The role of household panel surveys in evidence-based policy

This is a revised version of the paper presented to the Australasian Evaluation Society International Evaluation Conference, 2009, 31 August–4 September 2009 Canberra Australian Capital Territory.

Disclaimer

The opinions, comments and/or analysis expressed in this document are those of the author and do not necessarily represent the views of the Minister for Families, Housing, Community Services and Indigenous Affairs and cannot in anyway be taken as expressions of Government policy.

The role of household panel surveys in evidence-based policy

Abstract

Evidence-based policy and practice have now become a catchphrase in public policy in many countries, both developed and developing. As part of planning for services and monitoring and evaluating programs, policy-makers are now demanding reliable, valid, relevant and meaningful data for policy decisions. This paper explores the contribution of household panel surveys to the development of knowledge and evidence on which to base policy decisions. It discusses the strengths and weaknesses of panel surveys compared to cross-sectional surveys in providing evidence to assess the social and economic well-being of the population and its subgroups. An increasing number of countries now invest in household panel surveys to gain not only snap-shots of the population, but knowledge of the changes in socio-economic and behavioural characteristics of the people at the *individual (*or micro) level as opposed to *aggregate* (or macro) *level*. Drawing from the published data from the Household, Income and Labour Dynamics Australia (HILDA) Survey, and focusing on selected key social indicators, this paper will demonstrate how panel data analysis can contribute to informing policy.

Introduction

Evidence-based policy is an approach that helps people to make 'well informed decisions about policies, programmes, and projects by putting the best available evidence from research at the heart of policy development and implementations' (Davies, 2004:7)[1], although public sector policies are often not solely based on evidence. This approach is a shift away from opinion-based policies to a vigorous, rational approach that gathers, critically appraises and uses high quality research evidence to inform policy-making and professional practice (Davies, 2004). Evidence-based policy emerged as the centrepiece in policy-making and governance in Western societies in the later part of the 20th century. As part of this changing policy environment, governments began to look for satisfactory evidence to judge the impact of public policies and social programs on the marketplace outcomes and economic well-being of their citizens (Burkhauser and Smeeding, 2001). Accordingly, governments have been demanding better-quality evidence with the capacity to identify good and bad consequences of public policies and programs to inform voters and policy decision-making (Henry, 2009).

Government statistical agencies and the research community responded to this demand by producing cross-sectional surveys, with repeated elements in some surveys, to provide evidence for sound policy-making and assessment of program impact. The research community, using representative samples of cross-sectional data, provided snapshots focusing on priority areas of social and economic concerns, and using micro-simulation modelling methods, attempted to provide quantitative assessment of the effect of policy changes on the well-being of the people (Citro and Hansheck, 1991).

While research based on cross-sectional data helped to describe snapshots, and with repeated cross-sectional data, to measure broad trends at the macro level, they did not assist in understanding the dynamic aspects of the population change at the individual (micro) level. This limitation reflected, in part, the lack of capacity to theorise social change and in part the non-existence of data that could describe changes in the socio-demographic and economic characteristics of individuals and households (Rossini, 2002).

The recognition of this limitation paved the way for investing in longitudinal social surveys and more importantly household panel surveys. Although panel studies on special topics had been conducted before the 1970s, panel surveys on a large scale, based on nationally representative samples, emerged around the 1970s in the US, and in many European countries in the 1980s (Duncan, Juster and Morgan, 1987) or later. These include the US Panel Study of Income Dynamics (PSI), 1968; the British Cohort Study (BCS) 1970; the British Household Panel Survey (BHPS, 1991); the German Socio-economic Panel (GSOEP) 1984; the Canadian Survey of Labour and Income Dynamics, (SLID, 1993); and the European Community Household Panel (ECHP) 1994. In Australia, there are a number of longitudinal panel surveys, but the most widely used panel survey is the Household, Income, Labour Dynamics in Australia (HILDA) survey 2001.

Objectives

The objective of this paper is to highlight the importance of household panel survey data as a means for informing evidence for policy decision-making and to discuss briefly the advantages and disadvantages in using such data. Drawing from the HILDA survey of Australia this paper illustrates how the panel data have contributed to evidence-based policy. It also draws attention to the fact that while household panel data have the potential to provide policy-makers with valuable evidence on the socio-economic conditions and well-being of the population, their use will depend largely on the quality of the information, sample size, representativeness of the sample at different points in time, and the level of sophistication of the methodology used for the analysis of panel data.

Panel surveys

Panel surveys collect information over time on the behaviour of individuals, families and households. This means that with panel data it will be possible to undertake statistical analysis of the change and dynamic behaviour of the study population. While cross-sectional surveys can also collect data retrospectively to analyse change, the data are often subject to measurement errors.

As the panel surveys are conducted on a regular basis, they have both cross-sectional and time-series elements; each wave is similar to a cross-sectional survey and when data for more than one wave have been collected then it will become a time series. Thus, panel surveys allow cross-sectional analysis of a particular issue of policy relevance as well as providing time-series analysis to assess trends, at the individual level, as opposed to aggregate-level analysis that is possible with cross-sectional data. More importantly, panel surveys provide opportunity to examine transitions between states–flow data to analyse change.

Like panel surveys, repeated cross-sectional surveys can collect information on the target population at different points in time, but without the assurance that the subsequent surveys will include the same population covered in the previous rounds. The advantage of panel surveys, therefore, is that they cover the same persons at different points in time, including split families or households, and add new members ‘born into’ the sample when they become in-scope according to the criteria used for sample management. Depending on the method used for updating the sampled population, panel data are suitable for making point-in-time estimates as would be possible from a fresh cross-section of the population (Duncan, Juster, and Morgan, 1987).

Potential uses of panel data

Measure gross change and identify transitions

As panel surveys are designed to assess the behaviour of individuals, families and households over time they are useful for the statistical analysis of change and dynamic behaviour of the population. For example, using panel survey data one could examine the impact of a social policy on a population and identify population subgroups affected by the policy much more precisely than can be done using cross-sectional data, even if derived from repeated cross-sectional surveys. Cross-sectional surveys can be used to measure the net change of a socio-economic indicator across different points in time. Because panel surveys collect data on the same respondent over time they can be used to measure gross change at the micro (individual) level and identify transitions between states over time. As panel surveys typically collect data on core questions in each wave and on a large number of background and behavioural characteristics, it will be possible to find dynamic links of a specific issue, event (or an incidence) with a host of other factors.

In addition to the measuring of gross change, panel data are useful to measure change at other levels (Duncan and Kalton, 1987): average change for each individual; identifying average change or trends using multiple time points is useful to assess true trends discounting the effects of isolated factors affecting the observed change. In addition, the assessment of the instability for each individual provides a valuable input for modelling for further analysis, for example, the effect of an individual's income instability on consumption decisions.

Cross-sectional surveys can collect some types of information retrospectively, but such data are subject to errors, mainly recall errors due to telescoping: misdating of events that took place in the past (Mathiowetz and Duncan, 1984; Rose, 1993). Such data errors increase with the length of the reference period and these errors may occur even in panel surveys, which use longer interview periods (Givens and Massey, 1984).

Provide in depth information on policy issues

Panel data can provide valuable information for policy-makers on an issue of policy significance that cross-sectional data cannot provide. At times panel data provide information that is quite different to that derived from cross-sectional analysis. For example, using available evidence Burkhauser and Smeeding (2001) demonstrate, among other things, how cross-sectional and longitudinal panel data gave differing information to policy-makers on changes in the well-being of older people in the US. Using the results of the Panel Study of Income Dynamics (PSID), Burkhauser, Cut and Lillard (1999) demonstrated that contrary to the cross-sectional findings of the panel data which showed an improvement in the mean household size-adjusted income of older people (65 and over) from 1983 to 1989, the longitudinal analysis of data showed that the increased income was evident only among the new cohorts of older persons: the income of people already 65 years or over in both years had fallen between 1983 and 1989. Such an analysis brings to the policy-makers evidence that is quite different from that revealed in cross-sectional analysis. This is undoubtedly a good example of the role played by panel data in evidence-based policy.

Panel surveys can provide strong evidence of the effects of a policy change when the change occurs during the course of a panel survey, because it can measure an individual's circumstances before and after the policy change (Buck, Ermisch, and Jenkins, 1995). The strength of the evidence depends largely on the nature of the policy and any time lag between the changes in policy and the impact. However, if the interest is to examine the impact of policy on population groups then repeated cross-sectional data can also be useful, particularly when such surveys use larger samples (Buck, Ermisch, and Jenkins, 1995).

Determining the causality

It is well known that cross-sectional data are cannot resolve the issue of ambiguity in correlation and, more importantly, cannot confidently demonstrate the direction of causality (Davies, 1994:28). For example, it will be difficult to identify the direction of association between unemployment and poor health: which is caused by which? As panel surveys interview the same individual over different points in time, and panel data have a time order of measurement, they are suitable for assessing causality between variables. Panel data have the capacity to identify stability and change at the individual level. The causal association can be determined when the cause precedes the effect (or outcome of interest). Panel data have the advantage of convincingly identifying this.

Control for unobserved characteristics

The panel data contain repeated measures of the same variable for the same individual over time and because of this, panel data have the potential to control for the effects of some unobservable individual characteristics that do not vary over time (Hsiao, 2003), and information on which has not been collected. These unobserved characteristics may be influencing the criterion variable via some other explanatory variables.

Often the problems in making causal inferences are largely caused by unobservables and this becomes an important contribution of panel data in solving the problem of causal inferences (Halaby, 2004). These unobserved determinants include those that vary across individuals but remain the same across time for any individual (*unit-specific effects)* and those that could change across time (*disturbances*).

These determinants could affect explanatory variables and the presence of these is termed *population heterogeneity* (Rose, 1993). As individual information is collected at different points in time, panel surveys allow efficient modelling[2] of the impact of unobserved factors (Hsiao, 2003).

Although panel surveys offer the opportunity to measure unobservable fixed effects, any measurement error associated with the data could affect the quality of the estimates produced.

Having an increased sample by pooling data from individual waves

Panel surveys provide a valuable opportunity to increase the sample size by pooling data collected through different waves, without resorting to any oversampling, which is common in cross-sectional surveys, to achieve adequate representation of minorities or rare populations. Such cumulation is possible with some items such as new events over time for example, divorces or persons with a given disease. However, it cannot cumulate events of completely static characteristics such as persons with foreign–born parents (Duncan and Kalton, 1987:103).

Make cross-country comparisons

As Burkhauser and Smeeding (2001) point out, if panel data are collected uniformly in other nations with similar social and economic development, then such data can be used to observe the ways in which countries react to similar challenges with alternative polices and outcomes. Such exercises are valuable for social policy, because most advanced economies face similar issues of social concern (Rose, 2000). Now panel surveys are conducted in both developed and less developed countries and in some economic regions. The European Community Household Panel collects unified data across a number of countries to undertake comparative analysis across countries. Over time, comparative analysis that can provide evidence for decision-making will become available from these panel surveys. On the basis of these developments and the potential for panel data to inform policy, some have proposed that international comparability should be a criterion to judge the appropriateness of specific question items or panel data supplements (Burkhauser and Smeeding, 2001).

Estimate age, period and cohort effects

Household panel surveys are useful to assess age effects (that vary across age but are similar for all respondents of a particular age), period effects (that vary across time but are similar for any particular point in time) and cohort-specific effects (that are the same for individuals in the cohort but otherwise different across respondents). To decompose these effects using cross-sectional data is difficult. Panel data provide valuable information to assess such affects.

Measure intergenerational effects

Panel data contain measures of second-generation outcomes such as earnings, welfare dependence, and social behaviour as well as first-generation circumstances as reported by parents. Therefore, panel data could be useful in measuring the strength of the association between parental status and children's adult status or their pathways of transmission (Rose, 2000, Jenkings and Siedler, 2007), and social behaviour (Thornberry, Freeman-Gallant and Lovegrove, 2009). Panels with relatively long duration or panels with some retrospective information could be used for the study of intergenerational mobility of different statuses or examining the parental behavioural characteristics and their relationships with children.

The HILDA survey

The HILDA survey is a national household panel survey launched in 2001 aiming to provide nationally representative information on various aspects of socio-economic life of the Australian population with specific focus on households and families, income, employment, and well-being. The survey is initiated and funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs and is managed by the Melbourne Institute of Applied Economic and Policy Research at the University of Melbourne (Melbourne Institute). The HILDA is a household panel chosen by two stages: in the first stage Collection Districts (CDs) were selected by probability proportion to size (PPS) method and the sample is stratified by geographic areas (States and Territories and geographic regions within those jurisdictions). Owing to the expected high cost of survey operations, CDs in remote and sparsely populated areas have been excluded from the survey. Within each of the CDs chosen a random sample of 22 to 34 dwellings was selected (Watson and Wooden, 2002).

The final sample selected included 13,969 respondents aged 15 years and over chosen from 7,682 households living in private dwellings. The first wave of the data collection began in 2001 with subsequent waves conducted on an annual basis. Between 2001 and 2010 the survey completed 9 waves and Wave 10 is currently being conducted. Published data are available for the first six waves. Although there has been a progressive drop in subsequent waves from the number of respondents included in Wave 1, the numbers retained in the sample are sufficient to assess changes in socio-economic and behavioural characteristics of the population at the individual level.

The present analysis uses, the HILDA data from Wave 1 to Wave 6 published by the Melbourne Institute (Wilkins, Warren and Hahn, 2009). Drawing on the HILDA survey, the following sections illustrate how household panel data can be used to identify dynamics of the population that cannot be adequately covered in a cross-sectional setting. The analysis that follows looks at only a few aspects of the socio-economic status and well-being of the people, and is limited to examining the dynamics of change in marital status, household structure, labour force status and income. The analysis draws comparisons with the relevant estimates derived for the HILDA survey using cross-sectional perspectives. Unless otherwise stated, the information from the HILDA survey refers to those cross-tabulations provided in the report issued by the Melbourne Institute (Wilkins, Warren and Hahn, 2009).

Evidence from the HILDA survey

Marital status distribution levels and trends

The HILDA data used as cross-sections indicates that over the six-year period the marriage pattern in Australia has remained relatively stable with around 52 per cent legally married, 26 per cent never-married (and also not in *de facto* unions), 5 per cent each divorced or widowed (Table 1). *De facto* relationships have increased although modestly, and the percentage of persons separated has fluctuated over a narrow range from 2.3 to 3.0 per cent.

Table 1: Marital status distribution of population for individual years from 2001 to 2006: HILDA survey Waves 1 to 6 (%)

Marital status	2001	2002	2003	2004	2005	2006
Legally married	52.2	51.8	51.6	51.1	51.7	51.3
De facto	8.6	8.9	9.3	9.4	9.4	9.8
Separated	3.0	3.0	2.6	2.8	2.5	2.3
Divorced	5.2	5.5	5.6	5.5	5.5	5.6
Widowed	5.0	5.1	4.9	5.1	5.0	5.0
Never-married and not *de facto*	26.0	25.7	26.0	26.1	25.8	26.0
All	100.0	100.0	100.0	100.0	100.0	100.0

Source: Wilkins, Warren and Hahn, 2009; Table 2.2
Note: Weighted figures (cross-sectional population weights)

Examining the marital status of the population longitudinally reveals interesting patterns that have substantial policy interest. Of the persons legally married in 2001 only 92.0 per cent remained so in 2006. About 3.8 per cent of those married in 2001 later either separated (2.3 per cent) or divorced

(1.5 per cent), while 3.2 per cent have been widowed since 2001 (Table 2). On the other hand, by 2006, about a third of those in *de facto* relationships in 2001 had married and about half remained in a *de facto* relationship, which may or may not be the same relationship as in 2001. Approximately 2.3 per cent of persons in a *de facto* relationship in 2001 were separated, and 4.4 per cent had been divorced, from a relationship (legally married) that occurred after Wave 1 interviews. Some (11 per cent) moved out of *de facto* relationships and are recorded as not married in 2006.

Similarly, the longitudinal data reveal that approximately 10 per cent of the persons recorded as separated in 2001 had married by 2006 and a further 12 per cent were in *de facto* relationships, while another 30 per cent were divorced. Of those widowed in 2001, 98 per cent were still in the same category after five years while approximately 15 per cent were in a *de facto* relationship. Of persons identified as never-married and not in *de facto* relationships in 2001, 73 per cent remained as never-married in 2006. These dynamics in marital relationships and the pattern of relationship breakdowns can be judged by examining panel data longitudinally, and provide valuable information for policy-makers, which cannot be ascertained from cross-sectional analysis.

Table 2: Dynamics of marital status distributions 2001 and 2006: the HILDA survey Waves 1 and 6 (%)

	Marital status in 2006					
Status in 2001[a]	Married[b]	*De facto*	Separated	Divorced	Widowed	Not married[c]
Married[b]	92.0	0.9	2.3	1.5	3.2	-
De facto	32.4	49.5	2.3	4.4[d]	0.6	10.8
Separated	9.7	11.6	43.8	30.0	4.9	-
Divorced	8.6	8	0.8	76.8	5.8	-
Widowed	0.5	0.1	0	1.7	97.6	-
Not married[c]	11.9	14.5	0.7	0.3	0.1	72..5

Source: Wilkins, Warren and Hahn, 2009; Table 2.2

Note: Weighted figures (longitudinal population weights)

a Marital status in 2001; b Legally married; c Never-married and not in *de facto* relationship

d Implies that persons in *de facto* relationship in 2001 had married later and divorced by 2006.

Household formation and dissolution

Interest in household formation and dissolution processes has increased over the past few decades in most countries in the Western world. Consequently, bureaucrats, policy-makers and researchers alike are trying to gain insight into the formation and dissolution process, including the growth of one-parent families, non-family households and the traditional family.

Table 3: Household type of individuals, 2001–2006 (%)

Household type	2001	2002	2003	2004	2005	2006
Couple family	21.3	21.5	21.4	21.3	22.0	22.0
Couple family without children	21.3	21.5	21.4	21.3	22.0	22.0
Couple family without children	51.2	50.9	51.9	51.1	51.2	5-0.7
Lone parent households	11.6	12.0	12.1	12.4	12.4	12.0
Lone person	9.8	10.0	10.1	10.2	10.1	10.3
Group households	2.6	1.8	1.5	1.4	1.2	1.2
Other household types	3.5	3.8	3.0	3.6	3.1	3.8
All	100.0	100.0	100.0	100.0	100.0	100.0

Source: Wilkins, Warren and Hahn, 2009: Table 1.1

Note Population weighted data (Cross-sectional weights).

Identification of these trends is of major importance to national and state and territory policy-makers because of their perceived impact on social support payments, for example, female-headed one-parent families, child allowances, care for the aged, housing (1 and 2 person households), and income and consumption.

The cross-sectional analysis of the HILDA survey indicates that household composition in Australia has not changed significantly over the past six years. For example, couple-families with children remained stable at 51 per cent, lone-parent households at 12 per cent and lone-person households at 10 per cent. Although the net change in the proportions of household structure at the *aggregate level* has been more or less stable, there have been significant changes in household structure as shown in the longitudinal analysis of the HILDA data.

Table 4: Dynamics of household structure change: 2001–2006: the HILDA survey, Waves 1 and 6 (%)

Households by type in 2001	Households by type in 2006					
	Couple family without children	Couple family with children	Lone-parent	Lone-person	Group	All other
Couple family without children	73.9	16.0	0.9	8.5	0.4	0.4
Couple family with children	15.5	79.8	4.8	6.7	0.7	2.6
Lone parent	6.7	12.3	59.1	18.9	1.2	1.8
Lone person	10.3	5.3	2.5	80.1	1.1	0.6
Group	28.3	18.1	4.7	34.2	13.3	1.4
All other	31.2	14.0	11.9	12.0	2.6	28.2

Source: adapted from Wilkins, Warren and Hahn, (2009) Table 1.3

Note: Population weighted results (longitudinal weights) and percentages have been estimated using raw totals. Percentages may not add up to 100 due to rounding.

The household structure in Australia has shown major transitions in different directions, which will help to explain gross change in household structure in the population at the *individual level*. For example, 74 per cent of the couple-families without children in 2001 remained the same in 2006 (Table 4). About 16 per cent had children after the first wave and thus changed to couple-family households with children. Some nine per cent of the couple-families without children have become, within six years, lone-parent households. A further nine per cent have become lone-person households.

About 70 per cent of the couple-families with children in 2001 remained the same in 2006, whereas 16 per cent of them by 2006 did not have children residing with them. About 59 per cent of the total lone-parent households did not change between 2001 and 2006, however, about 19 per cent had moved to become couple families either with or without children.

Identification of these transitions in household structure is extremely useful for public policy, and transition rates derived from panel surveys are valuable for projecting households. The cross-sectional analysis on the other hand masks the presence of these transitions and provides limited and less precise information on the changes in household structure that took place during 2001–2006.

Changes in labour force status of the population

Cross-sectional data can give details such as how many or what proportion of working-age persons are employed or how many are unemployed at a particular point in time, whether these proportions have changed over time, and if so, in what direction. The labour force status of persons recorded from Wave 1 of the survey can be compared with subsequent waves and the results are presented in Table 5.

Table 5: Labour force status of population from 2001 to 2006, persons 15 years of age and over (both sexes combined): Waves 1 to 6 of the HILDA survey (%)

Labour force status	2001	2002	2003	2004	2005	2006
Employed	60.5	61.0	61.2	62.0	62.6	63.1
Full-time	42.0	41.6	41.9	42.4	42.9	43.4
Part–time	18.5	19.3	19.4	19.6	19.7	19.7
Unemployed	4.4	3.9	3.6	3.3	3.2	3.1
Not in labour force	35.2	35.1	35.2	34.7	34.2	33.8
All	100.0	100.0	100.0	100.0	100.0	100.0

Source: Derived from unpublished data from the Melbourne Institute.
Note: Population weighted results (cross-sectional weights)

The cross-sectional data reveal that the rate of employment (percentage of persons 15 years and over) increased gradually over the six-year period from

60.5 per cent in 2001 to 63.1 per cent in 2006. This increase was common to both full-time and part-time employment. During the same period, the unemployment proportion decreased, although slightly. The percentage of persons not in the labour force also declined, particularly after 2003.

Looking at the longitudinal data, of the persons having full-time jobs in 2001 only 88.4 per cent had full-time jobs in 2002 (Table 6). About 6.4 per cent had moved to part-time jobs and some (1.7 per cent) were unemployed; nearly 4 per cent were neither employed nor looking for work: not in the labour force. Over the years, the proportion of people holding full time jobs has gradually declined and those moving to part-time jobs has progressively increased.

Table 6: Labour force status mobility of persons (both sexes combined) aged 15 years and over: comparison of employment and unemployment statuses from 2001 to 2006–Waves 1 to 6 of the HILDA Survey (%)

Labour force category	Labour force status in				
	2002	2003	2004	2005	2006
Employed full-time in 2001					
Employed full time	88.4	84.9	83.0	80.6	78.7
Employed part time	6.4	8.7	8.9	10.4	11.2
Unemployed	1.7	1.2	1.1	1.2	1.3
Not in labour force	3.5	5.2	6.9	7.8	8.8
Employed part-time in 2001					
Employed full-time	18.3	24.7	29.4	34.5	37.0
Employed part-time	66.8	58.3	52.5	48.1	44.2
Unemployed	3.0	2.4	3.0	2.2	2.1
Not in labour force	11.9	14.7	15.1	15.2	16.7
Unemployed in 2001					
Employed full-time	23.2	31.7	32.3	37.5	38.0
Employed part-time	21.6	24.3	27.4	24.1	23.3
Unemployed	30.0	20.1	14.0	16.4	11.5
Not in labour force	25.2	23.9	26.3	22.0	27.3

Source: Wilkins, Warren and Hahn (2009) adapted from table 11.3.
Note: Population weighted results (longitudinal weights).

There has been an increase in the proportion of persons who were employed part time in 2001 and moved into full-time jobs in the subsequent years: the proportion who went to full-time jobs in 2002 was 18.3 per cent and by 2006, 37 per cent of them were in full-time jobs. The proportion of persons remaining in part-time jobs showed a corresponding decline. Like those holding full-time jobs in 2001, part-timers also became unemployed, and an increasing proportion of people were not employed and not looking for work in subsequent years. Without panel data evidence, job losses experienced during the subsequent five years by part and full-time workers employed in 2001 would be unknown to decision-makers.

Poverty dynamics

From cross-sectional data one can estimate indicators such as the percentage of persons below the relative poverty line, and income distribution across different socio-economic groups. Using measures of income distribution through indicators such as the Gini-coefficient, one can also assess inequality of income across time if repeat cross-section data are available for different population groups. Such data are useful for policy decision-making as they provide broad trends and socio-economic differentials in income distribution.

However, cross-sectional data do not provide information to distinguish individuals, families or households who are temporarily poor (or low-income) from those who are at risk of being poor over a long period (chronic poor). Similarly, if policy-makers need evidence of the full impact of social policies on levels and distributions of income it should be obtained by observing policy outcomes and individual behaviour responding to policies across multiple periods of time (Burkhauser and Smeeding, 2001). To gain a better understanding of poverty dynamics, cross-sectional data can be complemented by longitudinal analysis (Devicienti, 2001).

Table 7: Trends in poverty line and poverty rates (percentage of the population below the relative poverty line): HILDA survey Waves 1 to 6 (cross-sectional measurements)

Year	Poverty line ($)[a]	Relative poverty (%)	Absolute poverty (%)
2001	12,259	13.4	13.4
2002	12,727	12.9	12.5
2003	13,443	12.8	11.3
2004	14,138	12.5	9.8
2005	14,998	13.6	9.5
2006	16,120	11.9	6.7

Source: Wilkins, Warren and Hahn, 2009: Table 7.1

Note Population weighted data (Cross-sectional weights).

a The poverty line is set at household income half the median household income of the relevant year. Relative poverty is based on this measure and indicates the number of people below the poverty line. Income used in the computation of percentages is household size-adjusted income.

The HILDA data on income for the six years from 2001 to 2006 derived as a cross-section of the population demonstrate that the proportion of the population below the poverty line has declined over the period. The only exception was the relative poverty recorded for 2005, which showed an increase from the previous year. These increases occurred despite the progressive rise in the poverty line over the same period. For example, the poverty line at current prices in 2002 was 1.04 times the 2001 level and the level for 2006 was more than 1.3 times the 2001 level (Table 7). The annual poverty rate in Australia was around 13 per cent during the six-year period.

The longitudinal element of the HILDA survey can be used to identify the number of people who have been poor (with low income) throughout the six-year period and those who experienced low incomes for only a portion of the six years. Wilkins, Warren and Hahn, (2009) define poverty or low income as less than half the equalised median income of households. The question that interests to policy-makers is whether the observed poverty status is due to population homogeneity or addictive state dependence.

Table 8: Percentage of people according to the number of years of poverty (low income), 2001–2006: HILDA survey Waves 1 to 6

Number of years in poverty/low income	Per cent of people
Not in poverty in any year during 2001-06	69.1
Any one year	12.9
Any two years	5.7
Any three years	4.5
Any four years	2.8
Any five years	2.6
All six years	2.6
All	100.0

Source: Wilkins, Warren and Hahn, (2009): Table 7.2 (some figures have been rounded to maintain consistency).
Note: Population weighted data (longitudinal weights)

Overall 31 per cent of the population was in poverty (or with low income) in at least one year during the six-year period. By contrast, only 2.6 per cent of the population were identified as being poor consistently throughout the six-year period (Table 8). This proportion may be regarded as the chronically poor segment of the population. This kind of information is important for policy, and only longitudinal analysis can provide such information as the same individual is followed over time.

Table 9: Poverty persistence: HILDA survey, Waves 1 to 6

	Per cent of persons in poverty also in		
In poverty in the year	2001	2003	2005
2002	53.3	na	na
2003	49.7	na	na
2004	44.1	58.1	na
2005	46.7	55.1	na
2006	41.6	49.3	56.9

Source: Wilkins, Warren and Hahn, 2009: Table 7.4.
Note: Population weighted data (longitudinal weights); na = not applicable.

Of the persons identified as in poverty in 2001, 53 per cent were also in poverty in the following year and approximately half were also in poverty in 2003 (Table 9).The corresponding proportions for 2005 and 2006 were 47 per cent and 42 per cent respectively. Similarly 58 per cent of those identified as in poverty in 2003 were still in poverty in 2004. Approximately 55 per cent of those in poverty in 2003 were in poverty in 2005 and 49 per cent in 2006. More than half the persons identified in poverty in 2005 were in poverty in 2006. Such information is crucial for policy-makers as the persons who are likely to be in poverty for a medium to long term (chronic poverty) require policy options different from those in poverty for a short period (transition).

Controlling for unobservables

There are many research reports available which apply econometric models to the HILDA data controlling for unobserved variables. There are a few research studies such as the one carried out by Leigh (2007), that put a particular emphasis on estimating the impact of unobserved variables. Leigh's study used the HILDA survey Waves 1 to 5 to assess how employment outcomes change consequent to the change in informal care arrangements for the elderly. The research showed among other things that the cross-sectional data provided a misleading picture of the causal impact of informal care on labour force status outcomes. This was a result of the impact of omitted variables. Leigh fitted two regression models – one with individual fixed effects and the other without individual fixed effects – to assess whether the cross-sectional relationship between care-giving and labour supply is due to individual heterogeneity.

Problems with panel data

While panel data have numerous advantages over cross-sectional data they too are subject to limitations. The major limitation is the cost. Depending on the methods used to choose the samples, the cost of the first wave of a panel survey is expected to be not much different from that of the cross-sectional survey. Conducting subsequent waves in a Panel survey is expensive because it involves tracking all the original sample members who are in-scope for subsequent surveys. With the increased mobility of respondents, managers of longitudinal surveys spend additional resources to track respondents and get their co-operation to participate in the subsequent waves of the survey. Panel surveys, therefore, use incentives, both cash payments and gifts, to improve the response rates of the first survey and to maximize participation in subsequent survey waves. While their effectiveness in enhancing response rates is dependent on how they are designed and administered, these incentives will increase the cost of panel surveys.

Non-response and attrition

Non-response is common to both types of survey but its impact is greater in household panel surveys. Attrition occurs when sample subjects do not participate in successive waves. As the persons not contacted on the first wave of the panel are unlikely to be contacted in subsequent waves and others drop out in subsequent waves either by inability to be traced or refusal to participate, the attrition bias will be cumulative. Although some respondents could miss some waves and participate in others, the attrition in panel surveys is in general non-random and could affect the representativeness of the sample as time goes on.

Panel conditioning

This refers to the possibility that the members in the panel may change their attitudes or behaviour and respond to later waves of panel surveys accordingly because of their participation in previous waves. This means that the responses provided in later waves may be conditioned by the previous experience of taking part in the survey. The panel conditioning (also known as time-in-sample bias) can occur because of changes in the actual behaviour of panel members. For example, if the previous wave of interviews made some respondents aware of an activity, which was not familiar to them in the past, this awareness could lead some respondents to engage in those activities subsequently and to report on them in subsequent waves.

It is also possible that the experience in the previous survey of being exposed to certain survey questions can lead to giving different responses in later interviews, either through avoiding responding to a series of filtered questions or just through been interviewed in the survey previously. In both cases reported events, behaviour or characteristics affect only the panel members and not the general population (Lynn, 2009). On the other hand it is also possible to give accurate responses in later waves if the respondent has established a sense of trust with the interviewer, if the same interviewer conducts the interviews.

Seam effect

Panel surveys in general include questions designed to collect data for subperiods within the reference period of the last wave of data collection. This form of data collection has revealed that the levels of reported changes in behaviour (e.g. going on and off welfare) are likely to be greater at the seam between two adjacent waves than within two adjacent subperiods. The seam effect can be seen when data from two or more waves are combined. To minimise the impact of seam effects some panel surveys use the technique of dependent interviewing by feeding back to respondents their responses to some key questions at earlier waves of data collection.

Representativeness issues

Panel surveys are considered not as good as cross-sectional surveys at giving cross-sectional estimates if the subsequent waves of panel surveys are not representative of the population or subject to a high level of coverage errors, which are likely to accumulate over time (Lynn, 2009). Additionally, the response rates of subsequent waves could also be lower than those observed in cross-sectional surveys, partly because of dropouts from later waves and difficulty in tracing sample members.

Conclusions

This paper, citing some examples from the HILDA survey, has highlighted how panel surveys can provide valuable information that cross-sectional studies cannot provide on the socio-economic life of the population. A few examples provided in the paper reveal that panel data can be valuable in analysing the dynamics of a population: the most important contribution of panel surveys to social policy. While cross-sectional data are useful to gain an understanding of the levels and directions of change in a socio-economic phenomenon at the macro level, to understand the dynamic nature of gross change panel data are necessary.

For example, to develop and monitor policy outcomes to address poverty it is necessary to distinguish the poor according to whether they are experiencing poverty as transitory or chronic. Such comparisons can only be done with panel data. Cross-sectional comparisons of a particular state using data collected from different individuals at different points in time fail to answer these questions such as how many unemployed persons remain unemployed or the pattern of continuous welfare dependence among some segments of the population. To examine such changes, tracking of the same individual, and family or household is necessary.

Although the measurement errors associated with panel surveys should not be over-emphasised in view of the importance of such surveys in contributing to evidence-based policy-making, it is equally important not to over–emphasise the value of panel surveys as the sole source for evidence-based policy. As Hsiao (2005) stated, panel data offer numerous advantages but they should not be considered a panacea. To be useful, panel data analysis should be based on understanding their advantages for the issue of interest, the limitations of panel data and the methodology chosen for analysis, assumptions underlying the statistical inference procedures and the ways of increasing the efficiency of estimation parameters (Hsiao, 2005).

Even the identification of relationships between explanatory variables and socio-economic status, for which traditional cross-sectional studies are suitable, can be measured with accuracy if panel data are available.

While there are potential issues pertaining to panel surveys and panel data, the anticipated high cost of conducting such surveys and data management, there is a strong case for conducting panel surveys to provide valuable information on the socio-economic well-being of the population to inform public policy. What is needed is to understand such issues as any limitations in the sophisticated statistical methods applied to panel data and to inform the policy-makers accordingly, so that they will not have undue expectations.

The real advantage of panel data is that they can be used for detailed analysis controlling for many variables including unobservables. The quality and the usefulness of such analysis depend on the quality of the data collected and the appropriateness of the methods used. To improve the value of such analysis researchers should provide the policy-makers with information not only on the results but also on the likely quality issues that may be relevant in using such analysis for policy decision-making.

The value of the panel surveys can be improved if measures are taken to minimise the disadvantages and maximise advantages. According to Duncan, (1992: 4) these methods include the selection of the initial sample to the highest quality, clear and effective implementation of 'following rules', effective control over panel attrition, use of feedback methods in interviewing to minimise wave inconsistencies, and collecting continuous information about change between each wave.

References cited

Buck, N., Ermisch, J.F., and Jenkins, S.P. (1995) *Choosing a longitudinal survey design: the issues*, ESRC Research Centre on micro-social change, University of Essex.

Burkhauser, R.V., Cut, A.C., and Lillard, D. (1999) 'How older people in the United States and Germany fared in the growth years of the 1980s: a cross-sectional versus a longitudinal view', *Journal of Gerontology*: Social science 54B (5): S279–290 (cited by Burkhauser and Smeeding 2001).

Burkhauser, R.V. and Smeeding, T.M. (2001) 'The role of micro-level panel data in policy research' , *Journal of Applied Social Science Studies*, 121 (4): 469–500.

Citro, C.F. and Hansheck, E.A. (ed) (1991) *Improving information for social policy decisions – the uses of microsimulation modelling Vol. 1: Review and recommendations,* Panel to evaluate microsimulation models for social welfare programs, Committee on National Statistics, National Research Council, National Academy Press, Washington DC.

Davies, P. (2004) "*Is evidence-based government possible*?' Jerry Lee Lecture presented at the 4th Annual Campbell Collaboration Colloquium, 19 February, Washington DC.

Davies, R.B. (1994) 'From cross-sectional to longitudinal analysis' In Dale, A.S. and Davies, R. B. (eds) *Analysing social and political change*, Thousand Oaks, CA: Sage publications, 20–40.

Devicienti, F. (2001) 'Estimating poverty persistence in Britain', *Working Paper Series* No.1, Laboratorio R. Revelle Centre for Employment Studies

Duncan, G.J. (1992) 'Household panel studies: prospects and problems', *Working Papers of the European Scientific Network on Household Panel Studies*, Paper No. 54, Colchester: University of Essex (cited by Buck, N., Ermisch, J.F., and Jenkins, S.P. (1995).

Duncan, G.J., Juster, F.T., and Morgan, J.N. (1987) 'The role of panel studies in research on economic behaviour ', *Transportation Research*, 21A (4/5):249–263.

Duncan, G.J. and Kalton, G. (1987) 'Issues of design and analysis of surveys across time', *International Statistical Review*, 55 (1): 97–117.

Givens J.D., and Massey, J.T. (1984) 'Comparison of statistics from a cross-sectional and a panel survey. *Proceedings of the Research Methods Section of the American Statistical Associatio*n; 640–645.

Halaby, C.N. (2004) 'Panel models in sociological research: theory into practice', *Annual Sociological Review* 30: pp507–544.

Henry, G.T. (2009) 'When getting it right matters: the case for high–quality policy and program impact evaluation'. In S.I. Donaldson, C.A. Christie, and M.M. Mark (eds) *What counts as credible evidence in applied research and evaluation practice?* Newbury Park, CA: Sage, 32–50.

Hsiao, C. (2003) *Analysis of panel data* Cambridge: Cambridge University Press.

------------- (2005) 'Why panel data?' *IEPR Working Paper* 05-33, Institute of Economic Policy Research, University of Southern Carolina.

Jenkins, S.P., and Siedler, T. (2007) 'Using household panel data to understand the intergenerational transmission of poverty', *CPRC Working Paper* No.74, Chronic Poverty Research Centre, Institute of Social and Economic Research, University of Essex.

Leigh, A. (2007) 'Informal care and labour market outcomes' Report prepared for the Department of Families, Housing, Community Services and Indigenous Affairs (mimeo).

Lynn, P. (2009) Methods for longitudinal surveys. In P. Lynn (ed), *Methodology of longitudinal survey.* Chichester:John Wiley and Sons Ltd:1–19.

Mathiowetz, N. and Duncan, G. (1984) 'Temporal patterns of responsible errors in retrospective reports of unemployment and occupation', *Proceedings of the Research Methods Section of the American Statistical Associatio*n: 652–657.

Rose, D. (1993) 'European household panel studies', *Working Paper No 45 of the ESF network on Household Panel Studies*. Colchester: University of Essex.

——————— (2000) 'Household panel studies: an overview'. In D. Rose (ed) *Researching social and economic change, The uses of household panel studies*. London: Routledge: 3-35.

Rossini, E. (2002) *Introduction to longitudinal research*. London: Routledge.

Thornberry, T.P., Freeman-Gallant, A., and Lovegrove. P.J. (2009) 'The impact of parental stressors on the intergenerational transmission of antisocial behavior' *Journal of Youth and Adolescence*, 38: 312–322.

Watson, N. and Wooden, M. (2002) 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 Survey Methodology' *HILDA Project Technical Paper Series No. 1/02*, Melbourne Institute, University of Melbourne.

Wilkins, R., Warren, D., and Hahn, M. (2009)*Families, incomes and jobs, Volume 4: A statistical report on Waves 1 to 6 of the HILDA survey*, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

---

[1] Davies cites a model outlining other factors that influence policy-making in government, which includes experience and expertise, judgment, resources, values, habits and tradition, lobbyists and pressure groups, pragmatics and contingencies.

[2] To put the above arguments more clearly, let us consider the following simple model:

$Y_{it} = X_{it}\beta + Z_i\gamma + \alpha_i + \eta_{it}$ I = 1,2,..N (individuals) and t = 1,2,…T (time).

Where β and γ represent k and m vectors of coefficients associated with time-varying and time-invariant observable explanatory variables respectively, te disturbance η is assumed to be uncorrelated with (X, Z, α) where α is the unobservable time-invariant random variable and is distributed independently across individuals. Here our interest is on the potential correlation of α with X and Z variables. In the presence of such correlations, whether t = 1 (cross section) or t= 1,2,..T (panel), ordinary least squares (OLS) and generalized least squares (GLS) yield biased and inconsistent estimates of the parameters β and γ. The conventional technique to overcome this problem is to eliminate the individual effects in the sample by transforming the data into deviations from individual means in the above equation. Thus, the 'within' estimators of the parameters in the above equation can be obtained by regressing the 'within' group deviations of the explained variables on those of the explanatory variables. $Y_{it} - \overline{Y}_i = \left(X_{it} - \overline{X}_i\right)\beta + \left(Z_{it} - \overline{Z}_i\right)\gamma + \left(\eta_{it} - \overline{\eta}_i\right)$

$$\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$$

and

$$\overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_i$$

Alternatively, the dummy variable least squares approach, where individual specific dummies are included in the above equation, can be used to estimate the parameters of interest.

In either case, the fixed effects technique is a 'within-individual' regression. It uses the variability of the data within each individual through time and not the variability of the data across individuals at any given point in time. Thus one of the shortcomings of the method is that it fails to exploit fully the rich information panel data can provide. Nevertheless, panel data provide a method to control for the influence of unobservable individual time-invariant effects on explained variables which cross-sections cannot do.