple overlapping fixed life panels (i.e. a rotating panel)? In section 4, we address the issues underlying such a choice, comparing the relative advantages and disadvantages of the two design types in detail: see Table 2 for a summary. We also discuss issues common to both sorts of survey.

Table 2. Single panel studies vs rotating panel studies

	<i>Single panel</i>	<i>Rotating panels</i>
Advantages	<ul style="list-style-type: none"> - Lower cost - More completed spells - Sufficient duration for data on antecedents and consequences of events 	<ul style="list-style-type: none"> - Better current population representativeness - Possible to investigate impacts of attrition and conditioning - More frequent interviews possible
Disadvantages	<ul style="list-style-type: none"> - Potential attrition bias and small subgroup sizes after attrition - Miss immigrants to population - More sensitive to respondent burden 	<ul style="list-style-type: none"> - Higher cost - Insufficient length to collect information on antecedents and consequences of events - Small number and biased sample of completed spells - Difficult to merge data from separate panels

Our review indicates that the strengths and weaknesses of the two panel types are complementary to a large extent, and we are therefore reluctant to make a choice between them. This issue is best resolved by those seeking the new survey. In order to make their choice they will need to specify precisely the survey's purpose, plus the operational and budget constraints which they face.

1. INTRODUCTION

Our aim is to analyse the contributions which alternative longitudinal survey designs can make to information about the dynamics of family income, labour market and demographic behaviour, and related policy issues. We also give advice about assessing the most cost effective longitudinal survey designs for these purposes. In preparing our evaluation, we have also compared pure longitudinal surveys with repeated cross-sectional surveys since, as we shall show, these can sometimes provide the same information as the more complex longitudinal survey while also having some other advantages.

Choosing the most appropriate survey design requires assessment of the benefits of the different sorts of information provided and the different costs required to derive them. Our report discusses the relevant costs and benefits in the sections which follow. However the ultimate weighing up of these costs and benefits is something which can only be done by those seeking the new survey. It will require them to specify very clearly the purposes the survey is to serve, and to identify their operational and budget constraints.

2. TYPES OF LONGITUDINAL SURVEY INSTRUMENT¹

The essential feature of a longitudinal survey is that it provides repeated observations over time on a set of variables for the set of persons belonging to the survey. Different ways of deriving these repeated observations on the same people distinguish the three main types of longitudinal surveys: retrospective surveys, panel surveys, and record linkages. In conjunction with our descriptions of these, we discuss some real-world examples. (We should stress that our aim with these is not to provide a comprehensive catalogue; rather it is provide some illustrative examples.) But first we address an issue of particular relevance to all longitudinal social surveys, the unit of analysis.

2.1 The unit of analysis in a longitudinal survey

The ‘unit of analysis’ in virtually all longitudinal surveys is an individual person, not the family or household. (This contrasts with cross-sectional social surveys which, depending on their purposes, may use any one of these different units of analysis as their focus.) The reason for the focus on individuals is very simple: it is impossible to define a longitudinal family or household in any rigorous way which would enable the unit to be followed over time. New families and households are continually being created, and existing ones have ever-changing memberships (and may cease to exist). By contrast with this flux, the concept of an ‘individual’ is stable in a longitudinal context. This does not mean that longitudinal surveys cannot tell us about families and households and their dynamics—quite the opposite. But the necessary information is derived from individuals who are related to their family or household context (which changes over time).

A separate but related issue concerns how one targets these individuals and their households: the sampling unit and the unit about which the information is gathered may be different things. Possible sampling units for a longitudinal social survey include: the individual, the household, the dwelling (or address), or the establishment. For a household panel, for instance, the target group of households in the initial wave might be gathered from either a household-based or dwelling-based sampling frame. In the latter case the sample will be constituted by selecting the

¹ Also see Smith (1992) or Buck *et al.* (1994, chapter 1).

persons (and their households) within the dwelling and subsequently following them regardless of whether they continue to live in that dwelling or not.²

Let us now characterise retrospective surveys, panel surveys, and record linkages in more detail.

2.2 Retrospective surveys

In retrospective surveys respondents are typically interviewed only once and they are asked about the past. The advantages of this method are its simplicity and cheapness (primarily because there is only a single interview; respondents do not have to be tracked etc), and the immediate availability of longitudinal information (since one doesn't have to wait for a second interview to measure change). The principal disadvantages are that the information about the past is typically dependent on respondents' recall of events, and the accuracy of this is questionable for some variables of policy interest. People are unlikely to remember very well earnings or income levels (or other continuous variables) beyond the immediate past, or may do so in a biased way. On the other hand, the dates of significant, low frequency, lifetime events such as getting married or divorced, having a child, or changes in one's main job, are more likely to be remembered with reasonable accuracy. These events have therefore been the focus of retrospective social surveys to date.

Example (i): U.K. Women and Employment Survey (WES)

The WES (Martin and Roberts, 1984) is a national random probability sample of all U.K. women of working age (16-59 years) in 1980, irrespective of their current employment status. The 'overall focus or objective of the survey was ... to establish the place of employment in women's lives' (Martin and Roberts 1984, p.3); it aimed to give a full picture of women's past and current labour market activity and their current plans and priorities. There were 5588 respondents, which corresponds to a response rate of about 80%. Longitudinal histories of employment, marriage and fertility were collected using respondent recall, plus much data about current characteristics such as type of job, work hours and working conditions, earnings, job search, and attitudes to and intentions about paid work. (As in most retrospective

² The initially selected addresses (or establishments) cannot be used as the sampling unit throughout the survey because they are essentially fixed units through which individuals rotate over time: those who move out of the unit are excluded from the survey's coverage. Movers are however a group of great relevance to the topics of interest here, and they are likely to differ from stayers in interesting ways.

surveys, there was no attempt to collect historical information on continuous variables such as earnings or income.) The richness and breadth of the longitudinal data (which was unique for the U.K. at that time) meant that the WES has received extensive analysis. Its large sample size allows analysis of subgroups who are few in relative numbers but of much policy interest such as women who had ever been a lone mother. See section 3 for some examples of analysis based on the WES.

Example (ii): U.K. Barriers to Leaving Income Support Survey (BLIS)

Given the potential for recall errors in retrospective surveys, there is a tradeoff between the length of the longitudinal history sought and the level of detail which can be reliably collected. The WES aimed for complete life histories on relatively low frequency events. Our second example is chosen to demonstrate the opposite end of the trade-off: that when the recall period is kept relatively short, useful data about benefit receipt dynamics can be collected using retrospective methods.

The BLIS survey (Shaw *et al.*, 1995) is a national random sample from the population of all persons who had received Income Support (IS), the U.K.'s social assistance benefit, at any time between September 1992 and October 1994. The sample was derived from benefit administration records.

About 1,000 respondents were interviewed and provided information about whether in receipt of each of several welfare benefits, economic activity status, and family status for each week within the two year observation period. For episodes in progress at September 1992, start dates were also collected. There is also extensive information about earnings, income and benefit levels at the time of the interview (but no attempt was made to measure these levels retrospectively), plus job search activity and work attitudes and expectations. The first analyses of these topics are summarised by Shaw *et al.* (1995), who also model the lengths of time people receive IS, and IS spell recurrence times.

The innovative sampling design has some implications for modelling of benefit spell lengths which should be noted. The Income Support (IS) spells observed in the BLIS survey are a combination of a random sample of IS spells starting within the two year observation window, plus a larger non-random sample of spells which began before the period. Long spells are over-represented in the latter sample (since shorter spells finish before the observation period and so will not be selected by the sampling mechanism). Analysis must therefore take account of these 'selection

biases' (see Jenkins 1995 for details). Nonetheless an advantage of the sample design is that inferences can be based on information about both short and long spells.

2.3 Panel surveys: overview

In panel surveys a sample of persons (a 'panel') is followed over time, and data collected from a sequence of interviews ('waves'). There are very many variations under this general description, but the main distinction is between (i) surveys comprising a single panel of indefinite life, and (ii) surveys comprising multiple overlapping panels of fixed life, also known as 'rotating panel' surveys.³ Under both headings, surveys may differ in the time interval between interviews (and in the number of interviews in total). A second distinction refers to the sampling unit and the population which the survey aims to represent—whether the focus is entirely at the individual level, or on individuals within their household context (see section 2.1). This distinction helps define the rules about who comprises the panel beyond the initial sample: which people in the original panel are followed over time, and how (if at all) new panel members might replenish survey numbers. A third distinction refers to the types of longitudinal information collected by the panel survey, in particular the extent to data is collected about life prior to the first interview wave and about life between waves, in addition to concurrent information.

2.4 Single indefinite life panels: (a) cohort panels

Perhaps the simplest example of a single indefinite life panel is a cohort survey. This is an individual-level panel focussing on the persons comprising a specific birth cohort of the population (or some subsample of this). By construction, the definition of membership of this group cannot change over time and so the 'following rules' for this survey type are simple: attempt interviews with all original sample members. Information might be collected about the persons in a sample member's household, but no attempt is made to follow these people: they cannot become sample members in their own right. Sometimes however, depending on a survey's purpose, an original panel may be supplemented subsequently, e.g. by immigrants of the same birth cohort as the original sample members. Even in this case the focus remains on a specific class of individuals.

³ Types of panel designs are also discussed by Kalton and Citro (1993).

Example: the U.K. National Child Development Survey (NCDS)

The U.K. National Child Development Survey is based on all 17,000 persons born in Britain during the weeks 3-9 March 1958, with some later supplementation with immigrant children of the same birth cohort. There have been interviews at birth, and ages 7, 11, 16 (information collected primarily from the children's parents and schools), and at ages 23 and 33 (information collected primarily from cohort members). Approximately 11,000 persons were interviewed at wave 5, which is about four-fifths of the number of people traced from wave 4. Comparisons of later waves with preceding ones, and with other British surveys, suggest that the NCDS has remained broadly representative of the original birth cohort. See Ferri (1993) for further details about the survey and some reports on analysis.

In the earlier waves, the survey was mainly concerned with social and medical factors associated with child development, but the scope has become progressively broader in later waves. For example waves 4 and 5 built up histories of post-compulsory education and training, labour market experiences, housing and mobility, partnerships and family formation, together with much other socio-economic information pertaining to the current interview period (ranging from earnings to social attitudes). As a result the NCDS is now a major resource for social scientists.

The potential constraints arising from the long intervals between interviews are reduced by the collection of retrospective histories covering the intervening periods. Nonetheless the scope for examination of relatively high frequency transitions and earnings and income mobility is generally limited. Some other cohort surveys, such as the U.S. National Longitudinal Surveys of Youth, reduce some of these problems by using more frequent interviewing (e.g. annually).

2.5 Single indefinite life panel surveys: (b) household panels

The typical household panel requires a more complicated design than a cohort survey because it aims to remain representative of both individuals and the households within which they reside.⁴ It therefore needs mechanisms to ensure proper representation of the populations of persons and households as time passes, in particular for representing new entrants, persons and families, into that population. The following rule which ensures the on-going cross-sectional representativeness of

the (non-immigrant) population is as follows. Define all the adults and children in the representative sample of households in the first wave as ‘original sample members’ (OSMs). Then, at second and subsequent waves, attempt interviews with all adult members of all households containing either an OSM or an individual born to an OSM whether or not they were members of the original sample.⁵ This following rule underlies the design of virtually all household panels. Practice differs however in the treatment of new panel members who subsequently stop living with an OSM. In most surveys, including the U.S. Panel Study of Income Dynamics (PSID) and the British Household Panel Survey (BHPS), these people are not interviewed again (unless they retain an important relationship with a sample member, such as parent). By contrast the German Socio-Economic Panel (GSOEP) since wave 7 has followed and interviewed all panel members, regardless of their relationship to the OSM.

Household panels typically use the same sort of instruments to collect the information about each household and its members. Questions about the household itself (e.g. type of dwelling, housing costs, enumeration of household members and relationships between them) are answered by the household head or some other designated adult. Each adult member of each household answers an individual questionnaire which asks about personal characteristics and behaviour. Information is also derived from the administration of the panel itself (e.g. addresses, and response status), and from ‘cover sheet’ details filled in by the interviewer.

The number of household panels has increased significantly in recent years. They range from the PSID—the pioneer and the longest running—to the country panels comprising the European Community Household Panel (ECHP) survey which fielded their first waves within the last year. Other on-going household panels in Europe include the BHPS, the German Socio-Economic Panel (GSOEP)⁶, and panels in Luxembourg, Lorraine, and Hungary.

⁴ In most household panels, the population of individuals refers to the civilian non-institutional population.

⁵ For further details and justification of the claim about representativeness, see e.g. Duncan (1992). To ensure representativeness extends to the immigrant population as well, supplementary samples are required—e.g. the U.S. Panel Study of Income Dynamics in 1990 added a new sample of Latino households to the panel to supplement existing information about blacks and non-Latino whites (see Hill, 1992). Similarly the German Socio-Economic Panel after re-unification included a new probability sample of households from the former East Germany. Overall representativeness of household panels is also contingent on the development of suitable weights to account for differential non-response and attrition (more about this below).

⁶ For an overview of the GSOEP, see Burkhauser (1991) and Wagner *et al.* (1993).

The ECHP surveys are organised by the Statistical Office of the European Union (Eurostat), and carried out using a comparable questionnaire by research agencies in each of the member countries. The total sample size is around 65,000. Since only one wave has been collected to date, and because very little information about the survey outcomes has been publicly released so far, it is difficult to comment as yet on the value of the data for research and policy purposes. In so far as the questionnaire is relatively short (around 30 minutes in total per individual) and focussed on poverty and social exclusion issues, it might provide a model for situations where the motivation for the new panel survey is analysis of family income dynamics.⁷ However, in many other respects, the study design is similar to that for BHPS, described below.

We select two generic illustrations of household panels to discuss: the first is the PSID, an example of a mature panel; the second is the relatively new BHPS. Our experience with establishing this survey is drawn on in section 4.

Example (i): U.S. Panel Study of Income Dynamics (PSID)

The PSID is a representative sample of the U.S. individuals (men, women, and children) and the families in which they reside. The survey began in 1968 and, although originally planned to last only five years, it has continued ever since, and aims to do so indefinitely. For a detailed description of the PSID, see Hill (1992), from which we have drawn extensively.

Throughout its life the PSID has been made up of several subsamples. The original 1968 sample was the combination of two sub-samples. The Survey Research Center (SRC) sub-sample was a random sample of the U.S. non-institutional household population and comprised almost 3000 interviews. By contrast the Survey of Economic Opportunity (SEO) sub-sample was of some 1900 low income families with heads aged under 60 years old. (This over-sampling was motivated by the PSID's original focus on the dynamics of poverty.) To maintain representativeness and account for the composition of the post-1968 immigrant population, a sample of some 2000 Latino households was incorporated in 1990. The cumulative number of persons ever included in the PSID more than doubled between 1968 and 1988, from about 18,000 to 37,500. About 19,500 persons were covered by the 1988 wave. With

⁷ See Barreiros (1995) for some general background information about the design and purpose of the ECHP surveys. See Eurostat (1994) for the wave 1 questionnaires.

a quarter century's coverage, the PSID now contains information about three generations of American families.

The combined samples can be used to produce descriptive statistics representative of the U.S. population as a whole, as long as sample weights are used to account for the differential selection probabilities. Sample weights are also modified to take account of differential attrition rates, but this adjustment is minor relative to the selection weighting. Overall response rates in the PSID have remained consistently very high. Seventy-six per cent of sampled families were successfully interviewed in the first year, 1968; in 1969 almost 90% of heads of families containing an adult OSM responded, and the wave-to-wave response rate has exceeded 96% every year thereafter. However the cumulative impact of attrition over the long length of the panel means that the proportion of wave 1 persons still responding is now below 50%.

PSID interviews occur once per year, conducted using in-person interviews between 1968 and 1972, and telephone interviews thereafter.⁸ Interestingly, and by contrast with most other household panels, in the PSID a single person—the household head—serves as the sole respondent on behalf of the whole household. This of course reduces survey costs, though there may also be some potential loss in the quality of information derived about the household context.⁹

Information is gathered about all members of sample members' households, but the most detailed information refers to the household head and to the cohabiting spouse, if present. The questionnaire has core components—those used at each interview—plus variable components which are those used for only one wave or irregularly. Core components cover income, employment and family composition, whereas the variable components have addressed topics such as health, time use, wealth, and retirement plans. Although the measurement period for many variables is the previous year (e.g. annual work hours, or annual income), since 1984 the PSID has

⁸ Except for a small number of households without telephones, or for whom telephone interviewing is infeasible. The telephone interview last about thirty minutes on average. The PSID is now moving to computer-assisted telephone interviewing.

⁹ A household panel survey using a single informant per household (as in the PSID) is very different from a panel survey which follows only a single respondent per household. The latter might be useful were one's interest solely in measuring transitions between states described by a person's household context (e.g. household income). But usually one is also interested in investigating the causes of such household-context transitions—e.g. by relating the transitions to the labour force

also collected retrospective data about labour force participation status and government programme participation on a monthly basis for the year preceding each interview.

The richness of the PSID in terms of variable coverage and panel length has established it as the data set most widely used by U.S. empirical social scientists. Hill cites ‘labor supply, consumption, lifecycle earnings, unions, compensating differentials, dynamic aspects of income distribution, income transfer programs, savings and wealth, unpaid productive activities, taxes and various methodological studies’ (1992, p. 6) as areas of basic economic research addressed using the data. She also refers to further research on poverty, motivation and economic mobility, changes in family structure (e.g. births, divorce, remarriage), child support, intergenerational transmission of economic status, and many other topics. The PSID has also been the subject of several validity studies (summarised by Hill 1992, chapter 5), which have provided reassuring evidence about the quality of PSID data. A particular advantage for policy analysis of such a long panel is that having significant macro time series variation in addition to the micro longitudinal data helps analysts disentangle causal effects better.

Example (ii): British Household Panel Survey (BHPS)

The BHPS is an immature household panel survey by comparison with the PSID, having conducted its first interviews in the Autumn of 1991. Re-interviewing of panel members occurs annually, and the first three waves of data have been publicly released. Wave 4 data is currently being processed and is due for release in April 1996. The BHPS is planned to continue indefinitely (and currently has committed funding until 1999). It shares many structural features with the PSID (and other household panels), but also has several distinctive features. For an overview of the BHPS and analysis of wave 1-wave 2 changes, see Buck *et al.* (1994).

The first wave of the BHPS was designed as a nationally representative sample of the population of Great Britain living in private households in 1991.¹⁰ Households were selected using an equal probability sampling mechanism using a standard design

transitions of the various household members—in which case the single respondent panel type is unsuitable.

¹⁰ Thus households in Northern Ireland are not included. In common with other U.K. social surveys, the BHPS Wave 1 did not sample households north of the Caledonian Canal (northern Scotland).

for British household social surveys. The achieved wave 1 sample comprises about 5,500 households, which corresponds to a response rate of about 65% of effective sample size (69% if proxies are included), i.e. much the same as in other British household surveys. Over 90% of eligible adults, approximately 10,000 individuals provided full interviews. About 88% of eligible wave 1 respondents were successfully reinterviewed at wave 2 (Autumn 1992), and the wave 2-wave 3 response rate was about 90%. The BHPS data contain weights to make them representative of the relevant national populations. The 1991 population census was used to derive weights to account for some differential unit non-response at wave 1, and these weights are further adjusted in later waves to take account of differential attrition rates.¹¹

Data is collected annually about all members of panel households, using in-person interviews with all adult household members.¹²

Six core topic areas are covered: household demography and organisation, labour market activity, income and assets, housing, health, and socio-economic and political values. This coverage makes the BHPS of rather broader scope than, say, the PSID, which collects no detailed information about the intra-household division of labour and the control of money and access to consumption goods (included in BHPS topic 1), nor about health, values and expectations on a regular basis. BHPS income data is collected in about as much detail as in the PSID though in rather less detail than e.g. the cross-sectional U.K. Family Expenditure Survey (which is designed to have an income and expenditure focus). A special BHPS feature however is the detailed information about income receipts on a monthly basis covering the period between interviews. Annual income measures can therefore be constructed as well as current period ones. Individuals' movements on and off welfare programmes can be followed in detail (and compared with concurrent labour market and home activities).¹³

The variable components in the BHPS are also distinctive in several ways. For example the BHPS now also has a wealth of retrospective information about panel members' lives prior to wave 1. Wave 2 collected lifetime histories of marriage,

¹¹ See the Appendix to Buck *et al.* (1994) for details.

¹² The individual interviews last about 45 minutes. It is planned to move to computer-assisted personal interviews from BHPS wave 6.

¹³ See e.g. Ashworth and Walker's chapter in Buck *et al.* (1994).

cohabitation and fertility and major employment transitions; wave 3 collected a more detailed job history. To gain a fuller picture of family life, a young person's questionnaire was introduced at wave 4 interviewing household members aged 11 to 15 years.

The BHPS is already recognised as a major new resource for social scientists. Its life history and other retrospective data are beginning to be extensively analysed, both because they update similar data in other sources (e.g. women's work history data in the WES) and because they provide previously unavailable information (e.g. the detailed cohabitation histories, and the monthly income receipt data). The BHPS cross-section data is also being used, because several of its topics are not covered by other U.K. surveys. Analysis of the BHPS panel data has been limited by the small number of waves available so far, but research possibilities are now expanding substantially.

2.6 Multiple overlapping fixed life panel surveys (rotating panels)

A rotating panel survey comprises a succession of separate panel surveys with staggered starting times. An initial sample of respondents is selected and interviewed a pre-determined number of times, often at intervals shorter than for most household panels. During the life of this first panel, a new sample is selected, followed, and interviewed in the same way as the first. Third and subsequent panels are constructed similarly. Thus respondents are being continuously rotated out of the survey and their numbers replenished by those being rotated into the survey. Although each constituent panel has a pre-determined fixed life, the overall survey itself usually has an indefinite life.

Following rules for rotating panels are similar to those in household panels. Every adult in each household belonging to the original sample for each constituent panel is designated an 'original sample member'. All OSMs are then followed throughout the life of their panel, even if they move to a different household. At each interview, information is collected about the OSM and the other individuals with whom they reside.

The rotating panel design has three additional features of special note. One, the shorter interval between interviews relative to household panels can be used to help reduce recall errors about relatively high frequency events and details of variables

such as income. Second, the survey as a whole can also provide better cross-section information at a point in time from the combination data from the constituent panels (in so far as the measurement periods for each of these overlaps). The increase in sample size reduces sampling errors. Three, by restricting the duration of each panel to a finite period, often only a few years, problems of attrition are reduced and representativeness more easily maintained.¹⁴

The rotating panel surveys which are in operation have exploited these features in different ways. Of the two surveys we discuss, the U.S. Survey of Income and Program Participation (SIPP) has put greater emphasis on the first one, whereas Canada's new Survey of Labour and Income Dynamics (SLID) has put greater emphasis on the second.¹⁵ It is worth noting that these are the only large scale panel surveys to date to be specially created by government organisations explicitly for social policy analysis purposes.

Example (i): U.S. Survey of Income and Program Participation (SIPP)

As its name suggests, the SIPP was designed specifically to focus on incomes and programme participation. Prior to its introduction in the mid-1980s, the principal U.S. data source for these topics was the March Income Supplement to the Current Population Survey (CPS), a large annual cross-section survey primarily designed to examine employment and unemployment rather than incomes. The CPS collects no information about within-year changes in income and participation in welfare programmes; nor does it collect the data about assets, liabilities and other variables needed to properly assess respondents' economic well-being and programme eligibility. The SIPP is the outcome of a research programme and review instituted during the late 1970s to investigate how these limitations might best be overcome. The survey is described in detail in by U.S. Bureau of the Census (1991), on which we have drawn extensively for our summary.

Each SIPP panel is based on a nationally representative sample of households in the U.S. civilian non-institutionalised population. The first panel was recruited in late 1983, comprising adults in almost 19,900 households. They were first

¹⁴ These features are often gained at the cost of greater complexity and higher budgetary costs—more about this trade-off in Section 4.

¹⁵ Rotating panel designs are used for most labour force surveys around the world, including e.g. the U.K. Quarterly Labour Force Survey, and the N.Z. Household Labour Force Survey. However these are address-based panels rather than household panels (more on this feature below).

interviewed during the period October 1983-January 1984, and then at four month intervals thereafter, for up to nine waves, or 36 months in total. Interviews for the second panel began in February 1985, and continued for eight interviews four month apart (two and half years in total). Subsequent panels follow the same pattern, with interviews beginning each February.¹⁶ For the period February through July, three different panels, or some 30,000 to 40,000 households, are in the survey.

Response rates in the initial panel samples are high: e.g. only 5% of eligible households in the first panel were not interviewed. Cumulative wave on wave sample loss is thought to be of the same order as in other longitudinal surveys. See U.S. Bureau of the Census (1991, chapter 7) for details. The SIPP data contain weights to make them more nationally representative. These weights take account of differential unit non-response in each initial wave, and are further adjusted in later waves to take account of differential attrition rates.

Data is collected in each panel wave about every adult OSM and all the members of his/her household using in-person interviews with persons aged 15 years of age and over. There are four core topics in the SIPP questionnaire. The first collects data about labour force activity, participation in government programmes, receipt of various type of income and asset ownership. The second gathers information about employment and self-employment and earnings from them. The third details the amounts of income from the sources identified under the first heading. The fourth topic is a short section about participation in several specific programmes and addressed only to household heads.¹⁷

Information is collected under these headings for each of the four calendar months preceding the interview month, with the aim of building up a continuous record of data for every month during the life of the panel. SIPP is unique in developing monthly records for the *amount* received from each earned and unearned income source (except for asset income, which refers to the four month period). Two characteristics—whether the respondent is working and whether seeking work—are in

¹⁶ There have been some variations in this standard pattern: e.g. the number of interviews was reduced for some panels, and the sample size increased in others. The panel starting February 1990 over-sampled households types with high poverty risks. See U.S. Bureau of the Census (1991) for details.

¹⁷ There was also a fifth, 'missing wave' section, which was part of the core until mid-1987, which had questions for respondents to wave $n+1$ and $n-1$ but not wave n .

fact collected on a weekly basis. This very detailed monthly tracking, which requires the frequent interviews, allows detailed analysis of programme eligibility and take-up.

In addition to the core components, SIPP contains a large battery of variable components, some of which are asked of every panel but only in one or two waves and some asked on a once-off basis. The fixed topical modules currently includes the following subjects: health and disability status and use and access to health care; retrospective personal histories (data about education and training, brief histories of work, marriage and separation, fertility, migration, welfare reciprocity); and assets and liabilities (asked in two waves, one year apart). There are also two ‘annual round-up’ modules to gather e.g. annual earnings data of the sort declared on tax assessment forms. The occasional topics have included child care usage and child support.

The SIPP has become, as intended, the major U.S. data source about income and program reciprocity dynamics over short- and medium-term time periods. For these variables it provides a valuable and detailed complement to the longer term perspectives provided by the PSID. The major limitation of SIPP is the relatively short overall length of the observation period for any individual. This was recognized as a major constraint on academic and policy analysis in a review undertaken for the Bureau of the Census (Citro and Kalton, 1993). The length limits the degree to which the study can illuminate long spells of poverty or welfare dependency or multiple transitions between states. The review committee’s recommendation to extend the lengths of the panel is about to be implemented.

Example (ii): Canada’s Survey of Labour and Income Dynamics (SLID)

The SLID is a major new Canadian social survey, the purpose of which is ‘to map out patterns of labour market activity and changes in income, and to record the events which achieve these goals’ (Webber 1994, p. 7). It is the first Canadian household survey to provide national data on longitudinal income fluctuations, and is explicitly intended to support research about the nature and extent of poverty. The survey began in 1993 and is intended to continue indefinitely, using a succession of overlapping panels each six years long. Webber (1994) provides an introduction to SLID, and we have drawn on this extensively.

The initial sample of some 15,000 households comprising the first SLID panel was derived using the same probability sampling method as the existing national Labour Force Survey and was also chosen to represent all civilian non-institutional

households in the ten provinces. The first interviews, which collected background information, were conducted in January 1993 and are being followed by twelve further interviews, two per year for six years in total. Labour market information is collected in each January interview, and income information each May (the time of the year when taxpayers are preparing their tax returns). New panels are scheduled to begin every three years (and thus the second in January 1996), and will use exactly the same pattern of interviews. There will therefore be a three year overlap of each successive panel.

First estimates of response rates are encouraging. Of some 30,000 eligible adults in the first wave, 72% completed both an income and labour interview, 14% completed the former only, 6% the latter only, and 8% completed neither (Statistics Canada, 1994). The first publications about, and data from, the January 1993 preliminary interview, are due for imminent release.

As in the SIPP, data is collected in each panel wave about every adult OSM and all the members of his/her household using interviews with persons aged 15 years of age and over. By contrast, interviewing is not in-person, but by telephone (and computer-assisted).

SLID aims to inform seven core research and policy themes: employment and unemployment dynamics; lifecycle labour market transitions; job quality or quality of working life; family economic mobility; dynamics of low income; demographics; and education. To this end each interview asks for information about the previous year. For the variables which can be summarised in terms of transitions between states and spell durations, spell start and end dates data are collected for each relevant event. Examples include movements into and out of employment and unemployment, absences from work, job changes, geographic mobility, marriages and separations, births and deaths. Many other variables are collected using an annual measurement period, e.g. annual work hours, and annual income for up to 25 income sources, total income and taxes paid. However information about the receipt (but not level) of unemployment insurance, social assistance, and worker's compensation benefits is collected on a monthly basis. Assets and debt data are scheduled to be collected once or twice during the life of each panel. In addition the SLID contains a large amount of standard data about personal and family characteristics. No variable components to the questionnaire are currently planned to supplement these core topics. Moreover,

and of more serious consequence, it does not appear that any retrospective history data is being collected to summarise people's labour market and family life histories prior to recruitment into their panel.

In sum, the SLID is a significant new investment, and especially interesting because it combines several elements of household panels—an annual measurement focus in particular—with a rotating panel design.

2.7 Record linkages

Longitudinal data can also be collected without personal interviews, by linking together personal records from existing temporally-separate data sources. These data sets may be administrative records gathered for official purposes, e.g. of income tax returns, social security benefit administration records, or surveys such as national censuses.

Longitudinal data sets constructed in this manner have some significant advantages relative to the interview-based survey methods discussed so far. First, they usually have very large, often enormous, sample sizes, which means that analysis can be constructed for almost every population subgroup of interest, and sampling errors are minimised. Second, by not using interviews and respondent recall, there is no respondent burden or recall or reporting errors (apart from that in the original sources).

However there are also significant problems with using record linkages. First, linkage may simply be impossible, as a result of confidentiality or privacy restrictions relating to collection of the original data. For example the names and addresses of N.Z. quinquennial Census respondents—a prerequisite for linking records—are not recorded on the computerised census files, and all the original census forms are destroyed. A second problem is that analysis is constrained by the coverage of and the variables contained in the original surveys, which are usually designed for some other, often specialist, use. For example tax records refer only to those required to complete a tax return, and so may exclude many of the low income groups who are of particular policy interest. The range of variables collected is also rather limited. Tax returns often contain quite a bit of information about income, but relatively little about family and household variables, whereas the reverse is typically true for a census. (The U.K. census, by contrast with that in the U.S. or N.Z., collects no income information. But

even in N.Z. the income responses refer to income within a relatively small number of income bands.) A third problem is that the linked record data set only provides concurrent information; events or status during the intervening period (potentially several years long) may not be recorded. Many events of social policy interest with a relatively high degree of flux—labour market transitions or income variability—may therefore not be observed.¹⁸

Example: the Finnish Longitudinal Census File

This longitudinal data set was created by linking records from the Finnish censuses of 1970, 1975, 1980, 1985, and 1990. According to Starck (1994), from whom we take the detailed information following, the file contains all individuals included in at least one of the four census populations between 1970 and 1990, i.e. more than six million people.

Linkages were feasible from the 1970 census onwards because, since that time, each individual has had a unique numeric identifier—a social security number—and census records have been held in a suitable computerised form. Development of the data set was also facilitated by the substantial and increasing direct use of register data to compile the census. For example from 1980 onwards, the Finnish population has been counted from registers rather than enumerations from questionnaires, and for the 1990 census no questionnaires at all were used: all data were collected by combining data from registers.¹⁹

Moreover a range of new variables, including income, have been included in censuses from 1970 onwards, thereby substantially increasing the research usefulness of a linked file. The longitudinal file now contains information for each person about demographics (e.g. sex, age, place of birth, marital status, language, marital status, household size), economic activity (economic status, occupation, industry, work location), the dwelling, level and field of education, plus income from wages and salaries, self-employment, and other income subject to state taxation. There are also domicile codes enabling family data links to be established.

¹⁸ The U.K. Longitudinal Study, based on links between decennial censuses, reduces this problem slightly by also linking in registration data about births, deaths, and marriages. The problem is also a lot smaller in the Danish Longitudinal Data Base, which comprises annual observations on a 5% random sample of persons in the Danish population for the years 1976-1990, and derived by linking together data from administrative registers.

¹⁹ See Harala (1994) for an overview and evaluation.

This new data set is obviously a very rich one both in terms of numbers of variables and sample size: it will be used extensively. (For example trends in earnings and income inequality and mobility are already being studied by Markus Jäntti.) Against these substantial advantages, one must also remember that because of the five year interval between observations, the survey is not very suitable for analysing high frequency transitions. And in any case, the potential for linking of register data is much more constrained in most countries by privacy or ‘data protection’ legislation, than it is in Finland.

2.8 Repeated cross-section surveys

These surveys do not contain repeated observations on the same individuals and hence are not genuine longitudinal surveys. However we include a brief discussion of them here for completeness since, as we discuss in Section 3, they can sometimes be used to do longitudinal analysis.

The special case of these surveys which is perhaps closest in form to longitudinal surveys is that arising when they are converted into ‘pseudo-panel’ form, i.e. where within-group averages are derived for each of many cohorts, typically identified by age group, since e.g. all persons aged 25-30 years in 1970 are aged 45-50 in 1990. (In general cohorts may be defined as groups with a fixed membership which can be identified in the successive surveys.) The averaged data for the cohorts from the repeated cross-sections forms a time-series, just as one has in a genuine panel survey, and can hence be used to estimate many longitudinal models of interest (see below for details). In fact the pseudo-panel data have several advantages relative to genuine panel surveys. One of the main ones is that sample sizes are typically much larger, which helps reduce sampling errors. Moreover, the within-cohort averaging may reduce the impact of idiosyncratic measurement errors and, by construction, the data cannot be subject to attrition biases.²⁰

Examples

Most countries have repeated cross-section surveys of the type we refer to. For example in the U.K. there are two, the Family Expenditure Survey (FES), and the General Household Survey (GHS). The FES yields a time series of some thirty years with annual observations for some 7000 households (about 19,000 persons) and has a

²⁰ For a review and illustration, see Deaton (1985).

particular focus on income and expenditure. The GHS is a more general purpose social survey with a larger sample size, and can provide a two decade time series. Similar surveys exist in the U.S.A.: there is the annual Consumer Expenditure Survey (CES), and notably the annual Current Population Survey (CPS) which contains data for some 60,000 households.

3. COMPARISON OF INFORMATION PROVIDED

It is only meaningful to evaluate the advantages and disadvantages of different types of data in connection with the analyses for which they are to be used. And given the differences in the costs of data collection, it is clearly important to make such evaluations: it is desirable to use the ‘cheapest’ data that addresses the analytical issue at hand effectively. In this section, we review three generic types of analysis in relation to data: income dynamics, transitions between states, and impacts of policies. We also discuss a number of previous studies to illustrate how longitudinal data can be used to analyse policy impacts.

3.1 Analyses of income dynamics

An important topic of analysis is the way in which a person’s income or the income of the family in which he/she lives changes over time, and how these changes are related to other events, such as unemployment and marital dissolution. In particular, there is interest in how long people spend in poverty and how often they find themselves in low income situations.

Repeated cross-sections clearly cannot answer this question, and retrospective surveys are not feasible either because of the difficulty of accurately recalling income levels and changes in them. Census based record linkage panels are also inappropriate for analyses of income dynamics because of the long interval between observations, usually five or ten years. In sum, analysis of income dynamics must be based on genuine panel surveys.

An influential example of the use of panel data to address this topic is Bane and Ellwood’s (1986) study of the dynamics of U.S. poverty using the PSID. We return to the discussion of their analysis below.

3.2 Analyses of transitions between states

Another generic set of analyses concerns the transitions between states and the impact of policy variables and individual and family characteristics on such movements. Examples of these states include employment status (e.g. full-time, part-time, unemployed, out of the labour force), marital or partnership status, household living arrangements (e.g. living alone, with a partner, with friends), number of

children, occupational group, and whether or not people are in particular types of education or training.

Analysis of transitions *between* states is more complex than analysis of characteristics *within* states. The reason is that all transition analyses assume—either explicitly or tacitly—some model describing the underlying dynamic process. Otherwise the analytical procedures used would have little meaning. Estimation of these models generally requires some sort of history of the occupation of these states over time.

In the usual case, when life history information is required, the question arises whether one should collect it through a panel study or a retrospective history. The latter is certainly cheaper, as only one cross-section survey need be carried out with an appropriate life history component. The U.K. *Women and Employment Survey*, discussed above, is a retrospective longitudinal survey of precisely this form.

What are the advantages of panel over such retrospective data for the analysis of transitions between states?

Some panels only ask about current status (e.g. marital or employment status) rather than the dates of transitions during the time since the last interview. As a result, changes in status during this interval may be missed. For example, for a person who experienced an unemployment spell between two interviews but was employed at the time of each interview, the data would not record a change in employment status. Even when changes in status are recorded, the dating of the events within the period between interviews is ignored. In terms of the information that goes into the analysis, a person who changes state one month after the previous interview will be observationally equivalent to one who changed state in the month before the present interview. This makes it more difficult to estimate the pattern of dependence of transition rates on duration in a state, and this problem obviously looms larger the longer the interval between interviews. These problems can be eliminated by collecting data at each interview about important transitions since the previous interview, e.g. job changes, changes in marital status, and the birth dates of children. The BHPS does this.

A more general limitation of panels for dynamic analysis is that no information is available for events of interest that occur before the first round of the panel survey (unless extra data is specially collected—see below). This so-called ‘left

censoring' or 'initial conditions' problem can bias dynamic analyses because of non-random sample selection. For instance, if we were studying the chances of divorce, we would usually need to know how long ago each person married because divorce rates may depend on marriage duration. We would not know this for people who were married at the start of the panel. Restricting the sample to only those individuals who married after the start of the panel, and for whom we know their marriage date, may appear to solve this problem, but such a restriction would bias the sample in favour of younger people and those with shorter marriage durations. In general, pure panel surveys suffer from having inadequate information to account for the initial states in which panel members are observed. Getting around this problem when estimating dynamic models can be difficult and involve assumptions that may be difficult to justify (see e.g. Heckman, 1981).

The more straightforward solution for many analyses is to collect some salient life history information at the beginning of the panel. For example, partnership (marital and non-marital, birth employment and occupational histories were collected in the BHPS. This is, of course, quite costly, but a useful compromise position is to collect the particular retrospective information needed for one's analysis. Identification of these key pieces of information requires, however, that one has a very clear idea before collecting the data of the analyses one wishes to undertake in the future. For example, if one wishes to study receipt of social benefits, key information would be the starting date of benefit spells in progress at the start of the panel and whether or not it was first spell for the respondent. Summary measures of labour market experiences, such as starting date of current spell and details of first job, may be sufficient for some analyses of labour market transitions, but others which stress persistence of behaviour over people's lives may require more information. Neither the PSID nor the SLID collected life history information at the start of the panel.

Retrospective data on say marital and employment histories are, however, likely to suffer from recall errors, which is less of a problem with panel data particularly if the interval between interviews is relatively small. This is demonstrated by Peters' (1988) analysis of the young women's cohort of the U.S. National Longitudinal Survey of Work Experience (NLS). She compared two types of NLS longitudinal marital data, retrospective and panel, for the same set of people, and found systematic inconsistencies between them concerning the dating of marital

events which seemed to relate to factors that increase the difficulty of recall. Furthermore, episodic events, such as unemployment, can be grossly under-reported, and the severity of the problem appears to increase rapidly as the recall period increases. Job and cohabitation histories may also be affected by this recall problem. Again, some key summary measures, such as age at first cohabitation before marriage, or women's date of first major employment interruption may be sufficient for many analyses.

Panel data also have the advantage over retrospective data in terms of the explanatory variables available for models of transitions. Those characteristics which vary over time, like wages or other income, are particularly difficult to collect in retrospective surveys, but are collected in panels. These characteristics may be important explanatory variables, but analyses based on retrospective histories would omit them. Only variables fixed over time, like sex, race, and, for some analyses, education, or 'community-level' variables like the regional unemployment rate, or time-varying variables which can be constructed from other parts of the life histories, like births and age of youngest child when analysing employment transitions, can be used as explanatory variables in analyses using retrospective data.

Thus, panel data reduce problems of recall and can provide a richer set of explanatory variables. Are these advantages worth the cost of a panel study for analyses of transitions?

With respect to recall, statistical techniques for the analysis of transition data appear to be relatively robust to errors in dating events, but failure to recall events altogether is more problematic. Caution needs to be exercised when analysing episodic events collected through retrospective surveys. Recall is likely to be much more accurate for major life events like first cohabitation, marriages, births, divorces, primary jobs and major workforce interruptions. These are indeed the events which have been most analysed with retrospective data.

While it is always desirable to have more relevant explanatory variables in analyses of transitions, it is still possible to develop relatively rich transition models using retrospective data. This point is demonstrated by Ermisch and Wright's (1991a, 1991b) analyses of the duration of single parenthood and the dynamics of women's employment using event history data from the 1980 British Women and Employment Survey (WES). These two studies, discussed further below, utilise data from other

parts of the WES life history data, plus fixed characteristics and community-level variables, to construct relevant explanatory variables including a welfare benefits variable which varied over time and with the number of children. Moreover, the WES also allows relatively unstructured analysis of e.g. how the dissolution of a woman's marriage affects her employment using the relative dates of starts and ends of employment and marriage spells (as in Ermisch 1991, Chapter 7).

Although our discussion so far has focussed on comparisons of retrospective and panel data, it should be noted that the parameters of some transition rate models can be in fact be estimated from repeated cross-section surveys with the imposition of specific assumptions. For example, Markov models—which assume that transition rates depend only on a person's current state, and not on his/her history of states occupied in the past, nor duration in the current state, can be estimated if we make some assumptions about the variability of the Markov transition rates in the population. Take, for example, the simplest case, in which everyone has the same rates. Letting p_t be the proportion of the population in a given state (e.g. employment) at time t and μ and λ the inflow and outflow rates respectively, the fundamental accounting identity in the Markov model is $p_t = \mu + (1-\mu-\lambda)p_{t-1}$. Using aggregate frequency data from repeated cross-sections to measure p_t at a series of points of time, least squares estimation of this equation identifies the transition rates μ and λ from the estimated slope and intercept parameters. Clearly this strategy could be generalised by estimating the rates for a number of homogenous groups (e.g. defined by sex), but our example is primarily designed to provide the intuition into why repeated cross-sections may allow us to estimate dynamic transition rate models.

Using the same logic, Markov models in which explanatory variables affect the transition rates can also be estimated, provided we assume that measured explanatory variables capture all the differences between individuals which affect transition rates. Moffitt (1993) demonstrates this, and applies the technique to the estimation of Markov models of married women's entry rates to and exit rates from paid employment using data from the U.S. Current Population Survey (CPS) for the 21 years from 1968-88.

As mentioned in section 2, repeated cross-sections are likely to have much larger samples than panel studies and they may be more representative over time because they are unaffected by attrition. Nevertheless, retrospective or panel data

allow for more flexible and richer analyses of transition dynamics than is possible with repeated cross-sections. For example, transition rates can be allowed to vary with duration in the state and we can allow for unobserved individual heterogeneity. The repeated cross-sections' approach to estimating transition rate models is a clear example of substituting assumptions ('structure') for data.

Record linkage panels based on the population censuses are like panels with five years between interviews, and can be supplemented by information from registration records about the dates of vital events (births, deaths, marriages and divorces). With the exception of transitions associated with these vital events, census panels suffer from the disadvantages of panels discussed above, and these are reinforced by the long interval between "interviews". It is nearly impossible to use such data for the analysis of transitions between states involving relatively short durations of stay, like unemployment and employment. Furthermore, because of this long interval, they do not possess a sharp advantage over retrospective data in terms of explanatory variables for the analysis of birth and marriage transitions.

3.3 Analyses of impacts of policies

Under this heading we address topics such as the measurement of the impacts on subsequent earnings and employment of policies such as education and training, subsidised child care, and equal opportunity (affirmative action) programmes. We also consider assessment of the impacts of welfare benefits and housing subsidies on employment, housing consumption and household welfare.

When a policy changes during the course of a panel study, panel data provide an opportunity to observe the same individuals' circumstances or behaviour before and after the policy change, providing potentially strong evidence of the impact of the policy change. The strength of this evidence varies, of course, with the nature of the policy. If effects accumulate over long periods or there are substantial lags it may be more difficult to measure the policy impact from a relatively short panel, but for other policies with more immediate impact it will be much easier to measure policy effects. For instance, the impact of legislation mandating equal pay for men and women in the same jobs could be measured quite well from observing changes in the earnings of panel members before and after the equal pay legislation was implemented.

Panel data also allows measurement of the impacts of public programmes, like training, in which the people affected by the programme select themselves (e.g. by deciding whether to undertake training or not). Impacts could be measured using observations of earnings before and after participation in the training programme, taking into account people's decisions concerning whether to participate in the programme or not, such as in Ashenfelter (1978). Indeed, under some assumptions about how person-specific components of people's earnings evolve over time, it is sufficient to use information on earnings in two or more periods *after* receiving training to identify the effect of training (e.g. as in Heckman and Wolpin's (1976) study of the impact of affirmative action programmes on earnings).

Panel data are not, however, absolutely necessary to identify the impacts of such policy programmes. Heckman and Robb (1985) show that despite non-random participation in programmes like training, repeated cross-section data often can be used to identify the same policy effect parameters as longitudinal data. Estimators of policy effects which only can be implemented with longitudinal data require different assumptions than cross-section estimators, but it is not obvious which sets of assumptions are more plausible. The larger samples which are usually available in cross-sections may improve the precision of the estimates, and many of the cross-section estimators require only sample means of variables, making them very simple to compute and robust to measurement error in all of the variables. Thus, repeated cross-sections may provide the best solution to estimating many policy effects. Indeed, Heckman and Robb (1985) show that single cross-section estimators may provide consistent estimates of policy effects under similar or slightly more restrictive assumptions (e.g. that there is at least one variable which helps predict whether a person participates in a programme or not).

Repeated cross-sections also have an advantage over panel data when policies primarily affect groups which make up a small proportion of the population, like one parent families. The repeated cross-sections provide larger samples of these groups than could be provided by panel data, and changes in the policy variables, like welfare benefits and child care subsidies, over time may supply the variation required to identify the policy effects. Ermisch and Wright (1991c) use a time series of ten British General Household Surveys to identify the effects of welfare benefits on single parents' employment and Bingley *et al.* (1994) use a time series of British Family

Expenditure Surveys to investigate similar issues, including the impact reform in child support from absent parents.

In sum, it is difficult to make general statements about which type of data is best for analysing the effects of public policies. The choice depends on the particular policy impact that is to be addressed. To illustrate the applications of longitudinal data to policy analysis and evaluation, we consider a selection of examples spanning topics cited in the Terms of Reference.

3.4 Longitudinal data and policy analysis: examples

In many countries there is concern about the length of time which people spend below the poverty line or dependent on income maintenance programmes. This concern is often related to the question of whether there is an important section of the population who are part of a ‘poverty culture’, separate from the rest of society in terms of norms and behaviour. In order to study the *persistence of poverty*, panel data on incomes is essential.

Bane and Ellwood (1986) addressed this topic using U.S. PSID data, and Stevens (1994) extends and updates their analysis using additional PSID waves. Both sets of authors measured poverty using data on annual income deflated by a family needs standard. A spell of poverty was defined as beginning in the first year that income was below the needs standard and ending when income was above the needs standard after having been below. From the spells of poverty identified in the PSID, they estimate exit probabilities, from which they can estimate the distribution of completed spells for persons beginning a spell of poverty and the distributions of completed and uncompleted spells for poor persons at a point in time. When this spell data is combined with information on the dating of other events like changes in earnings or in family structure, it is possible to analyse why people move into and out of poverty.

Bane and Ellwood find that most of those who ever become poor will be so for only a short time, but the majority of people who are poor at a given time will have very long spells of poverty. These relatively few people with long spells of poverty account for the bulk of all poverty. Stevens (1994) found the same result applied when additional waves of PSID data were used. She also studies poverty spell repetition: of persons leaving poverty, she estimates that more than half of them will

return to poverty within five years, and such returns are more common among those who have recently ended long poverty spells.

Other analysts have used the monthly data in the U.S. Survey of Income and Program Participation (SIPP) panel to study similar issues, and receipt of Aid to Families with Dependent Children (AFDC) in particular. Fitzgerald (1991) finds that roughly one-third of spells of receipt of AFDC assistance end within six months, and that two-fifths last two years or more. He also found a small negative effect of AFDC benefit levels on the exit rate from AFDC. Blank and Ruggles (1994) study how long it is before former AFDC recipients end up back on the programme. They find that people return to AFDC assistance relatively quickly, with one-fifth of those leaving AFDC becoming recipients again within nine months. Shaw *et al.* (1995) derive related estimates for the U.K. (though using data from a retrospective survey design).

Bane and Ellwood (1986) find that of those who start a poverty spell, nearly half the starts of a poverty spell were associated with changes in family structure and other life cycle events. Although only a minority of poverty spell beginnings were associated with a fall in the household head's earnings, over half of the poverty spells ended with a rise in the head's earnings. Thus, family dynamics are just as important as earnings dynamics in understanding poverty. All of these analyses provide a crucial background to debates about policies to help the poor; clearly these analyses could not have been carried out without panel data.

A closely related issue concerns how programmes to assist the poor affect people's behaviour and ultimately the length of time they receive assistance from such programmes. In other words, how do public assistance programmes ('welfare benefits') affect the distributions of poverty spells and welfare benefit receipt spells discussed above? The focus of our examples will be on one parent families headed by a woman (i.e. 'lone mothers'), whose numbers and welfare dependency have been growing in a number of countries.

Clearly one way off of public assistance and possibly out of poverty for lone mothers is to obtain a job. Thus, we are interested in the *effect of welfare benefits on taking or leaving paid employment*. This issue could be addressed with panel data, but panels often contain relatively small numbers of women who have ever been a lone mother during the life of the panel. Retrospective lifetime employment, partnership and childbearing histories can be used to construct spells of lone

motherhood as well as employment for women who have ever been a lone mother. This produces relatively large numbers of spells of employment and non-employment among lone mothers, and as these span periods in which welfare benefits varied, it may be possible to identify their effect on lone mothers' entry rate to and exit rate from employment.

Analysis of the issue using longitudinal data constructed in this way has been carried out by Ermisch and Wright (1991b). In their case, they use the lifetime histories in the Women and Employment Survey (WES), and they estimate transition rates among full-time, part-time and no employment which depend on welfare benefits and other variables. While they find some evidence that higher welfare benefits increase exits from and reduce entry to paid employment, the model indicates a strong upward trend in lone mothers' exit rate from full-time employment which dates back to the mid-1950s and which cannot be accounted for by measures of changes in welfare benefits, women's real wages or unemployment. Policymakers should clearly be concerned about this trend, as it indicates a decreasing reliance on earnings as a means of support for lone mothers.

It is also noteworthy that this dynamic analysis confirms in most respects the results of the same authors' analysis of repeated cross-sections of the General Household Survey (Ermisch and Wright, 1991c), particularly the strong effect of woman's potential wage on the probability of employment in both analyses. The larger measured effect of welfare benefits in the repeated cross-sections analysis may arise from the inability to measure lone mothers' non-benefit, non-labour income from the retrospective data in the longitudinal analysis. As discussed above, the inability to measure important explanatory variables can be an disadvantage of retrospective longitudinal data for some analytical purposes.

Another possible route off welfare benefits and out of poverty for lone mothers is to find a new partner. The incentives to find a new partner could be affected by the level of welfare benefits, which raises an important policy question: what is the *effect of welfare benefits on the duration of lone motherhood*? Using the spells of lone motherhood constructed from the WES, Ermisch and Wright (1991a) address this issue by estimating the effects of welfare benefits and other variables on the (re)marriage rate of lone mothers. They find that, for the largest group of lone mothers, previously married mothers, there is no evidence that higher welfare benefits

prolong the length of lone motherhood. Among never married mothers, there is evidence that welfare benefits may prolong it indirectly by reducing the probability that the mother works, employment being found to be associated with a shorter duration of lone motherhood. The failure to find that higher welfare benefits significantly lengthen spells of lone motherhood probably reflects the low level of welfare benefits relative to the share of income in marriage going to a mother and her children.

There are other aspects of Ermisch and Wright's study which also may be of interest to policy-makers. First, the analysis suggests that lone mothers in poorer economic circumstances remain lone parents longer; that is, there is a 'selection' of poorer mothers into the population of lone parents at any point in time. Second, the study finds that the U.K. 1969 Divorce Reform Act, which made divorce easier and increased the number of people re-entering the marriage market, reduced previously-married mothers' durations of lone parenthood.

Our final example concerns measurement of the impact of training to disadvantaged or hard-to-employ individuals on their subsequent earnings. We consider the evaluation by Björklund and Moffitt (1987) of the Swedish government manpower training programme, for which people who are unemployed or at risk of being unemployed are eligible. The authors use panel data from the Swedish Level of Living Survey measuring earnings of individuals at a number of points in time, in conjunction with register data from the National Labor Market Board containing information on individuals who undertook manpower training. (Although the Swedish Level of Living Survey is a rather intermittent panel with intermittent interviews supplemented by register data, the same evaluative approach could be based on a panel with more frequent regular interviews which contained information about when people took part in training programmes.)

Björklund and Moffitt's analysis indicates that there is considerable variation in the rewards from training in the population, with people who gain more from the training being much more likely to participate in the programme. This heterogeneity in rewards has important policy implications. It indicates that bringing more people into the training programme (by, say, raising stipends to join) lowers the mean rate of return from training; that is, the marginal person entering the programme obtains lower rewards from training. The amount of reduction in rewards relative to marginal

programme costs is important in deciding on the size of the training programme. Björklund and Moffitt's estimates suggest that the Swedish programme is too large.

3.5 Conclusions from sections 1-3

Our general conclusion from the review so far is that the choice of longitudinal survey design depends critically on the topics which it will be used to analyse. However we can be more specific about survey design choice when information about family income dynamics is given priority. In this case the choice of design is effectively narrowed to some form of household panel survey—either a single panel survey of indefinite duration or multiple panels of fixed duration. Section 4 assesses the issues underlying this choice in more detail.

4. SURVEY IMPLEMENTATION AND RESOURCE REQUIREMENTS

In a discussion of the promise and problems of household panel studies, Greg Duncan writes: ‘If disadvantages are to be minimised and advantages maximised, the most important steps for conducting a household panel survey project are to: 1) ensure that the initial sample is of the highest possible quality; 2) employ the proper rules about whom to follow and interview; 3) minimize bias due to panel attrition; 4) use feedback techniques during interviewing and check for cross-wave inconsistencies to minimize errors in the measurement of change; and 5) wherever possible, gather continuous measures through the panel period’ (1992, p.4). In this section we discuss the implications of these desiderata—together with a number of others Duncan does not mention—for the implementation of a panel.

Although we assume in this section that a panel study is required, the previous sections also suggest that a number of alternative panel designs exist. The main distinction is between a single indefinite life panel (as with PSID, GSOEP and BHPS), and a set of overlapping limited life panels, also described as a rotating panel (as with SIPP and SLID). It will become clear that there are distinct advantages and disadvantages to these two methods, and a significant cost difference. Which is more suitable depends on the precise goals of those seeking the new survey. Before discussing this comparison in more detail however we first discuss design issues which are common to both types of design.

4.1 Sampling design

The first wave sample is critical to the future success of a panel study. It is usually extremely difficult to correct errors made at this point. This has a number of implications for the organisation of a panel project which are discussed below.

Conventionally panel studies use a wave one sample design closely akin to that used in high quality cross sectional surveys, i.e. a probability sample, *clustered* to reduce interview administration costs, and *stratified* to increase the precision of estimates given that clustering.

Clustering tends to inflate the variance of estimates, because observations are not truly independent. This reduces the effective sample size. For most types of data

however this reduction is substantially less than the reduction in sample size which would be enforced by the extra costs of using a simple random sample. There has been some discussion of the particular merits of clustering in a panel study: see e.g. Coxon and Corti (1991). While geographical mobility of panel members will reduce the cost advantages of clustered sampling to some degree, the cost advantages over simple random sampling are likely to remain. Moreover, there is an increasing accessibility of techniques for variance estimation in the presence of complex designs, and there are no intractable problems in applying these techniques for panel studies.

There does remain a cost trade-off to be made between the number and size of clusters and the precision of estimates. The more smaller the clusters which are used, the lower the variance inflation, and hence the higher the effective sample size, but the higher the fieldwork costs. The sorts of calculations to be made are similar to those for existing national cross-sectional social surveys which use clustering. This issue also has impacts on questions of oversampling—we discuss these below.

4.2 Sample size

The answer to the question of what the appropriate sample size should be depends heavily on the proposed use of the panel survey. The key issue is to have adequate numbers of observations on each event or persons in population sub-groups of interest. However there is no definitive answer to the question of what sample size is adequate: it depends on what level of precision of estimate is acceptable. To get a feel for the nature of the issue, consider the following two examples.

The BHPS has a sample size of 5500 households, and 9900 full respondents at wave one, of whom 242 were recorded as leaving unemployment between wave one and wave two, and 210 left the major social assistance benefit, Income Support. This order of sample size is adequate—especially where multiple events are accumulated across a number of waves—for investigating e.g. the factors associated with coming off welfare benefits. However it may be insufficient to investigate differential effects of factors on different subgroups, at a single point in time.

To give a New Zealand example, with a sample of 5000 households, we would expect to find at wave one approximately 500 non-Maori unemployed persons, and 130 Maori unemployed persons, assuming 1993 unemployment rates.²¹ With this sample

²¹ Calculations based on Statistics New Zealand (1994, Tables 13.2 and 13.10).

size it would not be possible to say whether an observed 10% difference in the rate of flow out of unemployment by wave 2 (e.g. between 60% and 50%) was due to sampling variation or reflected genuine differences between Maoris and others in the propensity to leave unemployment. A 15% difference would just be significant if one used 95% confidence intervals. If the sample was reduced to 3000 households, one would have to observe a 20% difference to be confident that differences in outflow rates were significant.

Two other factors should be borne in mind. First, attrition will tend to reduce the numbers over time, at least for those who provide data at all waves. Moreover some policy relevant subgroups may experience above average attrition (e.g. the young and mobile).

Second, when the concern is with events and transitions, observations may be accumulated from more than one wave. Hence, for example, although the number of transitions out of unemployment may be small between one pair of waves, a much more significant number will be accumulated over time. Of course, if events have to be accumulated from a number of waves, it may prevent the direct investigation of the impacts of period specific macro-economic or policy shocks.

4.3 Oversampling

One response to the potential problem of small sample sizes amongst key groups of policy interest would be oversampling of such groups. There are two issues which will require careful consideration before such a policy is embarked on: what characteristics are appropriate bases for oversampling? How should it be achieved?

There are relatively few conceptual problems in oversampling in a panel study on the basis of an invariant characteristics such as sex, year of birth or ethnicity, though there may be implementation problems, discussed below. However it is unlikely that such factors will alone enable the creation of large enough subsamples for all policy relevant groups. Oversampling on the basis of characteristics which vary over time and are of direct policy interest cause different problems. Oversampling on the basis of, for example poverty status, may be appropriate where the policy interest is in transitions out of that state.²² However if the concern is also with conditions leading to transitions into

²² Though it would be necessary to collect initial conditions data to establish when the respondent entered that state.

the state, then such oversampling is unlikely to be appropriate. In this case it would be more appropriate to use factors known to increase the risk of the state. If researchers designing the study are confident in being able to identify such factors then oversampling may lead to a more economic overall design, provided implementation problems noted below can be resolved. However if these assumptions were incorrect the sample might not be adequate for identifying the real factors.

There are two approaches to oversampling: through selection and weighting of the primary sampling units (PSUs) and through screening. PSU selection and weighting would undoubtedly be cheaper. The preferred method would be to draw a sample of PSUs in the normal way, and then use a higher sampling fraction in those containing higher proportions of the groups to be oversampled. This is clearly only practicable where there is adequate sampling frame information to determine the relative proportions of target groups. This method has one minor and one major drawback. The minor problem is that it will lead to uneven interviewer loads, which may present some fieldwork management problems. More significantly, it may lead to an unrepresentative subsample of the target group. This will arise if members of the target group (for example those at risk of poverty) found in high concentration areas differ systematically from members of the group found in low concentration areas or are exposed to different constraints and opportunities. This seems not at all implausible—e.g. we would expect systematic differences in the rate of arrival of new job opportunities. We would therefore argue against such an approach unless some method of controlling for the area differences was available.

The alternative of screening is therefore to be preferred, though it is likely to a higher cost alternative. There are three approaches: (a) to use a short screening interview for all respondents; (b) to issue a main sample and then a supplementary sample containing a screening interview to collect the oversampled group; (c) to use wave one as the screen, i.e. to collect full information from a larger initial sample, and then to select from this for continuation in the panel. The choice amongst these options would depend on the relative size of the oversample, and how easily the screening information can be collected. Thus e.g. if income was going to be a screening variable, then it would be more appropriate to use method (c), while if relatively simpler factors were to be used (e.g. household structure, labour force status, ethnicity) then methods (a)

or (b) would be possible. If the oversample is to be small in relation to the main sample, then method (b) might be preferred, but if relatively large method (a) may be preferable.

It should be noted that oversampling will lead to some inflation of variances for whole population estimates, owing to the large spread of weights that will result.²³

4.4 Following rules

The standard following rules for household panel studies have been outlined in section 2. Here we deal a number of more specialised issues.

Non-sample members The conventional design interviews non-members of the original sample when they are living with an original sample member (or a child of an original sample member), but not when they cease to do so. A number of panels have explored the possibility of continuing to follow these respondents, given that long runs of data may already have been collected from them, and GSOEP changed to doing so in wave seven. To do this adds to survey costs, and there must be a clear research rationale for following non-sample members. One of the major difficulties is in defining selection probabilities, and hence weights for including them in analysis. On balance it must be preferable to devote resources to increasing the original sample size and maximising panel response rates, rather than following non-sample members. There may however be situations where non-sample members can provide contextual information useful in the analysis of continuing sample members. This is rationale behind the BHPS decision to follow non-sample parents of new panel member births, even where they split from their partner.

Births There are two methods of including new births to panel members, to treat births where the mother is a panel member as a full sample member, or include as full sample members births where either parent is panel member. The former method should result in probability sample of births to the population, while the latter will result in some oversampling, though this may be desirable.

Immigrants There is a case for treating immigrants to the population since the start of the panel as full sample members if they join a panel household, since their selection probability is simple to calculate. ‘Immigrants’ here would also include those who moved out of institutions, if these were not sampled at wave one. This method of

²³ For further discussion of the issues discussed in this section, see Citro and Kalton (1993, chapter 4).

inclusion will not of course lead to a random sample of all immigrants since the start of the panel.

Emigrants It is important to retain emigrants within the sample at each wave, and collect what tracking information is available, so that any panel members returning to the survey country may be reinterviewed.

The institutional population While cross-sectional surveys frequently exclude the institutional population this is not necessarily appropriate in the case of a panel study. At the minimum it is necessary to track entries of panel members to institutions and to reinterview them when they leave. Serious consideration should also be given to collecting interviews from panel members in institutional settings, either directly or by proxy with a family member. Panel studies have been able to carry out interviews in some types of institution, such as university accommodation and old people's homes, though others, such as prisons and hospitals have proved less accessible.

Wave non-respondents It will be necessary to establish rules for continuing to attempt interviews with panel who have refused interviews or not been contacted at some previous wave. In principle every effort should be made to follow all such cases at subsequent waves. Of course ethical issues arise in seeking to persuade respondents who have made it absolutely clear that they have no wish to participate further. Wave non-response does create certain problems for data analysis, since their data can only be used for a limited subset of analyses, requiring in principle multiple sets of weights. Another option is to impute data for these missing waves. SIPP has done this, but it is relatively costly in terms of researcher time.

4.5 Instrument design

The most important decision in instrument design is clearly the scope of the panel study; whether it is to cover a wide range range of policy areas or a relatively narrow range. Given the need to cover each in adequate depth, the more areas covered the longer the instrument and the greater the respondent burden. On the other hand it would be a mistake to make the focus too narrow, both because one of the major benefits of panel analysis is the ability to relate change in different domains. Given the client's overall objectives coverage of income receipts and welfare programme participation, labour market transitions and job characteristics, household transitions, education and training seems to follow naturally, including full coverage of changes in

the period between each wave for each area. This would come close to exhausting a feasible questionnaire length, though some consideration might be given to including some questions on health related events, and there could be some coverage of other areas through questions only asked at a small number of waves.

A number of other considerations will drive instrument design.

First, at the heart of the panel questionnaire design should be a set of core questions to be repeated at each wave. The selection criteria here will be their policy and research importance, and the likelihood of change. Change in question wording at subsequent waves should be resisted as far as possible.

Second, at the start of the panel it is highly desirable to collect initial conditions information, i.e. sufficient information to establish the process by which the panel member arrived at their current state, including the dates of entry. This might be a full lifetime history across a range of domains, or it might be sufficient to establish the dates of start of various types of state (e.g. labour force status, place of residence, marital status, welfare reciprocity status). Such information is necessary to reduce problems of left censorship in analyses of transitions.

Third, as far as possible it is desirable to collect continuous information about change between each panel wave, rather than just the situation at each panel. This is necessary to establish the ordering of events and changes, and to have complete information on multiple changes which may occur between panel waves. For reasons relating to the estimation of measurement error, it is desirable that the reference period for this continuous information should overlap with that used at the previous wave.

Fourth, given the necessity of establishing the representativeness of the panel, some repetition of critical questions from other national social surveys is important to enable benchmark comparisons.

Fifth, at the second and subsequent wave it is desirable to feed forward information about responses at the previous wave. This helps to reduce substantial changes at the seam between panel. Outside a computer-assisted interviewing environment this feed forward is relatively complex to implement, and it needs to be thought about in the very early stages of design.

The first four of these factors will all serve to increase interview length, and therefore respondent burden. There is no very clear evidence on the impact of respondent burden on subsequent response rates, though it would appear likely that

some effect will exist. The BHPS has one of the longer interviews for a household panel study, with a main individual schedule lasting approximately forty minutes. The average contact time per individual, including collection of household level information, a self-completion questionnaire and tracking information is around an hour. This would appear to be around the tolerable limit for maintaining high response rates in the long term.

Sensitivity of questions is another concern. Even if income questions are sensitive and subject to high non-response, it is undesirable to omit these from early waves until the panel is established, as sometimes suggested. This is because attrition is likely to be highest between wave one and wave two in any case, and income information will be lost for all those panel members who drop out at this point. On the other hand it may be desirable to delay introduction of questions which are still more sensitive. Wealth and assets are an example for a number of countries.

Given the importance of the first wave, noted above, it is critical that design issues are fully resolved before this wave, implying a significant period of development and pretesting.

4.6 Issues of timing

A number of issues of timing arise, relating to the intervals between waves, the timing of interviews within each wave and the length and timing of the fieldwork period. Clearly, the interval between waves will have a major bearing on questionnaire design (e.g. in defining the reference period for retrospective questions).

The interval between waves will be driven by questions of cost, respondent burden, likelihood of recall problems, and the frequency of transitions of interest. The great majority of household panel studies have adopted a one year interval. Apart from its conventional appeal, this appears mainly to be driven by a focus on annual measures of variables of interest, especially income, and especially in countries which use self-assessment income tax systems. It seems likely that a longer interval would lead to a significant deterioration in the quality of some critical data, though it might be acceptable if the interest was primarily in labour market transitions. The most important example of a panel using a shorter interval is SIPP, which interviews every four months. The main concern driving this design is precision of monthly estimates of income level and welfare programme participation. Such a design is clearly costly. There is some

trade-off between response burden and panel maintenance. More frequent interviews will increase burden, but are likely to minimise losses due to non-contact with movers. These losses can be very severe in panels with intervals longer than one year.

While cross-sectional surveys often tend to spread their interviews over a whole year, there are strong arguments for having a concentrated field period in a panel study. Seasonality is a less serious problem in panel measures than in cross-sectional estimates, but also because of the substantial organisational load in running a continuous panel, since it would be likely that multiple waves would be in the field at the same time.

Panel fieldwork is likely to be of relatively long duration. While the majority of interviews can normally be collected over one or two months, the experience of most panels is that there is a relatively extended follow-up period to trace and interview movers and attempt to convert refusers.

The timing within the year of the initial field period may not be of critical importance, unless it is intended to collect information for a specific reference period (e.g. a calendar year), or it is believed that respondents will have better access to specific types of information at different points in the year (e.g. SLID collects its income information shortly after tax returns are due for completion).

4.7 Maintaining response rates

For a panel study to be viable in the long term high response rates are critical at each wave. First wave response rates are particularly important in order to ensure overall population representativeness, and this may be particularly critical if some of the groups of highest policy relevance are generally subject to low response rates in surveys. Attrition after the first wave may introduce biases, but even if it does not it may reduce the available sample in some subgroups sufficiently to vitiate analysis. It is possible to control for attrition bias using weighting or other techniques provided the attrition is related to known characteristics at a wave of interview, but this is difficult if the attrition is related to some change in characteristics between waves. The most important mechanism for controlling attrition bias must be to maximise response rates through fieldwork and panel maintenance procedures.

There has been significant variation in response rates between surveys. This reflects differences between countries and also differences between government and other surveys. Some indication of these response rates has been given above. The

figures quoted for PSID should be regarded as a realistic target. Maintaining these response rates will require consideration of a number of activities:

Interviewer selection and training Some additional training is likely to be necessary in refusal conversion methods in the context of a panel study. It may also be desirable to use ethnic minority interviewers to maximise response rates in minority communities.

Maintaining interviewer continuity There is at least anecdotal evidence from a number of panel studies that having the same interviewer returning each year to the same household tends to maximise response rates.

Collection of additional information to facilitate tracking of panel members who move In some countries with population registers this has not proved necessary, but where these do not exist some additional contact information is likely to be necessary.

Contacts with respondents between waves Some contact, usually in the form of mailings containing further information about the survey, together with requests to return information about movers, proves helpful in maintaining panel members' interest in the survey and in updating information about respondent locations.

Incentives Consideration should be given to giving panel members monetary incentives to participate, in recompense for the burden of participation in a long term survey.

4.8 Interview mode

Choices of interview mode (face-to-face or telephone, traditional or computer assisted) depend heavily on country specific factors, though certain ideals for a panel study can be specified. The main argument in favour of telephone interviewing is lower cost. However in a country with less than 100% telephone coverage it is unlikely to be acceptable as a mode for wave one. What should be considered seriously is the use of telephone interviewing from wave two onwards for that part of the sample where it may be acceptable, or even preferable. There are two strategies here, the approach followed by PSID, where the great majority of the panel is interviewed by telephone, with face-to-face interviewing only for that part of the panel for whom telephone interviewing is difficult or impossible, or the approach of GSOEP, where face-to-face is the principal mode, but telephone interviewing is available as an alternative. In the first case the main purpose is to lower overall costs, while in the latter case it is to raise response rates. The

use of telephone interviewing has major instrument design implications. In particular it reduces the acceptable interview length, with a maximum of perhaps 30 minutes. If mixed modes are used there is a further choice between a common instrument and a short form for telephone interviews. The latter presents analysis problems.

There are major advantages in using computer assisted interviewing in a panel study, quite apart from their probable data quality advantages, savings in processing costs and more rapid delivery of final data, which would apply to all surveys. The more rapid delivery of data is itself an advantage, since it allows more complete data cleaning and validation for each wave before the next wave begins. It also permits relatively sophisticated feeding forward of data from the last wave, which may reduce spurious apparent change, and with some designs reduce interview lengths. Computer assisted interviewing is however relatively complex to develop, and whether it was appropriate for a new panel would depend on whether there was a substantial body of experience with these methods on other surveys. We would advise against using a panel as the first major testbed for the development of computer assisted interviewing, given in particular the serious, often irretrievable, consequences of errors at the first wave.

4.9 Panel conditioning

Some evidence exists for changes in behaviour or reporting practices as a result of participation in attitudinal and voting panel surveys (Traugott and Katosh, 1979; Waterton and Lievesley, 1989). However the evidence for such panel conditioning is much more inconclusive for household panel studies focussing on economic behaviour (Kalton *et al.*, 1989). In principle problems are likely to be greatest where the intervals between waves are relatively short, and where the subject matter is highly specialised.

However it is necessary to be sensitive to the issue in questionnaire design. One particular problem to which it is necessary to be sensitive in questionnaire design is introducing too many questions which open long sequences of follow-up questions. These are difficult to avoid, but respondents may learn over time to trim their answers to the first question to avoid the follow-up.

Some aspects of panel conditioning may be positive advantages. It may be possible to train respondents to have available documentation relating to complex questions, especially income, improving data quality.

4.10 Measurement error

Given that the main focus of panel studies is on the measurement of change, the problem of measurement error may be serious. Random errors in a cross-sectional survey do not bias population estimates, but random errors in two repeated measurements will give rise to over-estimates of the extent of change. Errors in the recall of dates may give rise to incorrect chronological ordering of events, and therefore perhaps misplaced inferences about causal ordering. Measurement error is probably the most serious problem confronting the collection and analysis of data about change, and one for which no solutions are wholly satisfactory. The two main approaches to the problem are to avoid it through instrument design, and to measure the extent of the problem.

Probably the best solution is to feed forward as much data as possible from the previous wave so as to reduce spurious reporting of change. However there is some evidence that such strategies err somewhat in the opposite direction, and suppress reports of actual change (Hill, 1994). This effect is undoubtedly less than excessive reported change which comes of not feeding forward. It is also, as suggested above, desirable to have some overlap in the reference periods for retrospective information collected in successive waves (for example the account of labour market spells over the past year should cover the previous interview date, so that inconsistencies can be identified. This is probably better than the practice of SIPP, which has no overlap, so that large changes appear to happen at the seams between survey reference periods.

It is easier to estimate what effect measurement error may be having on estimates of change if there is some knowledge of the extent of the error. This may be provided by some of the techniques outlined in the previous paragraph. There are also a number of statistical modelling techniques for estimating measurement error. In some circumstance it may be worth undertaking a specific validation study to estimate the extent of error. These can be relatively costly to mount, but it may be possible to use linkage to administrative records to validate some of the data collected.²⁴

4.11 Data processing and data management issues

Here we focus only on data processing and data management issues for a panel which are substantially different from those met in the course of cross-sectional surveys.

²⁴ See e.g. the PSID validation studies summarised by Hill (1992).

The timetable for data processing will depend in part on the needs of the next wave of the panel, as well as the requirements of data analysis. The sample to be issued at any wave after the first will depend on the outcomes from the previous wave, and therefore the basic integrity of these data must be assured before the next wave is issued. Three other issues are covered here.

Cross-wave data checking Decisions must be made on the extent to which inconsistencies between data collected at different waves should be resolved. This will depend in part on the nature of the final users and the resources available. Given that most inconsistencies cannot be resolved with certainty, more sophisticated researchers would probably prefer to have the inconsistency left in order to allow analysis-specific treatments. Other users may be content for some resolution of inconsistencies to be made prior to data release. Cross-wave checking is relatively costly. But it is worth ensuring some consistency in basic demographic variables such as sex, year of birth at the very least.

Database design Household panel data are relatively complex, and the design of the database to hold them is an important step. The central complexity is that while single wave data forms a natural hierarchy of persons within households, multiple waves do not fit into a single hierarchy. This is because while individual persons remain the same, household units change. This means *inter alia* that there must be two independent sets of identifiers in the database: wave specific identifiers linking individuals to households, and cross-wave individual identifiers. Different panel studies have come up with rather different solutions to the database design problem, and there would need to be some review of the options and consultation with final data users.

Imputation Item non-response in panel studies will limit the number of cases available for analysis, but if substantial it may also impart biases. Many panel studies have therefore imputed missing data on critical variables subject to non-response (e.g. income). There are a great many imputation techniques available: see Kalton (1983) for an account of some of these. It is important to note that some of the simpler techniques, e.g. mean substitution within sub-class, which may be adequate for cross-sectional population estimates are not satisfactory in panel studies. They tend to reduce apparent variances of estimates, and may create spurious estimates of change if no account is taken of values of the same variable at other waves. Cross wave imputation using

techniques which preserve the error structure of the original variables are to be preferred, though they are clearly somewhat more complex.

4.12 Weighting

Good weighting strategies are critical to achieving population estimates from a panel study. There is some argument about the necessity for weighting in dynamic models, but it will certainly be required both for cross-sectional estimates and for panel estimates concerned with frequency of events or transitions. For weighting wave one may essentially be treated as a cross section, and best available methods used in cross-sectional weighting should be employed. Thus depending on what information is available from the sampling frame, from population estimates and about non-respondents to the first wave, the weights will be based on some combination of weighting for unequal selection probabilities, non-response adjustment and post-stratification.

After wave one it is necessary to adjust for losses through attrition (and in some instances, gains through new entrants). There are essentially two types of weights necessary—weights for longitudinal analyses and weights for cross-sectional analyses of each wave. In principle the two could be the same, but it is likely that there will be a significant number of individuals who can be included in the separate cross-sections, but not in longitudinal analysis. The latter, clearly, can only use respondents at all the separate waves to be included in the analysis, while the former can include all respondents in the single wave of interest. At least one of each of these weights will require calculation for each wave.

Both these weights can be constructed either as non-response adjustments or through post-stratification. It is likely that the former will be more important. This is in part because very much more is known about non-responders after the first wave than is typically known about non-response to a cross-sectional survey. Thus the models which will feed into both the longitudinal and the cross-sectional weights can be very sensitive to the particular patterns of attrition which have taken place. The only exception to this is where attrition is related to changes since the previous wave (rather than states at the previous wave). One such association—higher losses among movers—is likely, but this is a change which is likely to be known, so can be incorporated in models. Weights are normally constructed by analysis of attrition

patterns to identify weighting classes which are particularly predictive of non-response. The weight component is then simply the inverse of the response probability, and the final longitudinal weight at any wave will be the product of the sequence of individual wave adjustments up to that point.

Longitudinal weights adjust panel members who have responded at all waves up to the latest wave for those panel members who drop out before that point. By definition, non-sample members cannot be given weights in this manner (since they did not respond in wave one). Other original sample members with an intermittent pattern of response are also not included. It is possible to include such cases, but this will require calculation of separate attrition weights for each combination of waves to be analyzed. Cross-sectional weights can be calculated for any wave for all original sample members surviving to that wave, irrespective of their previous response history, and the weight would normally be calculated on the basis of a single response adjustment up to the current wave, rather than the sequence of individual wave response adjustments used for longitudinal weights. It is also possible to give new entrants such weights. The method known as 'fair shares' is used in SLID and BHPS. It is discussed in Ernst (1989) and Lavalley and Hunter (1992).

A number of other issues are liable to arise in weighting panel data, for example the treatment of births, and of deaths and moves out of scope, the treatment of children as they become full respondents, and the treatment of within household non-reponse. There will also be a number of technical questions, such as the appropriate spread of weights, and rescaling. In addition the question of whether to produce separate weights for panel members with an incomplete response history will have to be addressed—this issue is also known as that of wave non-response. In addition to the references cited above, the following sources may also provide some guidance: Kalton (1983), Kalton (1986), Lepkowski (1989), Lynn et al (1994).

4.13 Implications for survey organisation

The characteristics outlined above have three major implications for the organisation which carries it out.

First, the critical importance of success in the first wave means that the panel survey must have a relatively *long development phase to perfect the survey instrument and the study design*. It is also necessary to give proper thought at this early stage to

issues that likely to arise later. Examples would be the need to prepare in advance for a switch to telephone interviews; the need to design the instruments with the potential for future feed forward of data to the next wave in mind; the need to anticipate opportunities for methodological validation, for example linkage to administrative records.

Second, there will be a *relatively complex timetable of activities involved in running the panel*. In particular, some elements of data processing from each wave must be completed before the issue of the sample for the next wave. The more data there is to be fed forward the more complete this must be. For wave one, where the fieldwork process is relatively short, this presents few problems (on the assumption of an annual panel), but for subsequent waves where fieldwork may have an extended period for follow-up, some of the deadlines become rather tighter. The separate stages of activities from a number of waves will of course tend to overlap with one another, but except for the interdependence of data processing this is not in principle different from the situation of repeated cross-sectional surveys.

Third, there is a need for *continuity of experience of staff* involved with the panel survey. The importance of using the same interviewers at each wave has already been mentioned, but some of the panel maintenance and data processing skills are also relatively specialised, and there is considerable benefit in building and maintaining a core of staff with detailed knowledge of the operations of the panel.

4.14 Costs relative to cross-sectional surveys

Many aspects panel study costing can be determined from estimates of the costs of cross-sectional surveys of equivalent sample size and instrument length, and such costs are probably the best starting point for estimating the costs of a new panel survey. However, there are a number of factors which tend to inflate the costs of a panel, as well as a number which may serve to reduce them. How these would affect a new panel will depend in part on design decisions (e.g. in the use of telephone interviewing), in part on the structure of survey costs (e.g. the relative balance between travel costs, interviewer remuneration, and central administration costs), and the size and structure of the fieldwork organisation (e.g. how far the overheads of panel maintenance can be absorbed within the existing organisation or require additional investments).

Three factors may serve to increase costs relative to a cross-sectional survey.

First, population mobility will lead to a declustering of the original sample, removing some of the economies in interviewer travel costs and time created by the original spatial clustering. The impact of this factor will depend on the nature of migration patterns (i.e. the balance between short distance and long distance migration) and the density of population of interviewers trained to work on the survey. For example, New Zealand 1991 population census estimates indicate that around 44% of the population had moved over the previous five years. If this movement was mainly short distance within urban areas, the effects on costs would be modest, but if a significant proportion was longer distance and took respondents to localities remote from other sampling points the effects could be more substantial. Simple random sample costs would be an upper bound on the potential cost escalation, but panel costs would be unlikely to approach very close to this level.

Second, it has been stressed above that high response rates at each wave are critical to the long term viability of a panel study. Response rates do depend in part on survey effort, and hence survey costs. A higher than normal budget may therefore need to be set aside to allow for interviewer training, multiple calls, and refusal conversion.

Third, panel maintenance also requires continuing year round activities to maintain contact with respondents, to make repeated mailings, to track movers as suggested above. It may also be necessary to consider paying incentives to respondents to maintain long term response rates.

Two factors may lead to reduced costs of a panel (at least after wave one).

Once a panel is established, the effort required to contact the great majority of respondents is reduced. Wave-on-wave response rates are much higher than in a fresh cross-sectional survey, less effort is required in locating addresses, and interviewers are likely to need to make fewer calls to achieve their interviews. It is difficult to quantify these effects precisely, but the other related consequence of having made a contact with the panel member is to make it feasible to move to telephone interviewing. This could reduce costs substantially.

Depending on the purposes of the survey, and the capacity to feed forward information from one wave to the next, it may be possible reduce the length of interviews after initial conditions data are collected at the first wave.

4.15 Comparison between single panel and multiple rotating panels

Once it has been decided to have some sort of panel survey, the main design choice is between a single panel with an indefinite life²⁵ or a series of limited life panels with some overlap between successive panels. There are of course many design options if a rotating panel structure is decided on—relating to the length of individual panels, the degree of overlap, and the interval between waves.

The advantages and disadvantages of the two designs are largely complementary. The major distinction is one of cost. It is not practicable for most purposes to combine observations from the separate panels in a rotating structure.²⁶ This means that the each separate panel in a rotating structure will need to be of an adequate sample size to carry out analyses on its own. It is likely then that a rotating structure would imply individual panels of the same size as a single indefinite life panel. Thus a rotating structure in which there were always two panels in existence would imply costs almost twice those of a single panel—there would probably be some savings in administration and data processing overheads.

The central advantage of the rotating structure is that at any point in time there will exist a panel which has been subject to relatively little attrition and panel conditioning, minimising uncertainties about representativeness with respect to the current population. A recently drawn panel will also contain almost complete coverage of the immigrant population. Such a structure is therefore highly desirable if the aim is a set of period-specific estimates of transition rates. An older single panel can provide such estimates, but much greater care is required to ensure that attrition effects are not biasing measures of change. The fact that there is a relatively fresh panel available will also mean that attrition will not have reduced the sample numbers in policy relevant subgroups.

If there are overlapping panels available it is also possible to make estimates of the impacts of different levels of attrition on panel estimates.

²⁵ Though of course it may be terminated as result of becoming unrepresentative or budget exigencies—the point is that it is not designed to have a limited life.

²⁶ This is because the panels will represent different populations, and will be subject to different levels of attrition, making combined weighting problematic, and because very different amounts of information are likely to be available.

One further advantage of a set of limited life panels is that it is feasible to impose rather higher respondent burdens, for example more frequent interviews, in the knowledge that long term attrition is not a concern. It is also possible to include more sensitive questions in final waves. Of course these advantages are only available if the panels are intended to have a short life.

The potential disadvantage of the rotating structure (aside from cost) is that the observation period for any individual is likely to be relatively short. This may make it impossible to collect sufficient information on both the antecedents and the consequences of events, unless substantial retrospective data is collected. Moreover only shorter spells in some states will be completely observed, leading to severe biases where complete spell information is required. This suggests both that careful attention should be paid to initial conditions data, and that the panel length should be sufficiently long. It appears that effective length of SIPP (32 months) is too short for many purposes, and the SLID length is more appropriate. The recent SIPP review panel (Citro and Kalton, 1993) recommended lengthening each SIPP panel's life, citing e.g. the need to get more information about long-term receipt of welfare benefits.

It has been suggested to us that the choice between a single household panel design and a rotating panel design is one which can be deferred. Although this may be true in principle, it may not be so in practice. The problem is that there are reasonably long lead times involved in organisation (see section 4.13), and so choices cannot be deferred very long, and especially if a dual interview design like the SLID were adopted.

5. SUMMARY

- Longitudinal data may be of several types: retrospective, panel, and record linkage. Some longitudinal information can also be derived from repeated cross-section surveys.
- The relative advantages of each type of data set depend on which issues have the highest priority for those seeking the new survey and the potential data users.
- Where information about income dynamics is a minimum requirement, a panel survey is required.
- The subsequent choice between a single indefinite life panel design or a multiple limited life panel design is a decision best made in the light of specific information about the aims of the research and related policy purposes, and about design and budget constraints.

6. REFERENCES

- Ashenfelter, O. (1978) 'Estimating the effect of training programs on earnings' *Review of Economics and Statistics* 60, 47-57.
- Bane, M.J. and Ellwood, D.T. (1986) 'Slipping into and out of poverty: the dynamics of spells' *Journal of Human Resources* 21, 1-23.
- Barreiros, L. (1995) 'The European Community Household Panel (ECHP): its design, scientific and policy purposes' *Innovation: The European Journal of Social Sciences* 8, 41-52.
- Blank, R.M. and Ruggles P. (1994) 'Short-term recidivism among public-assistance recipients' *American Economic Review* 84(Papers and Proceedings), 49-53.
- Bingley, P., Symons E., and Walker I. (1994) 'Child support, income support and lone mothers' *Fiscal Studies* 15, 81-98.
- Björklund, A. and Moffitt R. (1987) 'The estimation of wage gains and welfare gains in self-selection models' *Review of Economics and Statistics* 69, 42-49.
- Buck, N., Gershuny, J., Rose D., and Scott, J. (1994) *Changing Households: the British Household Panel Survey 1990-92*. Colchester: ESRC Research Centre on Micro-Social Change, University of Essex.
- Burkhauser, R.V. (1991) 'An introduction to the German Socio-Economic Panel for English speaking researchers', Cross-national Studies in Aging Program Project Paper No. 1, Maxwell School, Syracuse University NY.
- Citro, C.F. and Kalton, G. (1993) *The Future of the Survey of Income and Program Participation*. Washington DC: National Academy Press.
- Coxon, A.P.M. and Corti, L. (eds) (1991) 'Sample Design Issues in a Panel Survey: the Case of the British Household Panel Study', Working Papers of the ESRC Research Centre on Micro-Social Change, Paper 3. Colchester: University of Essex.
- Deaton, A.S. (1985) 'Panel data from time series of cross-sections' *Journal of Econometrics* 30, 109-126.
- Duncan, G.J. (1992) 'Household panel studies', unpublished paper, Institute for Social Research, University of Michigan.
- Ermisch, J.F. (1991) *Lone Parenthood: An Economic Analysis*. Cambridge: Cambridge University Press.

- Ermisch, J.F. and Wright R.E. (1991a) 'The duration of lone parenthood in Great Britain' *European Journal of Population* 7, 129-158.
- Ermisch, J.F. and Wright R.E. (1991b) 'Employment dynamics among British single mothers' *Oxford Bulletin of Economics and Statistics* 53, 99-122.
- Ermisch, J.F. and Wright R.E. (1991c) 'Welfare benefits and lone parents' employment in Great Britain' *Journal of Human Resources* 26, 424-456.
- Ernst, L.R. (1989) 'Weighting issues for longitudinal household and family estimates' in *Panel Surveys* (D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, eds.), New York: Wiley.
- Eurostat (1994) *European Community Household Panel, Survey Questionnaires (Community version) Wave 1* Doc PAN 14. Luxembourg: Statistical Office of the European Community.
- Ferri, E. (ed.) (1993) *Life At 33: the Fifth Follow-up of the National Child Development Survey*. London: National Children's Bureau.
- Fitzgerald, J. (1991) 'Welfare durations and the marriage market: evidence from the Survey of Income and Program Participation' *Journal of Human Resources* 26, 541-561.
- Harala, R. (1994) 'Evaluation of the results of the register-based population and housing census 1990 in Finland' *Statistical Journal of the United Nations ECE* 11, forthcoming.
- Heckman, J.J. (1981) 'The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process' in *Structural Analysis of Discrete Data with Econometric Applications* (C. Manski and D. McFadden, eds). Cambridge MA: MIT Press.
- Heckman, J.J. and Robb R. (1985) 'Alternative methods for evaluating the impact of interventions' in *Longitudinal Analysis of Labor Market Data* (J.J. Heckman and B. Singer, eds). Cambridge: Cambridge University Press.
- Heckman, J.J. and Wolpin K. (1976) 'Does the Contract Compliance Program work? An analysis of Chicago data' *Industrial and Labor Relations Review* 29, 554-564.
- Hill, M.S. (1992) *The Panel Study of Income Dynamics: a User's Guide*. Newbury Park CA: Sage Publications.

- Jenkins, S.P. (1995) 'Easy ways to estimate discrete-time hazard models' *Oxford Bulletin of Economics and Statistics* 57, 129-138.
- Kalton, G. (1983) 'Compensating for missing survey data', Research Report Series, Institute for Social Research, University of Michigan.
- Kalton, G. (1986) 'Handling wave non response in panel surveys' *Journal of Official Statistics*, 2, 303-314.
- Kalton, G. and Citro, C.F. (1993) 'Panel surveys: adding the fourth dimension' *Survey Methodology* (Statistics Canada Cat 12-001) 19, 209-215. Reprinted in *Innovation: The European Journal of Social Sciences* 8, 1995, 25-39.
- Kalton, G. Kasprzyk, D and McMillen, D.B. (1989) 'Nonsampling errors in panel surveys' in *Panel Surveys* (D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, eds.). Wiley: New York.
- Lavalley, P and Hunter, L. (1992) 'Weighting for the Survey of Labour and Income Dynamics', *Proceedings of the Statistics Canada Symposium 92, Design and Analysis of Longitudinal Surveys*, 65-74.
- Lepkowski, J.M. (1989) 'Treatment of wave non-response in panel surveys', in *Panel Surveys* (D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, eds.). New York: Wiley.
- Lynn, P., Purdon, S., Hedges, B., and McAleese, I. (1994) *An Assessment of alternative weighting strategies*, Youth Cohort Survey Report 30, Employment Department, Sheffield.
- Martin, J. and Roberts, C. (1984) *Women and Employment: a Lifetime Perspective*. London: H.M.S.O.
- Martin, J. and Roberts, C. (1984) *Women and Employment: Technical Report*. London: H.M.S.O.
- Moffitt, R.E. (1993) 'Identification and estimation of dynamic models with a time series of repeated cross-sections' *Journal of Econometrics* 59, 99-123.
- Peters, H.E. (1988) 'Retrospective versus panel data in analyzing life cycle events' *Journal of Human Resources* 23, 488-513.
- Shaw, A., Walker, R., Ashworth, K., Jenkins S., and Middleton, S. (1995) 'Moving off Income Support? Obstacles, opportunities, choices' (Report to the U.K. Department of Social Security). Paper CRSP246, Centre for Research in Social Policy, Loughborough University.

- Smith, H. (1992) 'Longitudinal surveys', unpublished paper, Statistics New Zealand.
- Starck, C. (199?) 'The Finnish longitudinal census file 1970-1990' in F. Finnas (ed.) *Language Shifts and Migration*, Studies 109. Helsinki: Statistics Finland.
- Statistics Canada (1994) 'First wave rolls in' *Dynamics* 3 (September), 1-3.
- Statistics Finland (1991) 'Economic activity of the population' Population Census 1990:1. Helsinki: Statistics Finland.
- Statistics New Zealand (1993) *Incomes Statistics 1992*, Cat. 01.024.0092. Wellington: Statistics New Zealand.
- Statistics New Zealand (1994) *New Zealand Official Yearbook 94*, Cat. 01.001. Wellington: Statistics New Zealand.
- Stevens, A.H. (1994) 'The dynamics of poverty spells: updating Bane and Ellwood' *American Economic Review* 84(*Papers and Proceedings*), 34-37.
- Traugott M.W. and Katosh, J.P. (1979) 'Response validity in surveys of voter behaviour' *Public Opinion Quarterly*, 43, 359-377.
- U.S. Bureau of the Census (1991) *Survey of Income and Program Participation Users' Guide, second edition*. Washington DC: Bureau of the Census.
- Wagner, G., Burkhauser, R., and Bheringer, F. (1993) 'The English language public use file of the German Socio-Economic Panel' *Journal of Human Resources* 28, 429-433.
- Waterton, J. and Lievesley, D. (1989) 'Evidence of conditioning effects in the British Social Attitudes Panel Survey' in *Panel Surveys* (D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, eds.). New York: Wiley.
- Webber, M. (1994) 'Survey of labour and income dynamics: an overview' in *Report 1994*, Cat. 75-201E. Ottawa: Statistics Canada.