The Household, Income and Labour Dynamics in Australia (HILDA) Survey: An Introduction

Nicole Watson and Mark Wooden

Melbourne Institute of Applied Economic and Social Research,

University of Melbourne

Paper presented at the 30th Annual Conference of Economists,

University of Western Australia,

23-26 September 2001

Acknowledgement

The HILDA Project was funded, and is initiated, by the Commonwealth Department of Family and Community Services (FaCS). The views expressed in this article, however, are those of the authors and should not be attributed to either FaCS or the Melbourne Institute.

1. Introduction

Most industrial nations now conduct large-scale, representative panel (i.e., longitudinal) surveys of households and the members of those households. Australia, however, is a notable exception. While longitudinal data collections do exist in Australia, they typically focus on relatively small sub-groups of the population. Included here, for example, are the Longitudinal Surveys of Australian Youth (LSAY), which follow samples of young people,[1] the Longitudinal Survey of Immigrants to Australia (LSIA), which is restricted to recent immigrant arrivals, the Survey of Employment and Unemployment Patterns (SEUP), which had a focus on job seekers, and the Australian Longitudinal Survey of Women's Health (ALSWH), which follows three female age cohorts. Moreover, with the exception of the ALSWH, these surveys were all deliberately designed as relatively short-life panels. For example, the SEUP covered just four years while the first LSIA cohort was followed over a three- to four-year period. Furthermore, the focus in all of these surveys is on individuals rather than households and hence the data collected do not allow individual respondents to be linked to data on other related individuals. Even the Negotiating the Life Course Study, which commenced in 1997 and has a focus on family dynamics, conducts interviews with only one person per household (though information about other household members is collected).[2]

Australian policy-makers and researchers thus do not have access to data that are both representative of the Australian population (or at least a significant proportion of it) and provide information on the dynamic nature of events and how they interact in influencing the changing behaviour and fortunes of Australian households, families and individuals. This has been consistently identified as a major problem in reviews of policy-relevant research literature in areas such as the labour market and social policy (e.g., Creedy 1994; Barr 1999, pp. 22-23; Norris and Wooden 1996, p. 107; Wooden 1997, pp. 262-263).

Assessing the extent of poverty in Australia, for example, has long been plagued by the problem that income varies over the life cycle and hence at any particular point in time, households and individuals may appear income poor, even though this would not appear to be reflected in their consumption behaviour. Conventional cross-sectional data are not well placed to help with this type of research question since they only provide measures of net rather than gross change. Of course, retrospective histories could be collected from members of cross-sectional surveys. This approach, however, is very burdensome on participants and,

more importantly, suffers from problems associated with how respondents recall events in their past.

The obvious solution to this type of dilemma is to collect longitudinal data. Such data would, for example, facilitate a more accurate assessment of incomes from a lifetime perspective. More generally, longitudinal data facilitates the measurement and analysis of gross change. As observed by Rose (2000), Kalton and Citro (2000) and Duncan (2000), the ability to measure gross change is arguably the most important analytical advantage that longitudinal data provide. To return to the poverty example, cross-section data reveal that rates of poverty, as measured by the proportion of persons whose current income is less than 50 per cent of average income, have been relatively stable in Australia over the last two decades.[3] These data, however, concern net change, and almost certainly disguise a high volume of gross change – that is, the movement of people into and out of poverty. Consequently, cross-section data are not well placed to inform us about what sort of people are at most risk of falling into poverty and remaining there, and what factors might be assisting others to escape poverty.

It is against this background that, in 1999, the Commonwealth Government committed funds for the conduct of the first three waves of a major new longitudinal survey. Titled the Household, Income and Labour Dynamics in Australia (or HILDA) Survey, its primary objective is to support research questions falling within three broad and inter-related areas. These are:

(i) income dynamics — with a particular focus on how households respond to policy changes aimed at improving financial incentives, and interactions between changes in family status and poverty;
(ii) labour market dynamics — with a focus on low-to-middle income households, female participation, and work-to-retirement transitions; and
(iii) family dynamics — focusing on separation and divorce and socio-economic status, and on links between income support and family formation and breakdown.

The management of the HILDA Survey was subsequently put out to tender, and eventually culminated in the project being awarded (in August 2000) to the University of Melbourne. More specifically, a team based at the Melbourne Institute of Applied Economic and Social Research is managing the HILDA Survey. The winning tender also involves the significant involvement of the Australian Council for Educational Research (ACER) and the Australian Institute of Family Studies (AIFS).

This paper provides an outline of the HILDA Survey methodology. Its main focus is on the key parameters that underpin the design of the HILDA survey and on the development of the survey instruments. We do conclude, however, with a cursory discussion of some of the issues that could be examined with these data that are likely to be of interest to economists.[4]

2. Survey design

Overview

The panel design being implemented for the HILDA Survey shares many similarities with the designs used in household panel studies conducted overseas, and in particular the German Socio-Economic Panel (GSOEP) and the British Household Panel Survey (BHPS).[5] Thus, like the BHPS, the sampling unit would be the household, and members of those households would be tracked over an indefinite life. Assuming participants can be traced each year and are prepared to cooperate, individuals would only drop out of the sample in the event of death, emigration from Australia, the acquisition of some disability that prevented further participation (such as the onset of dementia), and incarceration.

However, and in contrast to the conventional longitudinal design involving a single cohort, in this design it is possible for the overall sample to grow in size. That is, the sample is automatically extended over time by "following rules" that add to the sample any new children of members of the selected households (including both biological and adopted children) as well as new household members resulting from changes in the composition of the original households.

Compared with other possible designs, such as a classic single cohort panel (e.g., the National Longitudinal Survey of Youth — NLSY — in the USA) or a repeated medium-life panel (e.g., the LSIA and LSAY in Australia, or the Survey of Labour and Income Dynamics — SLID — in Canada), this design is generally thought to be superior in terms of delivering high quality information about family, income and labour dynamics. The lengths of medium-life panels, for example, are often too short to provide an understanding of some of the issues and questions that of highest priority in the HILDA Survey.

An important feature of the proposed design is the hierarchical nature of the data structure, with all individual participants grouped into a larger unit — the household. Household-based data sets thus permit analysis of the behaviour of both individuals and households (and

indeed, other units of analysis are also possible, such as couples, families, and income units within households). Moreover, this data structure is ideal for the development and testing of models in which household characteristics and dynamics are thought to shape individual behaviour, and similarly where the characteristics and behaviour of individual household members are thought to influence outcomes observed for households.

Reference population

The reference population for the first wave of the HILDA Survey is all Australian residents who live in private households. That is, and consistent with most previous longitudinal surveys of this type, the scope of the population for sample selection at the first wave will exclude most persons living in institutions (such as hospitals and other health care institutions, military and police installations, correctional and penal institutions, convents and monasteries, and boarding schools) and other non-private dwellings (such as hotels and motels).[6] Further, to ensure that all members of the in-scope population have the same probability of selection, dwellings that are not primary places of residence (e.g., holiday homes) are also excluded.

For cost reasons, persons who live in remote and sparsely populated areas will be excluded from the sample. The Australian Bureau of Statistics (ABS) adopts the same practice in its supplements to the monthly Labour Force Survey. This exclusion results in about 80,000 persons being omitted from the scope of the HILDA Survey.

Note that while all members of the household are defined as members of the sample, interviews will only be conducted with those persons who are at least 15 years of age (on June 30 in the year the survey wave commences). Some limited information about younger persons (e.g., their date of birth, sex, presence of long-term health conditions or disabilities, and contact arrangements with parents if their parents are separated), however, will be collected from an appropriate adult member of the household.

Data collection unit

The data collection unit is the household. Following the ABS, this is broadly defined as "a group of people who usually reside and eat together" (ABS, *Statistical Concepts Library*, ABS Cat. No. 1361.30.001), with emphasis given to the making of common provision for food.

The ABS makes the further observations about households.

- A household resides wholly within one physical dwelling. A group of people who make common provision for food but are living in two separate dwellings are two separate households.
- Lodgers, who receive accommodation only (not meals), are treated as a separate household.
- Boarders, who receive accommodation and meals (board), are treated as part of the household.

In general, persons who live in more than one household will only be treated as members of the household where they spend most of their time.

We do, however, vary from the ABS practice in how children attending boarding schools and halls of residence while studying are treated. Specifically, while these dwellings are out of scope, such individuals are treated as members of sampled households provided they spend at least part of the year in the sampled dwelling.

Note that just because the unit of data collection is the household does not mean that the unit of analysis will also be confined to the household. As Creedy (1994, p. 63) has emphasised, since household composition will change over time, researchers will typically use the individual as the unit of analysis, attributing to individuals the characteristics of the household in which each person lives (see also Duncan and Hill 1985). Indeed, in terms of tracking sample participants over time, it is the individual that is the most relevant unit.

Sample representativeness and following rules

Obviously the intent is to select a sample that is, with the exception of the exclusions noted above, representative of Australian households and residents. Nevertheless, even if this achieved in wave 1, there is the strong possibility that the sample will become increasingly less representative of the population over time. One reason for this is that the nature of the sample changes systematically through attrition. The experience with the Panel Study of Income Dynamics (PSID), possibly the world's most well known household panel (and certainly one of the longest running), however, suggests that in well-managed surveys, such concerns take a long time to assume significant proportions. Fitzgerald, Gottschalk and Moffit (1998), for example, have shown that 21 years on, and despite a loss of 50 per cent of the original sample, the sample still retained its cross-sectional representativeness.

This outcome, however, was no accident, and reflects the presence of following rules that are designed to ensure that the sample replaces itself in the same manner as the population (Duncan and Hill 1989). In the BHPS, for example, persons not included in the first wave but who subsequently became members of households containing an original sample member (e.g., as a result of birth or marriage, or because of other changes in household composition and formation) become eligible for sample inclusion. We propose to adopt similar following rules in implementing the HILDA.

Essentially, eligibility for sample inclusion after wave 1 can occur in the following ways.

(i) A child is born to, or is adopted by, an 'original' or 'continuing sample member'. This child automatically counts as an original sample member and information about that child will be collected from parents until age 15 (after which they too will become eligible for interview).

(ii) An original sample member moves into a different household with one or more new people. These new people will now become eligible for interview, but are only treated as 'temporary sample members'.

(iii) One or more new people move in with an original sample member. Again, these new people will now become eligible for interview, and are counted as temporary sample members.

All temporary sample members remain in the sample for as long as they remain in the same household as the original sample member. Temporary sample members, however, are converted to continuing sample members if they become the parent of a new original sample member birth.

As in the BHPS, we would expect that all continuing sample members would be traced and followed in subsequent waves, including persons who move into institutions. The only exceptions here would be those persons who cease to have residence in Australia or are in prison at the time of the interview. These sample members, however, may still be interviewed at subsequent waves should their status change (i.e., they return to Australia or are released from prison). Temporary sample members, on the other hand, are only followed for as long as they remain living with, or are converted to, a continuing sample member.

It is important to realise, however, that the use of these following rules do not guarantee that the panel will remain representative in a cross-sectional sense. As explained in Henstridge (2001), most following rules systems will still cause the age distribution of the sample to diverge from that of the population. Ultimately, the only way to adjust for this divergence is through the use of appropriate weights.[7]

Other potential problems with the representativeness of the sample relate to the treatment of:

(i) persons with illness conditions and disabilities that render them unable to be interviewed;
(ii) persons with English language difficulties; and
(iii) recent immigrants.

The first problem will be at least partially addressed through obtaining assisted interviews wherever possible. That said, it is recognised that there is likely to be little value in persisting with interviews of elderly persons who, because of age-related conditions, are unable to participate meaningfully in the interview. The second problem will again be handled wherever possible through assisted interviews Where this is not possible, bilingual interviewers may be used, though financial considerations mean that foreign language interviewing will only be available for a handful of the most common languages and only in a small number of cases. The third problem is even more difficult to address, though in the short- to medium-term it is not expected that the under-representation of recent immigrant arrivals will seriously bias the representativeness of the sample. In the longer-term serious consideration will need to be given to adding a 'refresher' sample of recent immigrant arrivals.[8]

Interview frequency

Interviews will be conducted on an annual basis, at least during the first three years of the survey. This seems to be accepted international practice.

The main weakness with annual surveys, of course, is that the data collected may provide only limited information about dynamics that occur within the year. That is, while it will provide extensive information about the situations of households and their members at yearly intervals, it is less likely to provide detailed information about changes in status and behaviour between interviews. For example, many people who commence an unemployment spell exit unemployment within six months of entering it.[9] Further, many (about one quarter of job seekers in the year ended February 1999) experience more than one spell of

unemployment within the year. Given these dynamics, there is a danger of missing a significant amount of important information with survey dates a year apart. The main way this being dealt with is through asking questions at each wave about the labour market histories of the respondents over the course of the previous year.

Mode of data collection

In general, international practice is that during the first few waves at least, panel surveys are conducted on a face-to-face basis. Face-to-face interviewing is generally thought to be more successful in eliciting cooperation, which is vital during the earliest years of the panel when sample member identification with the study is still developing.[10] Unfortunately, the initial budget for this study together with the expected sample size precludes conducting face-to-face interviews in all three waves. At the moment the most likely scenario is that face-to-face interviews will be conducted in the first two waves followed by computer assisted telephone interview (CATI) in subsequent waves. This, however, could change.

Apart from eliciting greater rates of respondent cooperation, use of face-to-face interviews provides other advantages. First, more difficult questions that are not easily asked over the telephone can be posed with the use of show cards and other visual aids. This is of particular significance to the HILDA Survey given the difficulties likely to be associated with collecting income data, and especially the components of income, over the telephone. Second, face-to-face interviews provide more opportunity for respondents to check records to verify their answers.[11] Third, compared with CATI, more data can be collected with a face-to-face interview (i.e., direct face-to-face contact tends to be associated with greater interviewee attention, thus facilitating the collection of more accurate data over longer interviews).

The main downside of face-to-face interviewing is that it is much more expensive and thus will mean a smaller sample size. As a consequence, the data may not provide sufficient numbers of observations to analyse important sub-groups within the population, including those of particular interest to the Commonwealth (such as persons in receipt of certain types of benefits). Other disadvantages include:

(i) greater difficulties supervising interviewers and hence maintaining ‘control’ over the interview process;
(ii) respondent discomfort from the physical presence of strangers within the home; and
(iii) a loss of statistical efficiency from having to use a clustered design.

Sampling

The use of face-to-face survey methods in wave 1 dictates some form of direct physical sampling which, in turn, almost always involves a clustered sample design (to minimise travel costs). The sampling frame is thus an area-based frame consisting of Census Collection Districts (CDs), each of which typically consists of approximately 200 to 250 households.

In total, the initial sample in wave 1 consists of 488 randomly selected CDs. This sample of CDs was stratified by State, and within States, by metropolitan and non-metropolitan regions. Despite the region-based stratification, however, the smaller States and Territories were not over-sampled. This reflects both the focus of the HILDA on producing nation-wide population estimates, and our view that any benefits from a differential probability approach to sampling are outweighed by the negative impacts on overall statistical efficiency.

Trained interviewers visited each of the selected CDs (in June 2001) to provide a full listing of all private residential dwellings within the CD boundaries. A random starting dwelling was then chosen for each CD with further dwellings selected according to a pre-defined skip pattern and a randomly determined route. The intent was to select an average of 23 dwellings per CD. The actual number selected, however, varied depending on projected variations across CDs in response rates and in occupancy rates.

In total, the initial sample comprised over 11,000 dwellings, making it larger than the initial sample sizes in most other studies of this type, including the PSID, the GSOEP and the BHPS. Multiple households within dwellings means the effective sample is even larger, with all households at dwellings with three or fewer households selected into the sample, and up to three randomly selected households at dwellings with more than three households.

In addition to facilitating control over the sample selection process, the pre-enumeration of the selected CDs provided two other key advantages. First, it enabled us to mail a primary approach letter and brochure about the study to all selected households in advance of an interviewer arriving at the door. Second, it provided us with the necessary data from which to calculate sample selection probabilities.

3. Response and attrition

Wave 1

We have been aiming to achieve completed interviews with at least one household member at 70 per cent of all households selected into our wave 1 sample. Further, in 65 per cent of sampled households, we were hoping to obtain interviews with all members of those households aged 15 years and over. These target response rates were based largely on 'best-practice' international experience. For example, in the first wave of the BHPS (conducted in 1991) at least one interview was completed at 74 per cent of eligible households, with complete coverage of eligible adults achieved in 69 per cent of households. Similarly, at the commencement of the PSID, 76 per cent of households were reported as being successfully interviewed in the first year (Brown, Duncan and Stafford 1996, p. 158). [12]

The 65 per cent full household response rate, however, was based on the assumption that we would be using proxy interviews where in-scope sample members were unavailable during the interview period. Concerns about ensuring we had obtained informed consent, however, caused us to abandon pursuing proxy interviews and hence it is expected that we will fall short of our 65 per cent target.

Furthermore, outcomes from a Dress Rehearsal of the survey conducted in June 2001 suggest that these response rates are ambitious, with interviews only being conducted successfully at 55 per cent of sampled households. Nevertheless, there are at least three reasons not to get too pessimistic about this outcome. First, the Dress Rehearsal sample was concentrated in Sydney, and it is well established that survey response rates are relatively low in Sydney (see Bednall, Cavenett and Shaw 2000). Second, the fieldwork period for the Dress Rehearsal was relatively short. Third, unlike the Dress Rehearsal, a cash incentive – up to $50 for every household where all household members are successfully interviewed – will be used in wave 1 of the HILDA Survey to encourage respondent cooperation.[13]

Nevertheless, the difficulty of obtaining high response rates cannot be understated. Moreover, there are a number of factors that pose particular difficulties for HILDA compared with previous surveys. First, the fieldwork for the first wave of the HILDA Survey is being undertaken during a period when there are many other survey organisations competing for respondent time. Most obviously, it was commenced in the wake of the 2001 Census, which might be expected to create some hostility towards yet another government-sponsored data

collection initiative. Further, with a Federal Election taking place toward the end of the field period, there be a higher than normal risk of sample members being asked to participate in opinion poll surveys.

Second, in terms of delivering complete household coverage (i.e., obtaining personal interviews with all in-scope household members), the length of the fieldwork period for the first wave of the HILDA Survey is, at just 4 months, comparatively short. Most other household panel surveys have fieldwork periods that stretch to 9 months and beyond.

Finally, and most importantly, there is a growing amount of evidence from around the world that indicates that respondent cooperation with survey organisations is declining. Bednall, Cavenett and Shaw (2000), for example, report evidence from a large long-running US telephone-based opinion survey that reveals a marked rise in refusal rates during the 1990s, after being relatively stable during the 1980s. Even more relevant for the HILDA Survey, both the GSOEP and BHPS have in recent years added new samples to their studies, and in both cases the rates of response were well down on the initial response rates reported for their original samples.

In wave 9 of the BHPS, conducted in 1999, two additional household samples were recruited in Wales and Scotland. Partial coverage was only achieved at 63 per cent of the selected sample, which represents an 11 percentage point decline compared with the rates achieved in 1991. Further, the proportion of households where complete adult coverage was obtained fell even further – 54 per cent compared with the 69 per cent reported eight years earlier.

Similarly, in 2000 the GSOEP added a major new fresher sample. As with the BHPS, achieved response rates were much lower than those reported for the original sample surveyed in 1984. Interviews were obtained at only 51 per cent of their new households sampled in 2000, compared with 63 per cent in 1984.[14]

Non-response, and the consequent potential for respondent bias, is thus likely to be a major issue for the HILDA Survey.

Waves 2 and 3

In a longitudinal design there is potential for non-response at every wave. At one level this is less problematic than non-response at wave 1. In particular, since we have detailed information on the characteristics of all respondents at wave 1, it should be easier to apply

weights to the data to compensate for any bias this non-response gives rise to. Such procedures, however, are only likely to be effective in the short-run. Over the longer-term it is important to minimise attrition because of the potential for the 'movers' to be quite different from the 'stayers' in ways that may not be observable at wave 1. Further, high rates of attrition have obvious detrimental effects on sample size.

International experience tends to suggest that attrition is highest in the first two years of the survey and then stabilises. Indeed, in the PSID, attrition rates fall to as low two to three per cent by wave 3. Fourteen per cent of the sample, however, was lost in wave 2 (Brown et al. 1996, p. 158).

Attrition rates in most other studies are, if anything, higher. Again we can point to the experience with the BHPS. In that survey, interviews were conducted with 88 per cent of the wave 1 respondents in wave 2. By wave 8 the proportion had fallen to 68 per cent. A good proportion of the attrition by wave 8, however, was due to deaths or because the sample member had moved out of scope. In total, after adjusting for deaths and movements out of scope, 75 per cent of the original sample remained in scope at wave 8.

The preceding discussion, however, is based on surveys that used face-to-face interviews, at least in the first few waves; attrition rates can be expected to be higher when using CATI (which may be used for HILDA from wave 3 on). Certainly, and as noted earlier, response rates are lower in cross-section surveys that use telephone methods.

Strategies being implemented as part of the HILDA Survey which are intended to minimise attrition include the following:

- The inclusion of tracking questions in the questionnaire seeking contact details (postal addresses, telephone numbers, and email addresses), information on future movement intentions, and names of relatives and friends not living at the same address.
- Sending a thank you gift (a calendar) to all interviewees following the interview, together with change-of-address cards for notification of any intended moves.
- Maintaining contact with participants between survey waves through a respondent newsletter providing summary information about the study.
- Maintaining a 1800 telephone number so that participants can contact the fieldwork agency and/or the Project Director / Survey Manager.

- Seeking forwarding addresses or telephone numbers from non-sample members at the address or telephone number of the original sample member.
- Using the electronic White Pages and Australia Post to pursue contact details for persons who have changed address.

4. Survey content and instruments

In wave 1 the HILDA survey comprises four main instruments. These are as follows:

(i) the Household Form;
(ii) the Household Questionnaire;
(iii) the Personal Questionnaire; and
(iv) an individual Self-completion Questionnaire.

Household Form

The Household Form provides information obtained prior to interview or which can be observed prior to making contact with members of the household, as well as recording basic information about the composition of the household immediately after making contact. In effect, it is a type of "master document" that helps the interviewer determine who they need to interview, as well as alerting them to what sections of the Person Questionnaire are relevant for each respondent.

The type of information recorded on this form includes:

(i) dates and times of each visit;
(ii) outcome of each visit;
(iii) for refusing households, reason for refusal;
(iv) number of households at the address;
(v) whether the residence is in-scope (that is whether it is a private residence and is occupied on an ongoing basis);
(vi) household composition – name, date of birth, age and sex of all household members;
(vii) other selected personal characteristics of household members (e.g., health/disability status, marital status, English language ability, labour force status); and
(viii) relationships between household members.

Interviewer observations – for example, on the type of residence and its condition – are also recorded here.

Household Questionnaire

This questionnaire collects information about the household rather than about individual household members per se, and only has to be administered to one member of the household. In practice, however, interviewers are encouraged to be flexible. If more than one household member wishes to be present at the interview this is perfectly acceptable; indeed, it should improve the quality of data collected. Further, interviewers are given the flexibility to deliver part of this interview to one household member and part to another. Indeed, this is often required given this questionnaire includes questions on child care that needed to be asked of the primary care giver.

Key data items collected here cover the following:

(i) housing characteristics (number of bedrooms, ownership status and value);
(ii) details of mortgages, home loans and rent payments;
(iii) vehicle ownership;
(iv) household grocery and food expenditures; and
(v) child care arrangements.

The combined time spent in the household administering this schedule together with those details within the Household Form obtained directly from household members is expected to average 10 minutes.

Person Questionnaire

This instrument is administered to every person aged 15 years and over (on 30 June) in the household, and is expected to average 35 minutes. It will provide the bulk of the HILDA Survey data items. The structure of this questionnaire, together with examples of the types of issues covered, is summarised in Table 1.

Note that while explicit mention of 'life events' has not been made, major life events will be fully documented within the scope of the survey. Changes in household composition (e.g., arising from births, death, the formation and dissolution of cohabiting relationships, and children leaving home) will obviously be captured in the survey, as will other significant life

Table 1: Structure of the Person Questionnaire

Section		*Types of issues covered*
A	Country of birth and language	Birthplace Year of arrival in Australia (if born overseas) English language background Aboriginality
B	Family background	Parental presence at age 14 Age when first left home Parents' birthplace Parents' occupation
C	Education	Age left school and highest year of school attended Post-school qualifications obtained Current educational activities
D	Employment history and status	Numbers of years since left full-time education Summary measures of labour force history Current labour force status Current employment status Time since last worked, if not currently employed Characteristics of last job, if not currently employed
E	Persons in paid employment	Characteristics of main job (occupation, hours, tenure, etc.) Employer characteristics (industry, size, sector) Perceived employment prospects Job satisfaction Retirement intentions
F	Persons not in paid employment	Job search activity Types of difficulties experienced in finding a job Desire to work (if not looking) Reasons for not looking for work (if not looking) Reservation wage Perceived employment prospects
FG	Calendar	History of employment, job search and education history since July 1 of preceding year Mutual obligation activities
G	Income	Current income from: wages and salaries government pensions and benefits Financial year income from: wages and salaries government pensions and benefits businesses savings and investments other sources Credit card use

Table 1 (cont'd)

Section		*Types of issues covered*
H	Family formation	Characteristics of children Child support payments Patterns of contact parents and children (where they live apart) Likelihood of having children
J	Partnering	Marital history De facto relationships Marriage intentions (for persons in de facto relationships)
K	Life situation	Health status Life satisfaction Attitudes to life in Australia Reasons for moving (if moved in last 12 months)
T	Tracking	Contact details Movement intentions Names of friends / relatives
Z	Interview situation	Respondent cooperation and understanding Problems affecting interview

events, such as purchasing large assets (e.g., a home), moving house, changing employment, loss of a job, retirement from paid work and major changes in health status.

Self-completion Questionnaire

Additional data will be collected via a self-administered questionnaire. It is hoped that, in most cases, this questionnaire will be collected by the interviewer at a later date, but we are also allowing for these questionnaires to be returned to the fieldwork company by mail.[15]

The questionnaire comprises six main sections. These are as follows:

A General health and well-being (the SF-36 Health Survey);

B Lifestyle and living situation

C Personal and household finances

D Attitudes and values about work

E Your job and the place where you work

F Parenting

The questionnaire is 16 pages long and is expected to take 20 to 25 minutes to complete.

Topics for later waves

Of course, there is a range of other topics that ideally we would like to have information on, the inclusion of which has been constrained by a lack of space. It is anticipated, however, that at least some of these topics will be covered during the later waves of the survey. At this stage a dedicated module on savings and wealth has been proposed for inclusion in wave 2. Other possible topics for later waves include:

- literacy and numeracy;
- career aspirations;
- recent training and education experiences; and
- health-related conditions and behaviours.

5. Timetable

The fieldwork for wave 1 commenced in late August 2001 and runs until mid-December, facilitating a public release of the wave 1 data during the second half of 2002.

In later waves the fieldwork period will be extended through to April. That is, the fieldwork period will run from September through to April each year. This longer fieldwork period is consistent with the practice in the BHPS, and is needed to maximise the likelihood of locating and interviewing all sample members. This longer fieldwork period, of course, will mean data releases in later years will also occur later. At this stage it is expected that public releases of the wave 2 and wave 3 data, together with the relevant longitudinal data files, will take place by the end of 2003 and 2004 respectively.

6. Confidentiality and data access

Risk of disclosure

As with all unit record data sets, there is a risk that the identity of some respondents will be revealed through the information collected in the survey. This risk, however, is even higher with longitudinal data because of the large array of events and transitions that will be documented over the lifetime of sample members.

Solutions to the risk of disclosure are still being explored with the project sponsor, the Commonwealth Department of Family and Community Services (FaCS). It is inevitable,

however, that some of the information collected will not be made available on the public-use file. In particular, some variables may be provided in a form that is more aggregated than that in which it was collected. This, for example, would include occupation and residential location variables.

We would, however, expect that users who are able to demonstrate a strong need for any data that have been removed would still be able to access that data, but only after agreeing to a stringent set of use conditions concerning access, security and disposal of the data, and only if they can satisfy FaCS that those conditions are likely to be met.

Dissemination

The confidentiality issue can also be handled through regulating user access to that data. Again this is an issue that we are still exploring with FaCS. However, it is our hope that the data access policy will not be too restrictive.

The GSOEP provides one model. There the principal end product is a public-use micro-data file with front-end software (e.g., SAS or SPSS) that will allow users to select variables and sub-populations of interest. Data releases occur each year with data from subsequent waves merged into the earlier waves. These public-use micro-data files are made available to bona fide researchers and research organisations on a CD-Rom for a nominal charge. All users are required to sign an agreement that specifies conditions under which the data are to be used and stored.

7. What's in it for economists?

To illustrate the many and varied research uses that could be made of the HILDA data, we only have to look to the overseas experience. For example, according to the PSID website, at the time of writing more than 2000 articles had been published in academic journals and books using PSID data, with the number rising exponentially over time. Similarly, the literature database available from the GSOEP website currently contains 2250 different entries. Moreover, and despite the diverse range of topics covered by these surveys, economic research overwhelmingly dominates this literature.

A guide to the range of topics that economists have examined using the PSID data is summarised in a review by Brown et al. (1996). In particular, they highlight how significant the PSID data have been in improving our understanding of:

(i) intertemporal models of labour supply;
(ii) whether real wages are pro- or countercyclical;
(iii) labour mobility;
(iv) consumption dynamics;
(v) extended-family behaviour;
(vi) poverty, welfare and income dynamics;
(vii) the intergenerational transmission of economic status; and
(viii) the longer-run antecedents of events like marital breakdown, unemployment and retirement.

Of particular attention to Australian users, the HILDA data would appear well suited to a number of other purposes. Most obviously, one of the major purposes of HILDA is to help understand the impact of variations in income support arrangements on Australians and their families. The HILDA Survey is thus seeking detailed information about income received from government sources, including the type of benefit, the amount received and the length of time spent in receipt of those benefits. Together with other information collected within this survey, such data should be very useful in analysing a range of questions and topics, including:

(i) the risks and factors promoting welfare dependence;
(ii) the long-run redistributive impact of the tax and transfer system; and
(iii) the impact of variations in income support arrangements on work incentives.

The HILDA Survey also collects detailed information about employment and work, which should be of particular interest to researchers interested in the changing nature of work and its impact on individuals and families. There is, for example, considerable debate over the question of whether casual and part-time work experiences when young assist or hinder long-run employment prospects. Unfortunately, most longitudinal data available in Australia that might assist researchers analyse this question simply do not track people over sufficiently long periods of time. Gaston and Timcke (1999), for example, have attempted to analyse this question using data from the Australian Youth Survey, but were only able to consider transitions over a four-year period. Relatedly, there is currently debate about the impact of

new working arrangements on working hours and job security, and in turn, how changes in these factors affect job satisfaction, productivity and quality of life. These are all issues that the HILDA data should be well placed to help us analyse.

Another labour market issue is the distribution of work across households. The recent report of the Reference Group on Welfare Reform (2000), for example, noted the growing divergence between the number of 'work rich' households and 'work poor' households, but how permanent are these states and what are their consequences?

Another area of research where the HILDA will provide a large amount of information is family and household dynamics. Ermisch (2000), for example, has illustrated how panel data from the BHPS can be used to model the processes of family and household change, such as marital separation and divorce, cohabitation, and parental home leaving. More importantly, with each successive wave of data, the potential to analyse the consequences of such change (especially in terms of income) grows. Relatedly, the HILDA Survey stands apart from other surveys around the world in the amount of information it is collecting about children, and especially the decision to have children, care arrangements and contact with children. Such data, therefore, should facilitate analyses of topics such as the interrelationship between the timing of children and labour force participation, and the impact of child care availability and costs on labour supply decisions.

Ultimately, however, it is impossible to do justice to the research potential of these data in the brief space allocated to it here – there are just too many issues and topics that are being covered by the various HILDA Survey instruments. To fully comprehend the research possibilities all we can suggest is that researchers make themselves more familiar with the data, and the best place to start here is with the survey instruments.

References

Barr, N. 1999, 'Fundamentals of social security analysis', *Australian Social Policy* 1999/1, pp. 7-29.

Bednall, D., Cavenett, P. and Shaw, M. 2000, 'Response rates in Australian market research', Paper presented at the Market Research Society of Society Annual Conference, Sydney, October.

Brown, C., Duncan, G.J. and Stafford, F.P. 1996, 'Data watch: The Panel Study of Income Dynamics', *Journal of Economic Perspectives*, vol. 10, no. 2, pp. 155-168.

Collins, M., Sykes, W., Wilson, P. and Blackshaw, N. 1988, 'Nonresponse: The UK experience', in R.M. Groves, P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II and J. Waksberg (eds), *Telephone Survey Methodology*, John Wiley and Sons, New York.

Creedy, J. 1994, 'The statics and dynamics of income distribution: An introductory survey', *Australian Economic Review*, No. 108 (4/94), 51-71.

de Leeuw, E.D. and van der Zouwen, J. 1988, 'Data quality in telephone and face to face surveys: A comparative meta-analysis', in R.M. Groves, P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II and J. Waksberg (eds), *Telephone Survey Methodology*, John Wiley and Sons, New York.

Donovan, R.J., D'Arcy, C.D., Holman, J., Corti, B. and Jalleh, G. 1997, 'Face-to-face household interviews versus telephone interviews for health surveys', *Australian and New Zealand Journal of Public Health*, vol. 21, no. 2, pp. 134-140.

Duncan , G. 2000, 'Using panel studies to understand household behaviour and well-being', in D. Rose (ed.), *Researching Social and Economic Change: The Uses of Household Panel Studies*, Routledge, London.

Duncan, G.J and Hill, D.H. 1985, 'Conceptions of longitudinal households: Fertile or futile?', *Journal of Labor Economics*, vol. 3, no. 4, pp. 508-532.

Duncan, G.J. and Hill, D.H. 1989, 'Assessing the quality of household panel data: The case of the Panel Study of Income Dynamics', *Journal of Business and Economic Statistics*, vol. 7, no. 4, pp. 441-452.

Ermisch, J. 2000, 'Using panel data to analyse household and family dynamics', in D. Rose (ed.), *Researching Social and Economic Change: The Uses of Household Panel Studies*, Routledge, London.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. 1998, 'An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics', *Journal of Human Resources*, vol. 33, no. 2, pp. 251-299.

Gaston, N. and Timcke, D. 1999, 'Do casual workers find permanent full-time employment?', *Economic Record*, vol. 75, no. 231, pp. 333-347.

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

Haisken-DeNew, J.P. 2001, 'A hitchhiker's guide to the world's household panel data sets', *Australian Economic Review*, vol. 34, no. 3, forthcoming.

Harding, A. and Szukalska, M. 2000, *Financial Disadvantage in Australia – 1999: The Unlucky Australians*? The Smith Family, Sydney.

Henstridge, J. 2001, 'Weighting and Imputation Strategies for the HILDA Survey', HILDA Discussion Paper Series No. 3/01, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Kalton, G. and Citro, C.F. 2000, 'Panel surveys: Adding the fourth dimension', in D. Rose (ed.), *Researching Social and Economic Change: The Uses of Household Panel Studies*, Routledge, London.

McDonald, P., Evans, A., Baxter, J. and Gray, E. 2000, 'The Negotiating the Lifecourse Survey Experience: Implications for HILDA', Paper presented to the Panel Data and Policy Conference (Department of Family and Community Services), Rydges Canberra, 1-3 May.

Norris, K. and Wooden, M. (eds) 1996, *The Changing Australian Labour Market* (EPAC Commission Paper No. 11), AGPS, Canberra.

Reference Group on Welfare Reform (Chmn: P. McClure) 2000, *Participation Support for a More Equitable Society: Final Report of the Reference Group on Welfare Reform*, Commonwealth Department of Family and Communitry Services, Canberra.

Rose, D. 2000, 'Household panel studies: An overview', in D. Rose (ed.), *Researching Social and Economic Change: The Uses of Household Panel Studies*, Routledge, London.

Wooden, M. 1997, 'Rapporteur comments', in Industry Commission, *Changing Labour Markets: Prospects for Productivity Growth: Workshop Proceedings*, Industry Commission, Melbourne.

Wooden, M. 2001, 'Design and management of a household panel survey: Lessons from the international experience', HILDA Discussion Paper Series No. 2/01, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Endnotes

[1] This survey program combines and builds on the earlier cohorts within the Youth In Transition, Australian Longitudinal Survey and Australian Youth Survey programs.

[2] The sample employed in this study was relatively small, with just 2231 persons participating in wave 1 (819 of whom were incorporated from the pre-test). Consequently, even with modest rates of attrition, it is unlikely that it will be worthwhile following this sample for more than a few years. For more details about this survey, see McDonald et al. (2000).

[3] Indeed, according to Harding and Szukalska (2000, p. 6) the proportion has fallen, from 14.6% in 1982 to 13.3% in 1999.

[4] More information about the HILDA Survey can be obtained from the HILDA Survey web site. At this site updates about the progress of the survey in the field can be obtained. In addition, copies of all the survey instruments can be viewed and downloaded. Also available is a series of Discussion / Technical Papers that provide more specialised information about aspects of the HILDA. The web site address is: http://www.melbourneinstitute.com/hilda/.

[5] Extensive documentation on the GSOEP and the BHPS can be found on the world wide web. Their respective web addresses are:
http://www.diw.de/English/gsoep
http://www.irc.essex.ac.uk/bhps/index.htm.
The key features of these and other international household panel studies are also briefly summarised in Haisken-DeNew (2001) and Wooden (2001).

[6] Note that long-term caravan park residents are treated as residing in private dwellings.

[7] Henstridge (2001) outlines in much greater detail how such weighting factors might be constructed.

[8] Immigrant refresher samples have been added to both the PSID and the German Socio-Economic Panel. In both cases, however, these new samples were not considered warranted until after the survey had been running for many years. The German study commenced in 1984 with the immigrant refresher added in 1994 and 1995. The PSID has been running since 1968 and the new immigrant sub-sample was only added in 1997 (though a Latino supplement was added in 1990 only to be discontinued five years later).

[9] According to data from the ABS Survey of Labour Force Experience for the year ended February 1999 (ABS cat. no. 6206.0), 64 per cent of persons who had been looking for work during that year spent less than 6 months of the year engaged in job search. These data, however, will understate unemployment duration given they do not pertain to completed spells of unemployment.

[10] Response rates in surveys employing face-to-face interviews have generally been found to be higher than surveys using other methods. In a meta-analytic review involving 28 different studies where telephone and face-to-face methods were compared, de Leeuw and van der Zouwen (1988) reported an average response rate for face-to-face interviews of 75 per cent. This compared with an average of 69 per cent for the telephone interview method. Further, both Australian and UK evidence suggests that the gap may be even greater (Collins et al. 1988, Donovan et al. 1997). Indeed, in the Australian study reported by Donovan et al. (1997) the reported response rates for telephone interviews was just 46 per cent, compared with 57 per cent for face-to-face interviews.

[11] Though interview time constraints is likely to mean that such records checking is likely to be discouraged.

[12] The 76 per cent response rate reported for the PSID does not take into account the fact part of the sample was derived from an earlier survey, 25 per cent of which refused to allow their names to be sent to the PSID survey team. Adjusting for this would reduce the actual response rate to 69 per cent.

[13] Some indication of the potential effectiveness of cash incentives is provided by an experiment undertaken in conjunction with the National Adult Literacy in the USA in the late 1980s, and reported in Groves and Couper (1998, p. 281). That study found that a US$20 per head incentive was associated with a 71% response rate, compared with 64% among those not offered any cash incentive.

[14] As reported in personal correspondence from Dr John Haisken-DeNew. Note that unlike the BHPS, in the GSOEP incomplete households are excluded from the responding sample for the first wave.

[15] In the Dress Rehearsal, completed SCQs were received from 77% of all individuals interviewed.