



DEPARTMENT OF
**FAMILY AND
COMMUNITY
SERVICES**



THE UNIVERSITY OF
MELBOURNE

**HILDA PROJECT DISCUSSION PAPER SERIES
NO. 2/01, JUNE 2001**

**Design and Management of a Household Panel Survey:
Lessons from the International Experience**

Mark Wooden

**The HILDA Project was initiated, and is funded, by the Commonwealth
Department of Family and Community Services**



Executive Summary

- Household-based panel studies are now conducted in most industrial nations. These studies, however, vary quite markedly in terms of design, collection method and strategies adopted for both processing and disseminating data. There is no single approach that is universally accepted as the “best”.
- The data from many of these household panel studies have stimulated a large amount of research analysing change and dynamic behaviour. This, in turn, has inevitably fed into improved policy-making.
- The proposed survey design for the Household, Income and Labour Dynamics in Australia (HILDA) Survey is closely based on the design utilised in leading European studies, and especially the British Household Panel Survey (BHPS) and the German Socio-Economic Panel (GSOEP). The North American studies, on the other hand, are quite different in that data are only collected from one household member. This means that data collected about individuals within the household will tend to be subject to more measurement error in the North American studies. Perhaps, more importantly, this approach renders these studies less receptive to more subjective questions. Finally, collection of data from all household members permits more complicated analyses of family effects and intra-household dynamics. For these reasons, the approach advocated in the HILDA is preferred.
- Consideration needs to be given to eventually adding a refresher sample for recent immigrant arrivals. Such a refresher, however, will not be needed inside the current funding window.
- International experience suggests that survey frequency should be annual. Higher frequency is too burdensome on respondents, while less frequency is insufficient to adequately capture the dynamics of interest.
- 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. Some studies then convert to telephone interviewing at a subsequent wave.
- The time imposition on respondents is an issue that interviewers need to be sensitive to, and will need to be monitored during pre-testing of the HILDA Survey.
- The inclusion of a self-completion instrument to complement personal interviews would be advantageous for a number of reasons. First, it might provide a superior vehicle for collecting information about sensitive topics. Second, it may be a more time efficient vehicle for collecting attitudinal information. Third, and most obviously, it provides a cost-effective avenue for enriching the interview-based data.

- Based on international experience, the content of the HILDA can be expected to change significantly with the passage of time, as it evolves in line with what are the major economic and social issues of the day.
- The keys to achieving high response rates and low rates of attrition appear to be at least fourfold:
 - (i) use of respondent incentives;
 - (ii) a long field period;
 - (iii) a committed and motivated interviewer workforce; and
 - (iv) a fieldwork agency that works to academic standards.

Other strategies identified as influencing response and attrition rates include: distribution of marketing material in advance of each survey round; allocating considerable effort to converting refusers; distributing feedback to sample members; assigning the best interviewers to the most difficult cases; and providing financial incentives to the fieldwork agency.

- Recent experience with the introduction of new samples in both the BHPS and the GSOEP suggest that response rates have fallen markedly during the 1990s.
- The international consensus is that data linking that requires informed consent of participants should, at a minimum, be avoided during the first few waves of a panel survey. This reflects concerns about the potential impact of seeking consent for such linking on respondent cooperation and the strong likelihood that the linked data cannot be made available to many users anyway.
- International experience strongly suggests that a realistic timetable is important in delivering a high quality product. Most studies, for example, provide for a two-year planning period. Second, the fieldwork period for each wave typically extends up to anywhere from 6 to 9 months. Third, most studies allow at least 9 months for the processing of data from each wave.
- In line with both the British and German studies, the HILDA Survey will be marketed to sample members as the “Living in Australia” survey.

Introduction

As discussed in Wooden and Watson (2000), most industrial nations now conduct large-scale, representative household-based panel (i.e., longitudinal) surveys designed to collect a large amount of information about households and the members of those households. Examples of large national panel studies conducted in other countries where the case unit is a household or some other similarly related unit (e.g., family) include the:

- (i) Survey of Labour and Income Dynamics (Canada);
- (ii) German Socio-economic Panel Study;
- (iii) Indonesia Family Life Survey;
- (iv) Korean Labor and Income Panel Study;
- (v) Dutch Socio-economic Panel;
- (vi) Swedish Panel Study of Market and Nonmarket Activities;
- (vii) Swiss Household Panel Study;
- (viii) British Household Panel Survey; and
- (ix) Panel Survey of Income Dynamics (USA).

In addition, 15 European countries, including four of those referred to above, are currently part of the European Community Household Panel Study.

Australia, however, is a notable exception. While longitudinal data collections do exist, they typically focus on relatively small sub-groups of the population. Australian policy-makers and researchers thus do not have access to data that are both representative of the Australian population 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.

The absence of such data is a major problem for both social scientists and policy-makers interested in the analysis of change and of dynamic behaviour more generally. Without panel data such issues cannot be well analysed, at least not within a statistical framework. Cross-section data that contain retrospective elements, of course, can help, but as observed by Rose (2000, p. 20), responses to retrospective questions are affected by recall bias and by subsequent events.

The advantages and value of longitudinal data is ably demonstrated by the experience with the PSID. According to the PSID website, more than 2000 articles have 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. More importantly, panel data have clearly been important in influencing policy. It is well recognised that analyses of the PSID data have had a major influence on policy, especially in the area of poverty and welfare. Even more obvious, the European Community Household Panel study was established with the express aim of monitoring the effects of the Single European Market (Barreiros 1995).

It is thus against this background that Australia is finally about to get its first major household panel survey. In late 2001 the first wave of the Household, Income and Labour Dynamics in Australia (or HILDA) Survey will be conducted.

As a relative latecomer into this field, Australia has at least one major advantage – we are able to draw on the experience of the overseas studies referred to above. The principal aim of this paper, therefore, is to summarise the major lessons for the design and development of the HILDA Survey from the overseas experience.

The paper draws heavily on the outcomes from a series of consultations with leading figures associated with household panel studies that took place in the second half of 2000. The consultations were spread over two trips. The first trip covered the first two weeks of September and involved meetings in three countries: Germany, the United Kingdom and The Netherlands. The consultations covered persons associated with three leading European panel studies – the German Socio-Economic Panel (GSOEP), which commenced in 1984, the British Household Panel Study (BHPS), the first wave of which was conducted in 1991, and the Dutch Socio-Economic Panel, which, after one false start, got under way in 1985.

The second trip, to North America, took place in late October and early November. The principal purpose of this trip was to learn from the experience of both the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey (NLS) program conducted by the US Bureau of Labour Statistics. The PSID is the world's longest running panel focusing on household, income and labour dynamics, while the NLS studies have a reputation without peer in terms of achieving high response rates and low attrition rates. The trip also included attendance at a conference on the use of panel data for international comparisons and an associated workshop on the design and management of panel studies. Held at the University of Michigan, participation at this conference and seminar afforded contact with a wide group of persons with whom direct contact would not otherwise have been possible. It included, in addition to those mentioned above, representatives from other groups running panel studies (e.g., the Canadian Survey of Labour and Income Dynamics or SLID, the Panel Study of Employment and Labour in Luxembourg, the Indonesian Family Life Study, the Korean Labour and Income Panel Study, and the Japanese Consumer Finances Survey) or are in the planning stages for such a study.

The range of issues covered during these discussions was extremely broad, extending to all aspects of implementing a new household panel study. Included here were sample and survey design, data collection, response and attrition, survey content, the derivation of statistical weights, confidentiality requirements, data management and storage issues, dissemination, and the type of uses made of the data.

International Household Panel Studies: A Comparative Overview

A summary of six leading panel studies is provided in Table 1. While these studies have many similarities, no two studies are a like. In terms of overall design, the studies that come closest to what has been proposed for the HILDA are the three European studies – the BHPS, the GSOEP and the Dutch Socio-Economic Panel. These studies employ indefinite life designs and involve interviews with all adult

Table 1: Major International Panel Studies: A Summary

	GSOEP	BHPS	Dutch Socio-economic Panel
Host organisation	German Institute for Economic Research	Institute for Social and Economic Research, University of Essex	Statistics Netherlands
Design	Indefinite life panel	Indefinite life panel	Indefinite life panel
Commenced	1984	1991	1985
Initial sample size	5900 households	5500 households	5000 households
Reference population / data collection unit	All private households. All members aged 16 years or over are interviewed.	All private households. All members aged 16 years or over are interviewed.	All private households. All members aged 16 years or over are interviewed.
Over-sampling	At Wave 1 a separate sub-sample of foreign-born households was selected.	None at Wave 1, but a low-income sample from the European Community Household Panel Sample added in 1997 and new Scottish and Welsh sub-samples added in 1999.	None.
Frequency	Annual	Annual	Twice yearly prior to 1990. Annual since.
Collection mode	PAPI. Mix of personal and self-completion. Self-completed questionnaires typically collected by interviewer. Begun shifting to CAPI in 2000.	Personal PAPI with short self-completion questionnaire. Shifted to CAPI in Wave 10 (2000).	Personal PAPI before switching to CAPI in the early 1990s. Self-completion for income questions (returned by mail).
Proxy interviews	No.	Yes (3.4% of interviews in wave 1).	Yes – widespread.
Wave 1 response rates	61% West Germans; 68% foreigners (but note that incomplete households omitted). 1998 refresher sample – 54%. 2000 new sample – 51%.	69% including proxies (74% of households supplied at least one interview). 1999 Scottish / Welsh sample – interviews completed with at least one person at 63% of households.	Approx. 55%. Top-up samples added each year and average just 35%.
Attrition(a)	10% wave 2; 7% wave 3. 3% by wave 7. Stable since.	12% wave 2; 10% wave 3. 3.4% by wave 8.	High. Only 30% of original sample left after 12 waves.
Fieldwork	Full range of data collection, management and processing functions contracted out.	Only data collection contracted out. Management of panel and cleaning of data undertaken in-house.	Undertaken entirely in-house.
Data distribution	CD-Rom. Access restricted to bona fide researchers for specific purpose research.	Deposited in UK Data Archive.	A highly priced CURF.

Table 1 (cont'd)

	PSID	SLID	NLSY79(b)
Host organisation	Institute for Social Research, University of Michigan	Statistics Canada	Center for Human Resource Research, Ohio State University
Design	Indefinite life panel	Rotating medium life (6 years) panel	Single cohort panel
Commenced	1968	1993	1979
Initial sample size	4800 families	Approx. 15000 households in each panel	12686 individuals
Reference population / data collection unit	Heads of family units who have been continuously resident in the USA for at least 2 years.	Private households in the 10 provinces, with the exception of the Indian reserves. Interviews conducted with only one member of the household.	Persons aged 14-21 as of December 31, 1978.
Over-sampling	A sub-sample of 1872 low-income families was drawn from an earlier survey conducted by the US Census Bureau. A new Latino supplement was added in 1990 but discontinued after 1995.	Sample based on the Labour Force Survey and hence sample selection probabilities vary across regions (i.e., smaller regions over-sampled).	Supplemental samples were drawn so as to over-sample: (i) hispanic, black and economically disadvantaged youth; and (ii) members of the military.
Frequency	Annual until 1997; every other year since.	Annual (but with 2 interviews conducted 6 months apart).	Annual until 1994; every other year since.
Collection mode	Personal PAPI from 1968 to 1972. Mainly telephone since 1973. CATI introduced in 1993.	CATI.	Personal PAPI until 1992. CAPI from 1993. CATI also increasingly used.
Proxy interviews	The need for proxy information is a fundamental feature of the survey design.	The need for proxy information is a fundamental feature of the survey design.	No (not a household-based survey).
Wave 1 response rates	76% (but "real" response only about 69%).		87% (but "real" response rate only about 81%).
Attrition (a)	11.5% wave 2. Between 1.5% and 3.1% thereafter.		4.3% wave 2. Average of 1.9% thereafter.
Fieldwork	Undertaken by Survey Research Centre, a separate unit within the Institute for Social Research.	Undertaken entirely in-house.	Undertaken by NORC, a survey centre based at the University of Chicago. All data processing functions undertaken by CHRR.
Data distribution	Freely available from web site.	Currently available only via remote access or on-site access at StatCan. (First two waves had been released as CURFs.)	CD-Rom distributed at cost to all who request it.

Notes to Table 1

Acronyms

BHPS	British Household Panel Survey
CAPI	Computer assisted personal interviewing
CATI	Computer assisted telephone interviewing
CHRR	Center for Human Resource Research
CURF	Confidentialised unit record file
GSOEP	German Socio-economic Panel
NLSY	National Longitudinal Study of Youth
NORC	National Opinion Research Center
PAPI	Pencil and paper interviewing
PSID	Panel Study of Income Dynamics
SLID	Survey of Labour and Income Dynamics
StatCan	Statistics Canada.

- a Attrition rates are typically adjusted for deaths.
- b The NLSY79 is one of a number of longitudinal studies conducted as part of the NLS program within the US Bureau of Labour Statistics.
- c The Economically disadvantaged / non-Hispanic supplemental sample was dropped after the 1990 interview.

members of the household. The panels are indefinite in the sense that additions to the household (e.g., children) are automatically added to the sample.¹ In contrast, while the PSID also employs an indefinite life design (indeed, it may well have been the first survey of this type in the world), interviews are only conducted with the family head, who answers on behalf of all other members of the family unit. Similarly, the SLID in Canada also involves interviews with only person per household, but unlike the previously mentioned studies, sampled households are only followed for six years. Finally, consideration is given to a more conventional longitudinal study – the National Longitudinal Survey of Youth (NLSY) – in which specific cohorts of individuals, rather than households, are followed. The information in Table 1 relates to the 1979 cohort.²

There has also been some variation in mode of data collection, though most began with personal interviews using paper and pencil interviewing (PAPI) techniques and then eventually switched, in the 1990s, to computer assisted methods.³ Perhaps the most important mode distinction concerns whether interviews are conducted in person or by telephone. The NLSY and the European studies are all largely conducted in person with the assistance of a laptop computer (computer-assisted personal interviewing, or CAPI). The PSID, on the other hand, converted to telephone as long ago as 1973, switching to computer-assisted telephone interviewing (CATI) in 1993. Very differently, the SLID employed a CATI system from its very first wave.

The studies also vary in their experiences with respect to response and, though to a lesser degree, attrition. Wave 1 response rates vary from about 55 per cent in the case of the Dutch SEP up to a reported 87 per cent in the case of the 1979 cohort of the

¹ Though the rules by which new sample members are added and followed vary markedly across studies.

² Another cohort was commenced in 1997.

³ In the case of both the GSOEP and the BHPS, this change is still too recent for its impact to be assessed.

NLSY.⁴ With respect to the household-based panels, Wave 1 response rates appear to average somewhere around 70 per cent depending on how it is measured. In the BHPS, for example, interviews were completed with all household members at 69 per cent of cases. This compares quite favourably with the PSID, especially given that in the PSID an interview was only required from one family member. Indeed, while the Wave 1 response rate for PSID is typically reported as 76 per cent, this rate does not allow for the fact that the low-income sample was obtained from a larger population where consent to pass on names was required. One-quarter of these cases did not provide such permission. Including these non-consenting cases as non-contacts would reduce the response rate to just 69 per cent.⁵ Finally, while the GSOEP only reported a gross response rate in Wave 1 of 64 per cent, it needs to be noted that this study did not allow for proxy interviews, which the British experience suggests might have added up to four percentage points to the completed household response rate.

Attrition rates are less variable, with only the Dutch performing especially poorly. In most studies attrition is only of any significance in the second or third waves. Thereafter, attrition typically stabilises at quite low rates (under 4 per cent per year, and often closer to 2 per cent). Again the NLSY is the standout when it comes to achieving high completion rates.

Finally, the studies also vary markedly with respect to the approach taken to data collection. Both the SLID and the Dutch SEP are administered by national statistical agencies and hence obviously internalise all data collection functions. The PSID data are also collected by the host organisation, though in this case the host organisation – the Institute for Social Research – is based at a university. Further, the data collection group and the PSID management group are in an administrative sense quite independent of each other. The data collections for the GSOEP, the BHPS and the NLSY, on the other hand, are contracted out to quite separate entities and, in two out of the three cases these entities are private market research firms. The range of collection and data management functions that are contracted out is quite variable. In the GSOEP, for example, data editing and coding is largely left to the contractor. In the case of the BHPS and the NLSY, however, the host organisation retains considerable responsibility (total in the case of the NLSY) for these tasks.

Lessons from International Panel Study Experience: Sample and Survey Design

The Indefinite Life Household Panel Design

As outlined in Wooden and Watson (2000), the key features of the proposed survey design for HILDA were:

⁴ The reported initial response rate to the 1997 cohort (NLSY97) was even higher – 92 per cent.

⁵ A similar criticism can be made of how response rates are reported in the NLSY. The reported 87 per cent Wave 1 response rate in NLSY79 makes no adjustment for the fact that the sample was screened a year earlier and that successful screening interviews were not obtained from all selected units. Multiplying the successful interview rate by the reported response rate provides an effective response rate of around 81 per cent.

- (i) a household based sample;
- (ii) tracking members of the selected households over an indefinite life; and
- (iii) following rules that ensure the sample replenishes itself in the same way as the population.

This design is closely based on the design utilised in the three European studies and hence it is hardly surprising that it receives their endorsement.

The North American studies, on the other hand, are quite different in that none of them collect data from more than one household member. The PSID and the SLID, for example, both rely on a single household member to answer on behalf of other members of the household (and the NLSY is not a household-based survey). This, of course, means that data collected about individuals within the household will tend to be subject to more measurement error in the North American studies. Perhaps, more importantly, this approach renders these studies less receptive to more subjective questions. Thus, in contrast to the European studies, the survey instruments used in the North American studies tend to avoid questions about satisfaction and aspirations. Finally, collection of data from all household members permits more complicated analyses of family effects and intra-household dynamics. For these reasons, the approach advocated in the HILDA is preferred. Nevertheless, this approach has one major downside – it may be associated with higher rates of attrition and non-response. For example, Professor Randall Olsen, Director of the Centre for Human Resource Research at Ohio State University, which is responsible for managing the NLSY, argues that what matters most for response and attrition is not so much the length of a questionnaire, but the total time spent in the household. In his view, a long questionnaire administered to just one person will often be less intrusive than a shorter questionnaire administered to multiple members of the household.

There is also widespread support internationally for a design that provides for the tracking of individuals over long periods. Indeed, only the SLID employs a medium life design – household units are followed for a maximum of six years, with a new panel added every three years.⁶ While medium-life panels have their uses, the general consensus appears to be that in designing a survey that will provide a research vehicle for assessing a wide range of political and socio-economic developments into the future, most of which are not yet known, a classic design is best.

The third key feature of the survey design is the use of following rules to add new members to the sample in the same way as persons are added to households in the broader population. Use of such rules mean that in theory, the sample will be self-perpetuating, though this depends on rates of attrition, fertility rates, the degree of stability in household composition and success in recruiting these new sample members. The more important benefit of the application of following rules, however, is that it better facilitates analysis of both changes in household and family composition and inter-generational transmission mechanisms.

⁶ This decision by Statistics Canada to opt for a rotated medium-life panel design, rather than the classic indefinite life design, possibly reflected the fact that the SLID was to be used for multiple purposes. In particular, a number of existing cross-section supplements to the Canadian Labour Force Survey have been discontinued with the cross-section estimates from SLID being used in their place.

Reference Population

As noted above, the HILDA is closely modelled on the three European studies. In each of these the reference population for interview consists of ‘adult’ residents of private households. Further, the scope of the population is generally reduced to exclude persons living in institutions (such as hospitals, correctional institutions and boarding schools) and other non-private dwellings (such as hotels, motels and boarding houses).

For the most part, establishing population scope rules for wave 1 is fairly straightforward. More problematic is determining whether to follow all sample members as their status changes. For example, in the BHPS attempts are made to follow all sample members into institutions where it is practicable to do so. This thus typically means excluding prisoners. Very differently, in the PSID persons are not followed into institutions though their whereabouts are still tracked. If, however, a sample member is determined to have entered the institutionalised population on a permanent basis, then that person is removed from the sample entirely. Again we lean in favour of the approach adopted by the BHPS.

Some consideration will also need to be given to whether or not to follow persons who move overseas. The current PSID practice is to attempt to interview (by telephone) all persons who have not left the US permanently. This, however, leads to the question of how to establish whether a sample member is a temporary or permanent emigrant. Indeed, many of the sample members may not know the answer to this question.

Another issue concerns at what age to start interviewing children of adult sample members. Both the British and German studies begin interviewing persons from 16 years of age. In Australia, however, given the ABS definition of the labour force, a lower bound of 15 years seems more likely.

Interview Frequency

There is widespread agreement around the world that for a new survey, an annual frequency, as proposed for the HILDA, is best. More frequent survey periods (such as was employed in the Dutch SEP between 1984 and 1989) is generally not seen as feasible for most data collections given the time needed to collect the data. Less frequent data collection is possible, and indeed, both the PSID and the NLSY79 have moved to every other year collection. Nevertheless, in both cases the change occurred only after the sample had matured and stabilised. In general, annual frequency is seen as necessary for maintaining contact with sample members during its early years.

Sampling Frame

In most studies the first Wave begins only after a list of the names and addresses of the sampled households has been obtained, and in many cases this is based on a full enumeration of the relevant population.

A full enumeration offers a number of advantages. Most obviously, it enables pre-notification letters and materials to be sent to selected households prior to the interviewer arriving. Second, it provides the necessary data for ensuring that

interviewers do visit the dwellings they are expected to. Third, it provides the necessary data on from which to calculate selection probabilities for each household. Finally, pre-enumeration of the sample also permits the conduct of screening interviews, facilitating selective sampling of different population segments.

In light of these obvious benefits, a listing of the sample for the HILDA will now be obtained prior to interview.⁷ As is customary for surveys involving personal interviews, at least in countries with geographically dispersed populations, sample selection for the HILDA involves a multi-stage clustered design. That is, a sample of Census Collection Districts (CDs) is first selected (which consist of approximately 200-250 households). All private residential dwellings within each CD are then fully listed, and it is from this list that the sample will be selected.

Note, however, that it is not possible within the timeline and budget for the HILDA survey to conduct screening interviews (as used in the NLSY for example; see below). The sample listing thus cannot be used to help stratify the sample.

Over-sampling

One issue that almost all panel studies are forced to confront in designing their starting samples is whether to select an equal probability sample, or whether to vary sampling proportions with some important household characteristic. Most obvious here is income or some other measure of socio-economic disadvantage, with a number of international panel studies including low-income sub-samples. The most notable here is the PSID. This study was motivated by the “war on poverty” initiated in the US in the 1960s and hence a low-income sub-sample was crucial to the study objectives. However, it is important to realise that the PSID had originally been funded on the assumption that the *entire* sample would be comprised of economically disadvantaged groups and that it was never intended that the study would extend beyond its initial 5-year funding. The idea of including a general population sample came from Jim Morgan, the first Director of the PSID, and it is generally accepted that if he had not got his way, the PSID would indeed have ceased as planned and never have developed into the incredibly valuable research tool that it has become. It also needs to be borne in mind that the PSID had a ready-made sampling frame for drawing this population from – it had access to the names of participants in a US Census Bureau Survey on economic opportunities. As such, the probabilities of selection were relatively straightforward to calculate.

The NLSY cohorts also include over-samples for economically disadvantaged youth. In this case, however, sample selection involves a very expensive two-stage process. For example, in order to obtain the final civilian sample for the 1979 cohort (11,406 individuals), the National Opinion Research Center administered screening interviews in approximately 75,000 dwellings one year prior to Wave 1. These interviews were designed to elicit information that would allow the identification of persons eligible for inclusion in the sample and thus household income was necessarily a screening variable. It is also important to note that inclusion of an over-sample increases the cost of screening. The number of screening interviews for the equal probability civilian sample (6812 participants) was approximately 18,000. The supplemental sample designed to over-sample Hispanic, black and other economically

⁷ The initial proposal made no allowance for any pre-enumeration of the sample.

disadvantaged youth involved 57,000 interviews. As noted above, the funds provided for the HILDA render infeasible any first-round screening of the sample population. The proposed timeline also makes this impossible.

An obvious alternative to this approach which would not add to data collection costs would be to simply alter the composition of the sampling units (in our case, CDs) such that those sampling units with high proportions of low income households are over-represented. As Frank Stafford, the current Director of the PSID, observed, however, the problem here is that the correlation between household income and location may not be all that high. Consequently, the over-sampling proportions required in order to obtain a significant increase in the share of the sample accounted for by low income households would be extremely large (a view that has been strongly endorsed by our statistical consultant), with obvious adverse consequences for effective sample size.

This appears to be confirmed for Australia by 1996 Census data. In a random sample of Australian households approximately 28 per cent would be expected to have incomes below \$500 per week (in 1996 prices). By comparison, if we had restricted the sample to just those CDs which fell into the bottom three deciles on a socio-economic disadvantage scale (constructed by the ABS), the proportion of households expected to have incomes below \$500 per week would only have increased to 42 per cent. In other words, over-sampling CDs according to their degree of socio-economic advantage will not lead to a marked increase in the proportion of low-income households within the sample.

There is also the practical problem of how to draw a low-income sub-sample within budget that does not significantly affect the statistical efficiency of the sample.

Other problems that arise from over-sampling, especially on the basis of some non-fixed criteria (such as income), are listed below.

- (i) It is not obvious that household income is the appropriate selection criterion. In most cases we are interested in wealth rather than cash flow, and the latter is not necessarily well correlated with the former, especially when measured at a single point in time. This is especially true of the self-employed and young persons.
- (ii) Over-sampling low-income households necessarily means that other households are under-represented. This could compromise the value of the HILDA data in addressing other issues, such as transitions into low-income states from high-income states, and wealth and savings, one of the most important issues currently in both the UK and the USA, and one which other Australian users (such as the Reserve Bank and the Treasury) are almost certainly interested in.
- (iii) Related to the previous point, over-sampling low-income households is also likely to mean under-sampling workers (including low-wage workers, given it is well established that low-wage workers are spread across the income distribution).

Finally, it is worth again emphasising that the HILDA is a study of dynamics. Income states vary and hence in the long run the number of households which experience low-income states will be much greater than that in a single cross-section. That is, transitions into and out of low-income states will accumulate over time.

In light of the above, it was hardly surprising that almost all of the persons we discussed the issue with strongly advised against any such over-sampling, even though both the PSID and the NLSY did over-sample economically disadvantaged groups. Only Greg Duncan, the former long-time Director of the PSID thought over-sampling a low-income group had any merit, mainly because of his concern that low-income groups are more likely to attrit. But given our limited budget, he also suggested that we concentrate our resources on maximising survey completion rates rather than diverting resources to over-sampling.

Most panel study panel experts were of the view that over-sampling on income lines was not compatible with an indefinite life panel design. The main purpose of over-sampling low income groups is presumably to follow transitions out of low-income states. Since income is not fixed, this is best facilitated by a short to medium-life panel such as the ABS Survey of Employment and Unemployment Patterns, or even better still, by a medium-life panel with rotating samples, such as the Canadian SLID. With this design, new low-income sub-samples could be added with each cohort. This design maximises the benefits of over-sampling, and would provide the best framework for assessing the speed with which households move out of low-income states and the factors influencing that process.

The HILDA will, therefore, not include a selected sub-sample of low-income households.

Maintaining Sample Representativeness

Most international studies have, after some point, been compelled to address the question of whether or not their panels continue to be representative in a cross-sectional sense. With the exception of the Dutch, the international experience suggests that any loss of representativeness takes a long time to become a major problem. Fitzgerald, Gottschalk and Moffit (1998), for example, have demonstrated that 21 years on, and despite a loss of 50 per cent of the original sample, the PSID sample still retained its cross-sectional representativeness. The PSID, however, did add a new sample of Latino families in 1990. This sub-sample was subsequently dropped after 1995 and replaced with a small sample of post-1968 immigrant families and their adult children in 1997. The key point to note here is that it was over 20 years before concerns about representativeness were seen as serious enough for the PSID to have to augment its sample.

A new immigrant sub-sample was also added to the GSOEP in 1994 and 1995, but again the period between the first wave and the addition of this new sample was quite long.

The bottom-line is that it is likely to be quite a long time (and certainly not within the first 10 years) before new immigrant intakes are likely to be sufficient to warrant a new sub-sample being added.

Of course, refresher samples for the broader population may be warranted if rates of attrition and non-response (among new sample members) are sufficiently high. In the North American studies this has not been the case. In contrast, the Dutch are adding refresher samples every year. Indeed, one could seriously question whether the Dutch SEP is in fact an indefinite life panel. Arguably there are so many defects with the Dutch study that there are few lessons to be learned. Rates of non-response, however, are also relatively high in the GSOEP and have led to a new refresher sample of 1200 households being added in 1998. Note here that the problem with the German study is not caused by attrition – these are relatively good by international standards (about three per cent per annum). Rather the problem is one of obtaining cooperation at the first interview. As a consequence, members of new split-off households are under-represented.

Overall, it is not expected that a refresher sample will need to be added to the HILDA during the first three years of data collection.

Lessons from International Panel Study Experience: Instrument Design

Instrument Length

While there is a general consensus among the international household panel chief investigators that participants perceive long survey instruments as burdensome, there nevertheless is relatively little evidence from any of their panels to support the argument that length directly contributes to attrition or item non-response. Indeed, in the early years of the PSID this issue was examined and the correlation between interview length and attrition was found to be essentially zero. That said, it is important to recognise that interview length in the early years of the PSID was much shorter than it is today – 40 minutes compared with about 70 minutes. Further, just because attrition is not related to interview length does not mean initial response rates will not be. After all, persons responding in wave 1 share the common feature that they were not turned off by the prospect of having to spend time talking to an interviewer.⁸

For the HILDA, average interview times of 10 minutes for the household and 40 minutes for the individual have been established. These are on par with those reported as being typical in both the GSOEP and the BHPS.⁹ The perceived impact of the time imposition, however, will be monitored closely during the pre-testing phase.

⁸ It is often claimed that the relevant literature is inconclusive (see, for example, Bogen 1996). Taken as a whole, however, this literature is both of poor quality and dominated by the experience of mail surveys. In general, evidence from telephone and in-person surveys does support the hypothesised negative relationship between interview length and response rates, though the magnitude of the effect is arguably quite small (Frankel & Sharp 1981, Collins et al. 1988). It needs to be borne in mind, however, that these experiments were all conducted at a time when refusal rates tended to be relatively low.

⁹ In the GSOEP interview times for existing sample members average between 30 and 40 minutes for the individual questionnaire and between 15 and 20 minutes for the household questionnaire, giving a total interview time for a two-adult household of between 75 and 100 minutes. Interview times in the BHPS are almost certainly much longer, though the BHPS documentation suggests this is not so.

Timetable

A number of researchers emphasised the importance of time for ensuring a well designed and tested instrument or set of instruments. Instrument content in the NLSY, for example, is frozen a year before entering the field. This thus provides the time needed for ensuring the instrument will work as expected in the field. The timetable for the HILDA Survey, however, precludes adopting this approach.

Survey Content

While all the major household panel surveys have a common emphasis on work and income, they nevertheless differ enormously in their content. It is, therefore difficult to draw generalisations from the international experience about what topics should and should not be covered. Perhaps the main conclusion to be drawn is that beyond the survey core the content of these surveys does and will change significantly with the passage of time. This is hardly surprising – the content of these surveys evolves in line with what are the major economic and social issues of the day.

With respect to the actual content of the questionnaires, a large range of issues emerged in discussions. The most interesting of these, and which have informed development of the HILDA Survey instruments, follow.

- (i) Many data sets make the mistake of collecting high quality data on outcomes, but not on explanatory variables.
- (ii) Detailed data on many issues, such as health, education and wealth, do not need to be collected every year. Nevertheless, it is important to keep collecting data on standard items each year, even though many do not change. Thus, if it is identified that someone is employed in the same job as last year, details of that job should still be collected.
- (iii) Re-weighting for response bias in Wave 1 would be greatly assisted by collecting some data about all dwellings included in the sample (i.e., including those with which no contact is made).
- (iv) When collecting data on relationships, ensure data are collected on all relationships rather than just the relationship with the household head.
- (v) Data should be collected on both current labour market status and labour market activity over the course of the period between survey waves.
- (vi) Conventional monthly calendars, which are used to collect data on activity between waves, generate a “seam” bias – that is, a lot of spells appear to start at the beginning of the year and cease at the end. This potentially can be corrected for by using overlapping calendars.
- (vii) While calendar data for labour market activity is essential, the value of monthly income data is debatable. There may be greater value in simply collecting income-related data on an annual basis.
- (viii) There is a debate about the value of life history data and how it is collected. Both the BHPS and GSOEP attempt to collect spell-based work history data

that is very demanding and leads to recall problems. Greg Duncan favours summary type measures (such as were used in the SEUP).

- (ix) Questions on consumption and expenditure are unlikely to produce reliable data. No attempt should be made to construct a measure of total household expenditure.
- (x) Many studies admit to not having collected enough data about children and about the environment in which they grew up.
- (xi) Within Europe at least, there is a growing interest among researchers in the attitudinal data collected.

Finally, it needs to be recognised that there may be issues of importance to policy-makers in Australia that have not attracted much attention in overseas panels. FaCS, for example, has a strong interest in child care and in parent–child relationships, and these are issues that have not been central to most of the other panels. In contrast, they are likely to be the subject of a significant amount of interview time during the HILDA Survey.

Lessons from International Panel Study Experience: Data Collection

Mode of Collection

In general, international practice is that during the first few waves at least, panel surveys are conducted on a face-to-face basis.¹⁰ 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. Some studies then convert to telephone interviewing at a subsequent wave. The PSID, for example, switched to telephone interviewing at Wave 6.

The budget allocated for the HILDA Survey, however, only allows for conducting face-to-face interviews in the first wave. The most likely scenario, therefore, is that face-to-face interviews will be conducted in Wave 1 followed by computer-assisted telephone interview (CATI) in subsequent waves.

Telephone interviewing is problematic for the collection of certain types of data. Income data is most often mentioned, with both the amount and type of data that can be collected constrained by telephone methods. In the PSID, for example, income data is effectively only collected on an annual year basis – there has been no serious attempt made to generate an actual monthly calendar of income by source.

There is also an additional problem arising from the strategy being planned for HILDA. Specifically, switching modes so early in the life of the study may be problematic. The experience of the PSID with a switch in modes is that it is very taxing on the research team and unless carefully planned for, can create numerous problems with ramifications for the life of the panel. Such problems are best dealt with once the panel is established and once instruments have stabilised.

¹⁰ The Swiss Household Panel Study is a notable exception.

Self-administration

In most panel studies, trained interviewers control the survey situation. The GSOEP, however, is an interesting exception with the survey instruments designed so that self-administration is an option. The Germans argue that this can reduce interview time in multi-person households. Nevertheless, the negatives, especially in terms of item non-response and greater scope for misinterpretation of questions, would seem to outweigh the advantages.

Very differently, a separate short self-completion questionnaire is administered as part of the BHPS. Such an instrument can be used to ask questions that some respondents may not be prepared to answer out loud in a face-to-face interview. This might arise where the topic is considered sensitive or because of the presence of other household members at the time of the interview. It also might provide a good vehicle for collecting attitudinal data using fairly standard scaled responses (e.g., of the agree – disagree variety). Serious consideration is thus being given to the inclusion of such an instrument in the HILDA Survey.

Use of Computer Technology

Most of the large international household-based studies are now conducted using computer-assisted interview methods, though in the case of both the BHPS and the GSOEP the switch is very recent, and certainly too recent to assess its impact. The benefits of computer-assisted interviews are numerous and include: improved data quality through on-the-spot supervision and validation of the internal consistency of responses; reduced data entry and editing load; enabled delivery of complex questionnaire designs; and reduced data delivery times.

On the other hand, the experience of the PSID suggests that computer-assisted methods can generate considerable extra work at the data editing and validation stages. This needs to be budgeted for.

However, irrespective of the pros and cons, use of such methods for the HILDA will depend on their availability within the survey research market. In Australia such methods, when used in conjunction with telephone interviewing, are commonplace. In contrast, it is relatively rare for such methods to be used in conjunction with face-to-face interviewing. In large part, this simply reflects the difficulties generating a return on the initial capital investment in a market where large-scale surveys employing face-to-face methods are not widespread.

In summary, it is expected that during the first three waves of HILDA computer technology will only be used in conjunction with telephone interviewing.

Proxy Interviews

A key feature of both the PSID and the SLID is that one household member answers on behalf of other household members. In contrast, in the European studies, interviewers are expected to interview all household members. As observed earlier, it is proposed that the HILDA follow the European approach. This, of course, means that at some households interviews will only be completed with a sub-set of household members. This, in turn, gives rise to the question of whether or not to

permit proxy interviews – that is, one household member answering on behalf of another. In both the BHPS and the Dutch SEP this is permitted; in the GSOEP it is not. The main arguments against proxy interviews are twofold:

- (i) there are some questions, especially subjective questions, which one person cannot answer on behalf of another, at least not with any accuracy;
- (ii) permitting proxies provides an ‘easy’ option for interviewers given the incentives they face typically reward achieving high response rates.

The British, however, argue that some information is preferred to no information, and that incentive structures can be designed in such a way as to penalise the excessive use of proxies. In the BHPS, for example, less than four per cent of all individual interviews are conducted by proxy.

It is proposed that proxy interviews be permitted during the HILDA.

Lessons from International Panel Study Experience: Response and Attrition

Strategies for Maximising Response and Minimising Attrition

The keys to achieving high response rates and low rates of attrition in the international panel studies appear to be at least fourfold:

- (i) use of respondent incentives;
- (ii) a long field period;
- (iii) a committed and motivated interviewer workforce; and
- (iv) a fieldwork agency that is familiar with, and works to, academic standards.

Most studies use some form of incentive to encourage response. The US studies typically use cash. The PSID, for example, pays each respondent US\$50 plus a further US\$10 for returning address confirmation between waves. The NLSY also uses cash incentives. Unlike the PSID, however, the amount of the incentive has (at least until recently) varied depending on the difficulty of obtaining a response.¹¹ NLSY interviewers also have the scope to use non-cash incentives (in addition to cash incentives) where it would help. This might, for example, include purchasing a Disney video to keep the children occupied or buying pizzas (which appeals to younger sample members).

A cheaper alternative is to provide some type of gift. For example, every respondent to the BHPS is rewarded with a £7 gift voucher redeemable at a national pharmacy chain. Respondents who return completed change of address cards also receive a voucher. Very differently, participants in the GSOEP each receive a lottery ticket.

Long field periods also facilitate higher completion rates. Obviously the longer interviewers spend in the field chasing hard to get cases, the more likely it is that they will succeed in getting those cases. In most studies the fieldwork period is at least 6

¹¹ In 2000 the US Congress limited all cash incentive payments by the NLSY to just US\$20 each. In the opinion of Professor Randall Olsen, this decision has directly contributed to greater difficulty in obtaining responses and is reflected in an expected higher than average attrition rate and a much longer field period.

months, though longer periods – up to 9 months – are not uncommon. Indeed, in the 2000 NLSY interviewing commenced in February and is not expected to be complete until just prior to Christmas. A more typical schedule is that of the BHPS, with the main interview period running from September to mid-December, followed by a tail, involving far fewer interviews, which runs over the following February to April.

Most investigators also stress the need to have an interviewer workforce that is “sold” on the survey. Unless interviewers are genuinely convinced that the survey is worthwhile, then it is very unlikely that the interviewers will be able to encourage participation in the survey to sample members.

Other strategies identified as influencing response and/or attrition rates include: distribution of marketing material in advance of each survey round; allocating considerable effort to converting refusers;¹² distributing feedback to sample members; assigning the best interviewers to the most difficult cases; and providing financial incentives to the fieldwork agency (though there is some disagreement among the studies about the value of the latter).

Declining Response Rates

There is a recent but growing body 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 attempted to add 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.¹³

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.

Such findings suggest that even give the adoption of best practice strategies for dealing with response, it may be unrealistic to expect that the first wave of the HILDA Survey will obtain the same level of response rates achieved in some of the long-standing panel surveys.

¹² Customised persuasion letters are prepared for difficult cases in the PSID.

¹³ There was one other significant change in methods that may have impacted on response. In the brochure pre-mailed to respondents, an explicit indication of interview length was provided.

Lessons from International Panel Study Experience: Data Processing

Panel data collection places much greater demands on database management and data processing than cross-section surveys. For example, all panel study investigators emphasised the importance of maintaining a database for simply tracking sample members that was linked to the data collected.

The importance of clean data within a panel context was also emphasised strongly by a number of the panel study investigators. In comparison with a cross-section data set, measurement errors are much more problematic in panel data. Essentially the reason here is that noise has a much bigger impact on the precision of statistical estimates in models of change than it does on estimates of levels from cross-section data. The implication, therefore, is that the amount of resources needed for data cleaning, editing and validation cannot be under-estimated. A major issue to be determined here is whether editing is determined by algorithm (as in the NLSY) or whether algorithmic checks are complemented by manual checks. The latter, which is a central feature of the procedures used in the PSID, is obviously much more labour intensive.

Consideration also needs to be given to how much imputation of data will be undertaken. On some questions, such as those related to income, there may be considerable item non-response, leading to the question of whether responses can be imputed. In a cross-section collection, imputation is probably best avoided, but in a longitudinal context imputation is likely to be far more reliable.

The preparation of a user friendly data set, especially one that provides a large array of derived variables, is also very time intensive with most studies taking between 9 and 12 months to deliver public data sets.

Finally, there were clear differences among studies with respect to the extent that these tasks were the responsibility of the fieldwork organisation. The BHPS team, for example, assumed prime responsibility for sample management and data editing, and were heavily involved in interviewer training and briefings, fieldwork monitoring and in refusal conversions. In contrast, the GSOEP team left these tasks almost exclusively to the fieldwork organisation, arguing that this encouraged a more trusting relationship. There are clearly pros and cons associated with both approaches and respect to steer a middle course in the HILDA. Ultimately, however, the approach taken is likely to depend on the relative skills and expertise of the fieldwork organisation and the staff hired on to the HILDA team.

Lessons from International Panel Study Experience: Data Dissemination

There are a wide variety of approaches to data dissemination. At one extreme is the PSID where the main data set is available from a web site, with no restrictions on who can download the data. At the other extreme is SLID, where access to the unit record data can only be obtained through special agreement with Statistics Canada, and typically involves the researcher working with the data on the premises of Statistics Canada (or in official data laboratories).

The obvious and serious limitation with the latter strategy is that user access to the data is heavily constrained. In particular, if users can only obtain access within a small number of officially sanctioned secure sites, then the likelihood is that very little use will actually be made of the data. The experience with the SLID in Canada confirms this, with limited use being made of the unit record data outside of Statistics Canada. This is potentially of large importance for a survey like HILDA, with the enthusiastic support of as large a user community as possible likely to be important in ensuring continuing commitment from funding bodies.

In contrast, the PSID claim that there are some 30,000 users of PSID data, though this is certainly a very crude estimate. Less uncertain, the PSID report delivering around 6500 customised data sets to researchers every year. The PSID have thus been extremely successful in ensuring that their data are used. Nevertheless, some groups have expressed concern about the PSID strategy, arguing that the potential for misuse of the data means that public releases of the data should be restricted to users from a research background (be it in academia, government or private research institutions). This typically means the data will need to be encrypted with registered users being provided with individual passwords to decode the data. Further, such users will be required to sign confidentiality agreements designed to prevent them from passing the data on to others or to use the data in such a way that undermines the confidentiality and privacy of the sample members.

It is also clear that even those with the most liberal data dissemination strategies do not make all the data available to all the users. Thus it is common to find both detailed location and occupation data to be suppressed, though some groups are prepared to provide more detailed data files (typically at a cost) under very restricted conditions. Nevertheless, in general, the amount of data that is provided is still extremely generous compared with the types of strategies adopted by national statistical agencies, such as the ABS, Statistics Canada and Statistics Netherlands. In these types of organisations concerns about confidentiality invariably lead to the implementation of a range of statistical techniques to data to ensure no individual can be identified. This generally includes both altering the data (randomising responses, recoding outliers, top coding data items, creating artificial cases), removing data (e.g., by dropping cases or by dropping items), or reducing the level of detail supplied (e.g., by aggregating responses to some data items). Many of these strategies, however, are simply not possible with longitudinal data, and hence may explain why Statistics Canada has opted, in the case of SLID, for a policy of restricting access. Nevertheless, the question naturally arises: why collect the data if use of the data is to be discouraged?¹⁴ It is also not irrelevant that despite its long history and the ease of availability of its data, the PSID has still yet to report any serious misuse of its data resulting in the breach of respondent confidentiality.

It is also important that the data is accompanied by high quality documentation. Without such documentation there is again likely to be fewer users. User interest in

¹⁴ Interestingly, the approach taken by ACER and the Department of Education and Youth Affairs to the dissemination of data from the various waves of the longitudinal youth surveys under the LSAY program is extremely liberal. They adopt a similar approach to the GSOEP in terms of who they allow to have the data, but beyond that there are very few safeguards against breaches of confidence and security. For example, the only variables not provided with the main public unit record release are: name, date of birth, address / location, school, and achievement scores. All other variables are provided in full detail (e.g., occupation is provided at the 4-digit level).

the data can also be promoted through preparation of a wide array of derived variables.

A dissemination strategy for HILDA is currently being developed.

Lessons from International Panel Study Experience: Data Matching and Linking

With the exception of SLID, most international panel studies assiduously avoid matching their data. The US studies, for example, only link their data to Census geo-code data and deaths records, and in both cases, access to these linked data files is highly restricted. No attempt has been made to match the data to the administrative records held by government agencies. Such matching is seen as problematic for two key reasons.

First, a prerequisite for data matching is obtaining the informed consent of sample members, which in most instances will cause the studies to be linked, in the eyes of the participants to governments. In the US this is believed to have a very damaging effect on both cooperation rates and data quality.

Second, linked data of this type is typically too sensitive to be distributed to the wider community of researchers, partly explaining why Statistics Canada, which asks participants in the SLID for the equivalent of their tax file number, does not make their unit record data widely available. It is important to note here that many of the techniques used to protect respondent anonymity in cross-section data, such as randomisation of responses and the creation of artificial cases, are not viable options when dealing with panel data.

In summary, the international experts generally believed that given these concerns and given the lack of experience with panel data collections in Australia and the risky nature of embarking on a new panel, data linking which requires informed consent of participants should, at a minimum, be avoided during the first few waves of the HILDA.

Other Issues

There were a myriad of other issues that arose during the consultations with panel study investigators; indeed, too many to do justice to them in this paper. Some of the more interesting and most relevant are briefly summarised below.

The Pilot Test Sample

There may be value in maintaining the sample pilot for the first few waves, thus enabling some of the longitudinal aspects of survey design to be tested and evaluated. Pre-tests that are conducted each wave using new samples are unable to tell us much about how existing sample members will react to new questions or changes in questions. Maintaining a panel sample for the pilot may also assist in identifying influences on attrition.

It is thus our intention to revisit at least part of the sample used in the main pilot test (that is, the full dress rehearsal) during the pre-testing for later waves.

Between Wave Contacts

Effort needs to be devoted to maintaining contact with the sample between waves (e.g., through provision of feedback material, and circulation of materials providing advance notification of next wave). However, with the exception of obtaining details on any changes in address, between wave contact should not impose any more burden on respondents.

The Interview Situation

Interviewers need to be trained on how to control the interview situation. This is especially important for the individual interview where the presence of other household members should be avoided. This is to be built in to the two-day training course that all interviewers are required to attend.

It is also important to collect details about the interview situation, including for example whether other persons were present and influenced responses, degree of cooperation, and whether problems encountered in administering the interview.

Item Coding

While more expensive, occupation and industry should be coded at highly disaggregated levels. This facilitates dealing with changes in official classifications systems, which occur on an all too regular basis.

Identifiers

Absolutely critical for successful management of the data is the establishment of an appropriate system of identifiers. There obviously needs to be a unique individual identifier. Further the system needs to be able to quickly link related individuals to each other.

Pre-loaded Data

When using computer-assisted survey methods, it becomes possible to include pre-loaded data from earlier survey waves in the CAI programs. In practice, however, this is not really feasible. With annual interviews, there simply is not sufficient time to process the data to a state ready to load into CATI (or CAPI) programs. About all that can be done is the pre-loading of household composition (and this is essential).

Weights

Considerable resources are typically devoted to the construction of appropriate sets of weights.

At Wave 1 initial sample selection probabilities are typically adjusted according to established population benchmarks, though it is widely agreed that more could be done to correct for non-response bias.

Weights for subsequent waves need to be corrected for attrition. This typically requires estimating regression models of the probability of sample members staying in the sample each year.

Finally, weights need to be assigned to new joiners. In the BHPS this is done by sharing the existing 'household' weight with new joiners. An alternative method involves actually estimating the probability of new persons joining the sample, as is done in the GSOEP.

These issues are discussed at greater length in Henstridge (2001). A strategy for dealing with weights will be developed during 2001 and will lead to a Technical Paper to be released early in 2002.

Software

Assuming some use is made of computer assisted methods (e.g., when using telephone), a key issue will be the choice of software. There appears to be widespread dissatisfaction with the most popular products such as CASES or SurveyCraft. Of the commercially available packages, the preferred choice now appears to be Blaise, the product developed by Statistics Netherlands.

Also important is the choice of software for database management. Again there appears to be widespread dissatisfaction with the most well known software – certainly none of the studies use SPSS, which is not regarded as a useful tool for managing longitudinal data. Indeed, both the GSOEP and the BHPS use the little known packaged, Statistical Information Retrieval, or SIR, while the NLSY use a program developed in-house, based on an Oracle engine, which is integrated into the CAPI system.

The choice of appropriate database software is currently under consideration.

Brand Names

Many studies market their surveys to sample members using a different brand name than the official study name. Thus the GSOEP is marketed as "Living in Germany" while the BHPS is marketed as "Living in Britain". It is believed that these brands are easier for respondents to identify with. Of course, it also means that the survey must cover the major influences on life satisfaction.

The HILDA is thus to be branded as "Living in Australia". This brand name was market tested and received overwhelming endorsement from those who participated in that test. A logo has been developed to assist this brand name identification and will be used on materials and products sent to sample members (such as brochures, pens, fridge magnets and the like). A Living in Australia web site has also been established (www.livinginaustralia.org).

Conclusion

In the main, the design proposed for the HILDA appears to reflect what is regarded as international best practice. International experience, however, suggests there at least two major weaknesses in the initial design. First, the proposed switch to telephone

interviewing in Wave 2 is too soon, potentially compromising both response and data quality. Second, no provision was made for providing incentives to sample members for participating. Such concerns have led FaCS to seek additional funding to redress these concerns. At the time of writing, for example, additional funds had been secured to make the payment of at least a modest incentive to respondents in wave 1 possible.

References

Barreiros, L. (1995), 'the European Community Household Panel: Its Design, Scientific and Policy Purposes', *Innovation: The European Journal of Social Sciences*, vol. 8, pp. 41-52.

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.

Bogen, K. (1996), 'The Effect of Questionnaire Length on Response Rates – A Review of the Literature', Paper presented at the annual conference of the American Association for Public Opinion Research, Salt Lake City.

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.

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.

Frankel, J. and Sharp, L. M. (1981), 'Measurement of Respondent Burden', *Statistical Reporter*, vol. 81, no. 4, pp. 105-111.

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.

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. and Watson, N. (2000), 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: An Introduction to the Proposed Survey Design and Plan', HILDA Project Discussion Paper Series No. 1/00, Melbourne Institute of Applied Economic and Social Research, University of Melbourne, December.

Other Source Materials

Bureau of Labor Statistics, US Department of Labour (2000), *NLS Handbook, 2000*, US Department of Labor, Washington DC.

Centre for Human Resource Research (1999), *NLSY79 User's Guide: A Guide to the 1979-1988 National Longitudinal Survey of Youth Data* (Report prepared for the US Department of Labor), CHRR, Ohio State University, Columbus.

Haisken-De New, J.P. and Frick, J.R. (eds) (2000), *Desktop Companion to the German Socio-Economic Panel Study (GSOEP)*, Version 3.0, German Institute for Economic Research Berlin.

[Available at: <http://www.diw.de/english/sop/service/dtc/>].

Institute for Social and Economic Research, *BHPS User Documentation*.

[Available at: <http://www.irc.essex.ac.uk/bhps/doc/index.htm>].

Panel Study of Income Dynamics webpage (at <http://www.isr.umich.edu/src/psid/>).

Statistics Canada (1998), *Survey of Labour and Income Dynamics: An Overview*, Reg. Product # 75F0011XIE, June.

[Available at: <http://www.statcan.ca:80/english/freepub/75F0011XIE/eng.pdf/>].

Statistics Canada, *SLID Electronic Data Dictionary*, Reg. Product # 75F0026XIB.

[Available at: <http://www.statcan.ca/english/SLID/diction.htm/>].