

Sample Attrition in the HILDA Survey

*Nicole Watson and Mark Wooden**

The Melbourne Institute of Applied Economic and Social Research,
The University of Melbourne

Abstract

This paper provides an assessment of the nature of attrition in the HILDA Survey sample between wave 1 and wave 2. More specifically, the paper presents summary statistics on the levels of response achieved in wave 2, reports on observable differences between wave 1 respondents who did not respond in wave 2 and those that did, and examines briefly the impact of attrition on population estimates.

1. Introduction

It is widely recognised that surveys that employ longitudinal, or panel, designs have many advantages over simpler cross-sectional designs (see, for example, Rose, 2000 and Duncan, 2000). Nevertheless, compared with cross-sectional surveys, panel surveys are still relatively uncommon.¹ One obvious reason for this is the large financial commitment that a long-running panel survey entails. The other major detracting feature of panel surveys is that they are almost always affected by relatively high levels of attrition over time. The [US] Panel Study of Income Dynamics (PSID), for example, had lost 25 per cent of its original sample after seven years and 51 per cent after 21 years (Fitzgerald, Gottschalk and Moffitt, 1998). Rates of attrition have been even higher in more recent studies. The British Household Panel Survey (BHPS), for example, which started in 1991, had lost 33 per cent of its original sample after seven years.²

As should be obvious from the foregoing, attrition occurs when a person interviewed at one wave ceases to participate in the survey at a later wave. There are many reasons for attrition, but essentially these can be classified into two types. First, there is attrition that results because of changes in the underlying population, such as death and moves overseas. Second, there is attrition that arises because sample members cannot be located or because they refuse to continue participation. The first type of attrition is unavoidable and from a statistical perspective at least, gives rise to fewer problems for analysis of data. Potentially far more problematic is the second form of attrition, and it is this type of attrition that is the focus of this paper.

Address for correspondence: Nicole Watson, Melbourne Institute of Applied Economic and Social Research, Faculty of Economics and Commerce, University of Melbourne, Vic 3010. Tel: 61 3 8344 2088. Email: n.watson@unimelb.edu.au.

*The HILDA Project was initiated, and is funded, by the Australian Government 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 For a brief review of major household panel studies that have been established around the world, see Haisken-DeNew (2001). For some discussion of the Australian experience with panel surveys, see Wooden and Watson (2002).

² Derived from Buck, *et al.*, 2002, table 52. Proxy interviews have been excluded from this calculation.

© The Centre for Labour Market Research, 2004

High rates of attrition due to non-response per se, however, are not necessarily a major problem. Obviously attrition can lead to declining sample sizes and thus gradually reduce the efficiency of panel data estimates. This problem, however, is largely non-existent in panel studies that employ an infinite life design where the sample is constantly being augmented by new sample members (through the use of predetermined following rules). Far more problematic is attrition that is non-random. If the persons and households that attrit from the panel have characteristics that are systematically different from those who remain, then analyses of these data that do not account in some way for the selective nature of the attrition will most likely lead to biased inferences.³

The main objective of this paper is to assess the nature of attrition due to non-response in the Household, Income and Labour Dynamics in Australia (HILDA) Survey between wave 1 and wave 2. We begin, in section 2, by presenting summary statistics about the levels of response achieved in wave 2. In section 3 we report on observable differences between wave 1 respondents who did not respond in wave 2 and those that did. The key feature of this section is the presentation of results from the estimation of logit models of the probability of response. We then, in section 4, briefly examine the effect of attrition on various population estimates. In the concluding section some brief observations are made about the options researchers have for correcting for the potential bias arising from attrition.

2. Response and Attrition in the HILDA Survey, Wave 2

The HILDA Survey began in 2001 and, as detailed in Watson and Wooden (2002), commenced with a responding sample of 7682 households. In wave 2, conducted one year later, the objective was to obtain interviews at all of these households and with all adult members (that is, persons aged 15 years or older on the 30th June preceding interview) of those households.⁴ Household composition, however, is not stable, and will often change. In recognition of this, and in line with the following rules adopted in other major household panel surveys overseas, such as the BHPS, interviews are also sought with all other adults living with a wave 1 sample member. These new sample entrants, however, are distinct from the permanent sample members from wave 1 in that they only remain in the sample for as long as they reside with a permanent sample member.⁵

As reported in table 1, of the 7682 responding households from wave 1, 68 were known to have moved out of scope (due to death or moves overseas). To these, an additional 712 households were added to the sample as a result

³ For a formal statistical model of attrition bias, see, Fitzgerald, Gottschalk and Moffitt, 1998.

⁴ The fieldwork for wave 1 was conducted between August 2001 and January 2002 and the fieldwork for wave 2 was conducted approximately one year later, between August 2002 and March 2003.

⁵ There are a couple of notable exceptions to this rule. First, any child born to a permanent sample member is also treated as a permanent sample member, though obviously it will be another 15 years before they are eligible for interview. Second, any new entrant who has a child with a permanent sample member is converted to permanent sample member status.

of changes in household composition. There were, therefore, a total of 8326 households in the wave 2 sample, and interviews were obtained at 7245 of these, resulting in a household response rate of 87.0 per cent. As in wave 1, we were unable to obtain interviews with all household members at all households, with 10 per cent of responding households being incomplete. Table 1 also shows that of the 1081 non-responding households, 58 per cent actively declined to participate, 12 per cent were contacted but no interview was successfully conducted during the available fieldwork period, and 30 per cent were not contacted (mostly because they could not be located).

In table 2 we report a summary of person-level response in wave 2 cross-classified by the respondents' status in wave 1. A total of 13,041 persons responded in wave 2, 92 per cent of whom were interviewed in wave 1, just under two per cent were non-respondents in wave 1 but were eligible for interview, two per cent were children aged less than 15 years in wave 1, and over four per cent were new sample members.

Table 1 Wave 2 Household Response Summary

| <i>Sample Outcome</i> | <i>Number</i> | <i>Per cent</i> |
|---|---------------|-----------------|
| Households issued | 7682 | |
| plus split households | 712 | |
| less out-of-scope households (deaths or moves overseas) | 68 | |
| = Total in-scope households | 8326 | 100.0 |
| Refusals | 622 | 7.5 |
| Non-response with contact | 134 | 1.6 |
| Non-contact, not lost in tracking | 75 | 0.9 |
| Lost in tracking | 250 | 3.0 |
| Fully responding | 6540 | 78.5 |
| Partially responding | 705 | 8.5 |

Table 2 Wave 2 Person Response Categories Against Wave 1 Response Categories

| | | <i>Wave 2</i> | | | | | |
|---------------|-----------------------|-------------------|-----------------------|-------------------------------------|---------------------------------|--------------|--------------|
| | | <i>Respondent</i> | <i>Non-Respondent</i> | <i>Lost to Tracking^a</i> | <i>Out of Scope^b</i> | <i>Child</i> | <i>Total</i> |
| <i>Wave 1</i> | <i>Respondent</i> | 11993 | 1591 | 233 | 152 | - | 13969 |
| | <i>Non-Respondent</i> | 222 | 812 | 92 | 32 | - | 1158 |
| | <i>Child</i> | 250 | 58 | 133 ^c | 19 ^d | 4327 | 4787 |
| | <i>New Entrant</i> | 576 | 209 | - | - | 346 | 1131 |
| | <i>Total</i> | 13041 | 2670 | 458 | 203 | 4673 | 21045 |

- Notes: a People that are 'lost to tracking' are those that have presumably moved from their last known address and their new address cannot be identified via the range of contact details provided in wave 1 or other methods used to locate the individual.
- b People who are deceased or who have moved overseas are considered 'out of scope'.
- c Of the 133 wave 1 children that were lost to tracking, 3 had turned 15 for wave 2 (thus eligible for interview) and the remaining 130 were still children.
- d Of the 19 wave 1 children that have moved out of scope, 3 had turned 15 for wave 2 and the remaining 16 were still children.

The unadjusted attrition rate, which includes natural attrition (i.e., deaths and movers overseas), is calculated as the proportion of wave 1 respondents that did not respond in wave 2. This is 14.1 per cent (1976 / 13969). For this paper, the key statistic is the attrition rate after adjusting for movements out of scope – 13.2 per cent. While high, this rate is not much higher than the rate achieved by the BHPS in its second wave – 12.4 per cent (after adjusting for deaths and movements out of scope). Indeed, given the BHPS was commenced a decade earlier, and given evidence of a decline in survey response rates during the 1990s, a greater differential might have been expected.⁶

3. Observable Correlates of Attrition

In this section we consider the question of whether or not the persons who did not respond in wave 2 are systematically different from those who did. We begin by comparing the distribution of the responding wave 2 sample with the non-responding sample on selected characteristics. We then report results from the estimation of binary choice models for the probability of response in wave 2.

Distributions of Wave 1 Characteristics

Tables 3 and 4 show, for selected sample characteristics, measured at wave 1, both the composition of the sample eligible for interview in wave 2 (all wave 1 respondents less deaths and movers overseas) and the attrition rate. Further, we also report the result of a simple non-parametric test for significance of difference in attrition for each characteristic.

Table 3 presents results for key demographic variables. Specifically, for each characteristic we report the distribution of the population eligible for interview at wave 2 and the attrition rate. Thus if we consider the area in which people reside (based on the address recorded at the wave 1 interview), we can see that attrition rates vary from just 4.6 per cent in the Northern Territory (though the sample size involved here is very small) up to 17.6 per cent for residents of Sydney. Further, since Sydney residents represent a sizeable proportion of the sample (17.2 per cent), it suggests that this relatively high rate of attrition cannot be easily ignored.

In general, the most striking feature of table 3 is how different attritors and non-attritors are on most characteristics, as reflected in the reported P-values. Compared with non-attritors, attritors are more likely to have been, at wave 1: living in Sydney, young, single or alternatively living in a de facto relationship, born overseas but not in one of the main English-speaking countries, an Aboriginal or Torres Strait Islander, without a post-school qualification and/or living in a flat, unit or apartment. Of the variables considered, only sex and an index of socio-economic disadvantage were not significantly related to attrition.

⁶ The wave 2 attrition rate of 13.2 per cent obtained in the HILDA Survey compares favourably with the wave 2 attrition rates of 15.0 and 12.2 per cent reported for the Welsh and Scottish top-up samples, which were introduced into the BHPS in 1999.

Table 3 Wave 2 Attrition Rates by Selected Wave 1 Demographic Characteristics

| Wave 1 Characteristic | Wave 2 Eligible (%) | Attrition Rate (%) | P-value | Wave 1 Characteristic | Wave 2 Eligible (%) | Attrition Rate (%) | P-value |
|--------------------------|---------------------------|--------------------------|---------|--|---------------------------|--------------------------|---------|
| Area | | | <0.001 | No. of adults in h'hold ^a | | | <0.001 |
| Sydney | 17.2 | 17.6 | | One adult | 16.3 | 8.7 | |
| Rest of NSW | 13.7 | 11.8 | | Two adults | 54.5 | 11.3 | |
| Melbourne | 18.0 | 14.6 | | Three adults | 16.2 | 17.2 | |
| Rest of Victoria | 7.7 | 11.9 | | Four or more adults | 13.0 | 21.8 | |
| Brisbane | 8.4 | 12.5 | | No. of children in h'hold ^a | | | <0.001 |
| Rest of Qld | 10.7 | 10.8 | | No children | 64.2 | 12.6 | |
| Adelaide | 6.3 | 10.8 | | One child | 14.4 | 16.7 | |
| Rest of SA | 3.1 | 11.3 | | Two children | 13.7 | 13.0 | |
| Perth | 7.2 | 11.4 | | Three or more | 7.7 | 11.9 | |
| Rest of WA | 2.7 | 13.8 | | Country of birth | | | <0.001 |
| Tasmania | 3.0 | 14.2 | | Australia | 74.7 | 11.9 | |
| Northern Territory | 0.5 | 4.6 | | O/S: Main English-spkg | 10.8 | 12.9 | |
| ACT | 1.7 | 9.1 | | O/S: Other | 14.4 | 20.1 | |
| Sex | | | 0.117 | Indigenous status | | | 0.002 |
| Male | 47.5 | 13.7 | | Indigenous | 1.9 | 19.8 | |
| Female | 52.5 | 12.8 | | Non-indigenous | 98.1 | 13.1 | |
| Age group (years) | | | <0.001 | Education attainment | | | <0.001 |
| 15-19 | 8.9 | 20.3 | | Year 11 or below | 36.6 | 15.1 | |
| 20-24 | 7.1 | 23.4 | | Year 12 | 11.8 | 16.6 | |
| 25-34 | 18.7 | 14.2 | | Certificate | 26.1 | 13.1 | |
| 35-44 | 21.6 | 11.2 | | Diploma | 8.1 | 10.5 | |
| 45-54 | 17.4 | 11.5 | | Degree or higher | 17.3 | 8.3 | |
| 55-64 | 12.0 | 10.4 | | Dwelling type | | | <0.001 |
| 65-74 | 8.6 | 8.0 | | House | 82.1 | 12.8 | |
| 75+ | 5.6 | 12.8 | | Semi-detached | 8.1 | 12.8 | |
| Marital status | | | <0.001 | Flat, unit, apartment | 9.0 | 17.7 | |
| Married | 54.1 | 11.4 | | Other | 0.8 | 10.6 | |
| De facto | 9.7 | 16.5 | | Index of disadvantage ^b | | | 0.082 |
| Separated | 3.1 | 11.8 | | Lowest quintile | 18.8 | 12.7 | |
| Divorced | 5.4 | 10.8 | | 2nd lowest quintile | 20.0 | 14.6 | |
| Widowed | 4.8 | 7.9 | | Middle quintile | 20.1 | 13.2 | |
| Single | 22.9 | 17.9 | | 2nd highest quintile | 20.9 | 13.3 | |
| Relationship in h'hold | | | <0.001 | Highest quintile | 20.2 | 12.1 | |
| Married couple | 53.6 | 11.4 | | | | | |
| De facto couple | 9.5 | 16.7 | | | | | |
| Lone parent | 5.6 | 14.4 | | | | | |
| Child/relative | 14.4 | 20.2 | | | | | |
| Lone person | 13.9 | 8.0 | | | | | |
| Unrelated | 3.0 | 23.8 | | | | | |

Notes: a An adult is defined here as anyone aged 15 years or over. A child is therefore anyone under the age of 15 years.
 b The index of disadvantage used here is the index of relative socio-economic disadvantage constructed by the ABS (ABS cat. no. 2039.0) and is based on the place where people live.

Table 4 is similar to table 3, but all the characteristics listed here are better thought of as outcome variables. Again, attritors and non-attritors varied significantly on most characteristics. Attritors were more likely to have been unemployed when interviewed at wave 1, were more likely to have reported relatively low levels of life satisfaction, were less likely to own or be buying their own home and, as we would expect, were much more likely to have changed address between waves. There are also significant differences by income, with attritors more likely to have relatively low levels of personal income, but in general were not from low-income households. This latter finding no doubt reflects, at least in part, the difficulty securing interviews

with teenagers and adult children in many households. The only outcome variable where there was no marked difference between attritors and non-attritors was benefit status. Indeed, the rate of attrition among recipients of government benefits and pensions was identical to that of non-recipients. That said, there are almost certainly differences depending on the type of benefit claimed. Most obviously, the higher rate of attrition among the unemployed points to relatively high rates of attrition among unemployment benefit recipients.

Table 4 Wave 2 Attrition Rates by Selected Wave 1 Outcome Variables

| Wave 1 Characteristic | Wave 2 Eligible (%) | Attrition Rate (%) | P-value | Wave 1 Characteristic | Wave 2 Eligible (%) | Attrition Rate (%) | P-value |
|--------------------------|---------------------------|--------------------------|---------|--------------------------|---------------------------|--------------------------|---------|
| Labour force status | | | <0.001 | Annual household income | | | <0.001 |
| Employed full-time | 41.6 | 13.3 | | <\$20,000 | 16.8 | 11.0 | |
| Employed part-time | 19.6 | 12.6 | | \$20,000 - \$39,999 | 19.9 | 13.2 | |
| Unemployed | 4.4 | 18.9 | | \$40,000 - \$59,999 | 19.4 | 14.5 | |
| Not in labour force | 34.5 | 12.7 | | \$60,000 - \$79,999 | 16.0 | 13.8 | |
| Life satisfaction | | | 0.001 | \$80,000 - \$119,999 | 17.6 | 11.8 | |
| Low (0-4) | 3.3 | 17.8 | | \$120,000+ | 10.3 | 15.9 | |
| Medium (5-7) | 28.4 | 14.0 | | Annual personal income | | | <0.001 |
| High (8-10) | 68.3 | 12.6 | | Negative / nil | 4.6 | 15.8 | |
| Household tenure | | | <0.001 | \$1-\$9,999 | 24.4 | 15.1 | |
| Own / purchasing | 72.4 | 11.8 | | \$10,000 - \$14,999 | 12.7 | 11.0 | |
| Rent | 25.2 | 17.9 | | \$15,000 - \$29,999 | 21.4 | 14.3 | |
| Other | 2.5 | 7.8 | | \$30,000 - \$59,999 | 26.4 | 12.4 | |
| Benefit recipient status | | | 0.907 | \$60,000+ | 10.4 | 10.0 | |
| Benefit recipient | 33.1 | 13.2 | | Moved b/w wave 1 and 2 | | | <0.001 |
| Not benefit recipient | 66.9 | 13.2 | | Moved | 17.6 | 21.4 | |
| | | | | Did not move | 82.4 | 11.4 | |

Note: The income variables are for the financial year preceding the wave 1 interview (2000-01) and include imputed values for missing cases.

Attrition Logits

We now turn to a consideration of the determinants of attrition within a multivariate framework. Specifically, we estimated logit equations for the probability of response at wave 2.⁷ We also estimated a two-part model which distinguished between the two major stages of response – making contact and obtaining an interview.

The results from the single-equation model predicting response are provided in table 5. As can be seen, the coefficients and standard errors from three different specifications are reported. The coefficients, or log odds, indicate the relative effect of different characteristics on the likelihood of an individual responding in wave 2. Specification 1 only includes basic personal and demographic characteristics. The list of characteristics is the same as those considered in table 3 and all are specified in binary form, except age and the number of adults and children in the household, which

⁷ Alternatively, we could have used a probit model. We opted for the logit model on the grounds that a logistic distribution seems more appropriate when the dependent variable contains a marked imbalance between responses (Y 's equal to one) and non-responses (Y 's equal to zero), as is the case here.

are specified as continuous variables. Specification 2 augments this equation with the outcome variables considered in table 4, but with labour force status interacted with hours of work for those in paid employment. Finally, specification 3 adds to this model an array of variables describing the interview situation. With one exception, all explanatory variables are derived from data collected at wave 1. The exception is the mobility variable which is based on the observed movement of sample members between waves 1 and 2.⁸

Looking first at specification 1, it is immediately apparent that the probability of response at wave 2 does vary significantly with numerous individual and household characteristics. In particular, it rises with age (but at a declining rate) and educational attainment, falls with the number of adults in the household, is relatively low among indigenous Australians relative to non-indigenous Australians, persons born overseas relative to persons born in Australia, and in households living in flats, units or apartments relative to households in separate houses. There is also a strong regional effect, with response rates lowest in Sydney relative to most other parts of Australia. Further, the magnitudes of the effect of most of these variables are arguably quite large.⁹ The value of the pseudo R-squared term, however, is very low which, while only providing a crude measure of goodness of fit, suggests that most of the variation in attrition probabilities is either due to other factors or is random. This is a positive finding, implying that weights based on this equation would not greatly affect analyses of outcomes (Fitzgerald, Gottschalk and Moffitt, 1998, p. 276).

The additional outcome variables included in specification 2 raise the explanatory power of the model, but not by much. Indeed, of the variables considered, only mobility and housing tenure have effects that are statistically significant at conventional levels. Again, this is a positive finding. While most of the outcome variables considered here co-vary significantly with attrition probabilities, once we condition on a broad range of covariates these effects decline to insignificance. Furthermore, easily the most important variable for attrition – changing address – is one where, a priori, we would expect very large effects. Most obviously, households that change address between waves will simply be harder to find. Indeed, as revealed below, in table 6, once movers are located, there is no evidence that they are any more likely to refuse to participate.

⁸ We have assumed that all households lost in tracking have moved. It is possible, however, that contact might not have been made with a household and then that household subsequently be deemed untraceable, when in fact it had actually not changed address. We believe that the number of such cases is likely to be very few, especially given the wide range of tracking measures employed.

⁹ A guide to the size of these effects is provided by looking at the estimated odds ratio for each variable, which is simply the inverse log of the coefficient. Thus if we take the variable 'Other NSW', the coefficient of 0.312 gives an odds ratio of 1.366, meaning that the odds of people living in New South Wales but not in Sydney at wave 1 responding in wave 2 were 36.6 per cent greater than the odds of persons living in Sydney (the base group) responding.

Table 5 Wave 2 Response, Logit Results (n = 13,817)

| | (1) | | (2) | | (3) | |
|--|---------|----------------|---------|----------------|---------|----------------|
| | Coeff. | Standard Error | Coeff. | Standard Error | Coeff. | Standard Error |
| Constant term | 1.136 | 0.273 | 1.253 | 0.446 | 1.000 | 0.456 |
| <i>Personal / demographic characteristics</i> | | | | | | |
| <i>Area of residence (base = Sydney)</i> | | | | | | |
| Other NSW | 0.312 | 0.097 | 0.301 | 0.098 | 0.129 | 0.102 |
| Melbourne | 0.205 | 0.082 | 0.172 | 0.083 | 0.061 | 0.086 |
| Other Vic | 0.271 | 0.118 | 0.269 | 0.119 | 0.065 | 0.123 |
| Brisbane | 0.288 | 0.109 | 0.338 | 0.111 | 0.163 | 0.114 |
| Other Qld | 0.485 | 0.107 | 0.542 | 0.109 | 0.434 | 0.112 |
| Adelaide | 0.548 | 0.127 | 0.509 | 0.128 | 0.314 | 0.132 |
| Other SA | 0.391 | 0.171 | 0.354 | 0.173 | 0.194 | 0.178 |
| Perth | 0.411 | 0.118 | 0.402 | 0.118 | 0.259 | 0.122 |
| Other WA | 0.153 | 0.166 | 0.151 | 0.168 | 0.023 | 0.176 |
| Tasmania | 0.124 | 0.160 | 0.075 | 0.161 | -0.119 | 0.165 |
| Northern Territory | 1.592 | 0.601 | 1.742 | 0.606 | 1.724 | 0.617 |
| ACT | 0.597 | 0.241 | 0.578 | 0.242 | 0.263 | 0.245 |
| Female | 0.080 | 0.054 | -0.016 | 0.059 | 0.025 | 0.060 |
| Age | 0.055 | 0.010 | 0.060 | 0.010 | 0.075 | 0.010 |
| Age squared | -0.0005 | 0.0001 | -0.0006 | 0.0001 | -0.0007 | 0.0001 |
| <i>Marital status (base = Married)</i> | | | | | | |
| De facto | 0.676 | 0.550 | 0.715 | 0.564 | 0.778 | 0.580 |
| Separated | 0.323 | 0.359 | 0.352 | 0.365 | 0.322 | 0.376 |
| Divorced | 0.312 | 0.346 | 0.295 | 0.352 | 0.245 | 0.363 |
| Widowed | 0.697 | 0.359 | 0.662 | 0.364 | 0.646 | 0.374 |
| Single | 0.405 | 0.340 | 0.349 | 0.346 | 0.316 | 0.357 |
| <i>Relationship in household (base = Married couple)</i> | | | | | | |
| De facto couple | -1.051 | 0.552 | -0.962 | 0.566 | -0.976 | 0.582 |
| Lone parent | -0.694 | 0.332 | -0.562 | 0.341 | -0.364 | 0.353 |
| Child/relative | -0.268 | 0.347 | -0.347 | 0.351 | -0.296 | 0.362 |
| Lone person | -0.352 | 0.339 | -0.295 | 0.372 | -0.151 | 0.381 |
| Unrelated | -1.016 | 0.354 | -0.729 | 0.361 | -0.561 | 0.372 |
| Number of adults in h'hold | -0.287 | 0.031 | -0.313 | 0.034 | -0.218 | 0.043 |
| Number of children in h'hold | -0.024 | 0.027 | -0.034 | 0.027 | -0.057 | 0.028 |
| <i>Country of birth (base = Australia)</i> | | | | | | |
| Overseas: Main English-speaking | -0.290 | 0.087 | -0.261 | 0.088 | -0.251 | 0.090 |
| Overseas: Other | -0.618 | 0.071 | -0.618 | 0.073 | -0.353 | 0.089 |
| Indigenous | -0.399 | 0.167 | -0.308 | 0.171 | -0.179 | 0.175 |
| <i>Education (base = Year 11 and below)</i> | | | | | | |
| Year 12 | 0.175 | 0.082 | 0.212 | 0.084 | 0.210 | 0.086 |
| Certificate | 0.137 | 0.068 | 0.195 | 0.069 | 0.185 | 0.071 |
| Diploma | 0.371 | 0.109 | 0.432 | 0.111 | 0.354 | 0.113 |
| Bachelor or higher | 0.732 | 0.091 | 0.800 | 0.095 | 0.754 | 0.097 |
| <i>Dwelling type (base = Separate house)</i> | | | | | | |
| Semi-detached | -0.091 | 0.100 | 0.022 | 0.103 | 0.054 | 0.105 |
| Unit/apartment/flat | -0.396 | 0.092 | -0.259 | 0.096 | -0.221 | 0.099 |
| Other dwelling | 0.041 | 0.327 | 0.194 | 0.329 | 0.313 | 0.335 |
| <i>SEIFA disadvantage (base = Lowest quintile)</i> | | | | | | |
| 2nd lowest quintile | -0.242 | 0.083 | -0.258 | 0.084 | -0.261 | 0.087 |
| Middle quintile | -0.069 | 0.086 | -0.091 | 0.088 | -0.097 | 0.090 |
| 2nd highest quintile | -0.097 | 0.086 | -0.131 | 0.088 | -0.130 | 0.090 |
| Highest quintile | -0.007 | 0.092 | -0.062 | 0.094 | -0.070 | 0.096 |
| <i>Employment and LF status (base = Not in Labour Force)</i> | | | | | | |
| Unemployed | | | -0.149 | 0.124 | -0.106 | 0.128 |
| Employed PT (1-24hrs) | | | 0.145 | 0.096 | 0.136 | 0.099 |
| Employed PT (25-34hrs) | | | -0.133 | 0.126 | -0.157 | 0.129 |
| Employed FT (35-44hrs) | | | -0.154 | 0.096 | -0.151 | 0.099 |
| Employed FT (45-54hrs) | | | 0.011 | 0.120 | 0.024 | 0.123 |
| Employed FT (55+hrs) | | | -0.236 | 0.125 | -0.205 | 0.128 |
| <i>Housing tenure (base = Own/purchasing)</i> | | | | | | |
| Rent | | | -0.172 | 0.073 | -0.153 | 0.074 |
| Rent-buy, rent-free | | | 0.461 | 0.209 | 0.429 | 0.211 |

Table 5 (continued) Wave 2 Response, Logit Results (n = 13,817)

| | (1) | | (2) | | (3) | |
|--|---------|----------------|---------|----------------|---------|----------------|
| | Coeff. | Standard Error | Coeff. | Standard Error | Coeff. | Standard Error |
| Negative or zero personal income | | | -0.033 | 0.277 | -0.040 | 0.285 |
| Log personal income | | | -0.041 | 0.029 | -0.043 | 0.029 |
| Negative or zero income of others in h'hold | | | 0.510 | 0.315 | 0.740 | 0.320 |
| Log income of others in h'hold | | | 0.048 | 0.029 | 0.081 | 0.029 |
| Benefit recipient | | | 0.054 | 0.074 | 0.031 | 0.076 |
| High life satisfaction (8+) | | | 0.092 | 0.057 | 0.057 | 0.058 |
| Moved between w1 and w2 | | | -0.564 | 0.067 | -0.598 | 0.069 |
| <i>Interview situation (wave 1)</i> | | | | | | |
| Partially cooperating h'hold | | | | | -1.059 | 0.092 |
| Number of calls made to h'hold | | | | | -0.072 | 0.010 |
| Same interviewer in both waves | | | | | 0.024 | 0.056 |
| Interview time in h'hold | | | | | 0.000 | 0.001 |
| Interview time in h'hold unknown | | | | | 0.422 | 0.150 |
| Assistance and language difficulties during ivw (base = No assistance or difficulties) | | | | | | |
| Language difficulties, no assistance | | | | | -0.265 | 0.147 |
| Assistance for language reasons | | | | | -0.698 | 0.159 |
| Assistance for non-language reasons | | | | | -0.084 | 0.216 |
| Non-cooperative in interview ^a | | | | | -0.842 | 0.144 |
| Suspicious of study ^b | | | | | -0.554 | 0.110 |
| Log likelihood | -5090.2 | | -5019.6 | | -4809.3 | |
| Chi-squared | 602.1 | | 743.5 | | 1163.9 | |
| Pseudo R-squared | 0.056 | | 0.069 | | 0.108 | |

Notes: The number of cases included in model is 13969 wave 1 respondents less 152 deaths and movers overseas.

The pseudo R-squared equals 1 minus the ratio of the log likelihood of the fitted function to the starting value for the log likelihood (a function with only an intercept).

- a Equals 1 if the respondent's cooperation was described as fair, poor or very poor and 0 if described as excellent or good.
- b Equals 1 if the interviewer reported that the respondent was suspicious (either somewhat or very suspicious) of the study after the interview was completed.

Finally, in specification 3, we include a range of variables which describe the interview situation at wave 1. A priori, we expected that all of these variables would be significantly related to attrition. Laurie, Smith and Scott (1999), for example, in their analysis of attrition over the first four waves of the BHPS, found that response was significantly linked to interviewer observations recorded in wave 1 about the level of respondent cooperation and the presence of health and language problems that affected the interview, to whether the respondent was from a household where other household members did not cooperate in wave 1 and to interviewer continuity. We were able to include controls for all of these variables in this analysis of HILDA Survey data. In addition, we also included the total number of calls that had to be made to the household during wave 1, which we argue is a good measure of how difficult it will be to reach a household in wave 2.¹⁰ We also included a measure of total time spent by the interviewer in the household in wave 1 given it is typically assumed that interview lengths can have an influence on survey participation.¹¹ Finally, we included an interviewer-assessed measure of the degree of suspicion of the study exhibited by the respondent.

¹⁰ This variable will also be a function of the size of the household.

¹¹ In general, evidence from telephone and personal surveys support the hypothesised negative relationship between interview length and response rates, though the magnitude of the effect is arguably quite small (Frankel and Sharp, 1981; Collins, *et al.*, 1988).

As a group, these variables are clearly of considerable importance. The overall explanatory power of the model is somewhat enhanced, and most of the new variables exhibit large and statistically significant effects in the expected direction. Thus, in line with the results (Laurie, Smith and Scott, 1999) for the BHPS, we see that coming from a partially responding household is a major risk factor for non-participation at the next wave. Indeed, the estimates suggest that the mean predicted probability of a sample member from a partially cooperating household responding in wave 2 was about 84 per cent of that of sample members from fully responding households.¹² Also as expected, households that were more difficult to reach in wave 1 were much more likely to be non-respondents in wave 2. Similarly, interviewer assessments collected in wave 1 about the degree of respondent cooperation and suspicion were found to be good predictors of wave 2 non-response. Respondents whose interviews had to be assisted by others because of English language difficulties in wave 1 were also found to be more likely to be non-respondents. Finally, respondents with language difficulties, even though they did not need assistance to complete the interview, were also less likely to respond in wave 2.

Not all of the results here, however, were in accord with expectations. The findings in the BHPS data (Laurie, Smith and Scott, 1999) place great emphasis on the fact that respondents who were assigned the same interviewer each year were more likely to respond. The HILDA Survey also pursued a policy of maintaining interviewer continuity wherever possible, and indeed about 44 per cent of all households in wave 2 were assigned the same interviewer from wave 1. Nevertheless, we find no evidence that interviewer continuity matters.

The insignificance of the interview time variable might also surprise some. Interview time, however, is not a direct measure of instrument length. Indeed it is a product of instrument length, respondent interest in the survey, and respondent difficulty with the questions. Consequently, we expect the longer interviews to comprise a mix of respondents, some of whom found the interview very difficult and others who enjoyed the experience. Given this, the insignificance of interview time should not be surprising.

Of course, from a data user perspective, the more important question is not what influences attrition, but whether these influences are correlated with variables of interest (i.e., the outcome variables listed in table 4). The results presented in table 5 suggest that, for the most part, it is reasonable to conclude that there is little correlation between the variables of interest and the interview situation variables. The only outcome variable greatly affected by the inclusion of these interview situation variables is the income of other household members. This is a direct reflection of the correlation between this variable and the variable for partially cooperating households which, in turn, simply reflects the fact that single-person households cannot be partially cooperating households.

¹² The mean predicted probability for a dichotomous variable is the mean of the predicted probabilities where the variable is set to 0 (and then 1) for all cases, keeping all other variables constant at their reported values. The mean predicted probability of a person from a fully responding household in wave 1 responding in wave 2 was 0.88. This compares with a mean predicted probability of 0.74 for a person from a partially responding household.

Table 6 Two-equation Wave 2 Response Model, Logit Results

| Variable | Contact | | Response | Contact |
|---|---------|----------------|----------|----------------|
| | Coeff. | Standard Error | Coeff. | Standard Error |
| Constant term | 1.319 | 0.931 | 1.763 | 0.510 |
| <i>Personal / demographic characteristics</i> | | | | |
| Area of residence (base = Sydney) | | | | |
| Other NSW | 0.656 | 0.224 | 0.003 | 0.112 |
| Melbourne | 0.449 | 0.191 | -0.021 | 0.094 |
| Other Vic | 0.615 | 0.291 | -0.041 | 0.133 |
| Brisbane | 0.294 | 0.225 | 0.102 | 0.127 |
| Other Qld | 0.227 | 0.197 | 0.549 | 0.134 |
| Adelaide | 0.867 | 0.300 | 0.177 | 0.143 |
| Other SA | 0.145 | 0.337 | 0.195 | 0.204 |
| Perth | 0.052 | 0.226 | 0.336 | 0.141 |
| Other WA | -0.394 | 0.303 | 0.203 | 0.210 |
| Tasmania | 1.640 | 0.609 | -0.383 | 0.172 |
| Northern Territory | 0.437 | 0.774 | 2.511 | 1.023 |
| ACT | -0.078 | 0.435 | 0.388 | 0.291 |
| Female | 0.268 | 0.125 | -0.031 | 0.067 |
| Age | 0.067 | 0.023 | 0.072 | 0.012 |
| Age squared | -0.0004 | 0.0003 | -0.0008 | 0.0001 |
| Marital status (base = Married) | | | | |
| De facto | 2.244 | 1.195 | 0.392 | 0.659 |
| Separated | 0.405 | 0.593 | 0.405 | 0.458 |
| Divorced | 0.388 | 0.585 | 0.205 | 0.436 |
| Widowed | 1.043 | 0.729 | 0.514 | 0.439 |
| Single | 0.738 | 0.570 | 0.265 | 0.431 |
| Relationship in household (base = Married couple) | | | | |
| De facto couple | -2.772 | 1.194 | -0.522 | 0.662 |
| Lone parent | -1.126 | 0.578 | -0.199 | 0.422 |
| Child/relative | -1.038 | 0.586 | -0.243 | 0.435 |
| Lone person | -1.317 | 0.634 | 0.072 | 0.457 |
| Unrelated | -1.517 | 0.593 | -0.327 | 0.453 |
| Number of adults in h'hold | -0.018 | 0.088 | -0.236 | 0.046 |
| Number of children in h'hold | -0.148 | 0.053 | -0.040 | 0.032 |
| Country of birth (base = Australia) | | | | |
| Overseas: Main English-speaking | -0.436 | 0.180 | -0.187 | 0.100 |
| Overseas: Other | -0.278 | 0.191 | -0.343 | 0.097 |
| Indigenous | -0.657 | 0.249 | 0.308 | 0.247 |
| Education (base = Year 11 and below) | | | | |
| Year 12 | 0.313 | 0.175 | 0.134 | 0.095 |
| Certificate | 0.029 | 0.145 | 0.195 | 0.079 |
| Diploma | 0.324 | 0.251 | 0.334 | 0.124 |
| Bachelor or higher | 0.459 | 0.200 | 0.773 | 0.108 |
| Dwelling type (base = Separate house) | | | | |
| Semi-detached | -0.043 | 0.189 | 0.051 | 0.122 |
| Unit/apartment/flat | -0.280 | 0.169 | -0.133 | 0.118 |
| Other dwelling | -0.400 | 0.470 | 0.780 | 0.478 |
| SEIFA disadvantage (base = Lowest quintile) | | | | |
| 2nd lowest quintile | -0.179 | 0.160 | -0.297 | 0.099 |
| Middle quintile | 0.071 | 0.176 | -0.170 | 0.102 |
| 2nd highest quintile | 0.047 | 0.180 | -0.179 | 0.102 |
| Highest quintile | 0.249 | 0.199 | -0.159 | 0.108 |
| Employment and LF status (base = Not in Labour Force) | | | | |
| Unemployed | -0.025 | 0.203 | -0.041 | 0.157 |
| Employed PT (1-24hrs) | 0.275 | 0.192 | 0.079 | 0.112 |
| Employed PT (25-34hrs) | 0.667 | 0.294 | -0.364 | 0.141 |
| Employed FT (35-44hrs) | 0.615 | 0.196 | -0.367 | 0.112 |
| Employed FT (45-54hrs) | 0.763 | 0.254 | -0.194 | 0.137 |
| Employed FT (55+hrs) | 0.188 | 0.257 | -0.325 | 0.144 |
| Housing tenure (base = Own/purchasing) | | | | |
| Rent | -0.206 | 0.142 | -0.111 | 0.085 |
| Rent-buy, rent-free | 1.052 | 0.603 | 0.314 | 0.223 |

Table 6 (continued) Two-equation Wave 2 Response Model, Logit Results

| Variable | Contact | | Response Contact | |
|--|---------|----------------|--------------------|----------------|
| | Coeff. | Standard Error | Coeff. | Standard Error |
| Negative or zero personal income | 0.654 | 0.584 | -0.192 | 0.316 |
| Log personal income | -0.025 | 0.060 | -0.044 | 0.033 |
| Negative or zero income of others in h'hold | 1.610 | 0.636 | 0.547 | 0.359 |
| Log income of others in h'hold | 0.155 | 0.059 | 0.059 | 0.032 |
| Benefit recipient | 0.181 | 0.146 | -0.011 | 0.086 |
| High life satisfaction (8+) | 0.215 | 0.114 | 0.000 | 0.066 |
| Moved between w1 and w2 | -2.035 | 0.130 | 0.005 | 0.087 |
| <i>Interview situation (wave 1)</i> | | | | |
| Partially cooperating h'hold | -0.594 | 0.204 | -1.100 | 0.098 |
| Number of calls made to h'hold | -0.054 | 0.021 | -0.075 | 0.012 |
| Same interviewer in both waves | -0.190 | 0.120 | 0.068 | 0.062 |
| Interview time in h'hold | 0.000 | 0.002 | 0.000 | 0.001 |
| Interview time in h'hold unknown | 0.485 | 0.324 | 0.380 | 0.163 |
| Assistance and language difficulties during ivw (base = No assistance or difficulties) | | | | |
| Language difficulties, but no assistance | -0.893 | 0.258 | -0.101 | 0.170 |
| Assistance for language reasons | -1.063 | 0.309 | -0.630 | 0.174 |
| Assistance for non-language reasons | -0.056 | 0.474 | -0.070 | 0.237 |
| Non-cooperative in interview ^a | -0.999 | 0.261 | -0.730 | 0.163 |
| Suspicious of study ^b | 0.051 | 0.253 | -0.652 | 0.117 |
| Log likelihood | | -1367.2 | | -4077.1 |
| Chi-squared | | 943.6 | | 889.3 |
| Pseudo R-squared | | 0.257 | | 0.098 |
| N | | 13817 | | 13409 |

Notes: The number of cases included in model of contact is 13969 wave 1 respondents less 152 deaths and movers overseas. The number of cases included in the model of response given contact is a further 408 wave 1 respondents less as no contact was made with these people.

a See note on table 5.

b See note on table 5.

Our final set of analyses of attrition involved estimating logit equations that distinguished between two key stages in the response process – establishing successful contact (column 2) and then obtaining a successful interview (column 3). The results, using the full specification are presented in table 6.

These results add insights into the response process. Most obviously, a number of the explanatory variables are of much greater importance in explaining variations in the probability of making contact while others are of much greater importance in explaining the variation in response probabilities once contact has been established. The clearest example of the importance of this distinction is the variable identifying those who move house. As noted earlier, the relocation of a household has a marked impact on the likelihood of finding the members of that household. Indeed, based on the parameter estimates the mean predicted probability of making contact with a mover is 92 per cent. While seemingly high, this is well below the mean predicted probability of making contact with a non-mover – almost 99 per cent. Movement, however, has no influence on the likelihood of obtaining an interview once contact is established.

Other variables which have markedly different impacts at the two separate stages of the response process include the following:

- sex – females are easier to make contact with but are no more or less likely than males to agree to an interview;
- household type – married couple households are easier to find than other household types, but again refusal probabilities do not vary with household type;
- number of persons in the household – a greater number of adults in the households is associated with lower response probabilities but has no bearing on the likelihood of making contact, while the number of children has the opposite effects;
- indigenous status – the lower response probabilities of indigenous Australians are entirely due to greater difficulties making contact; and
- employment status – workers (and more specifically, those working between 25 and 55 hours per week) are relatively easy to make contact with but, with the exception of short-hours workers (persons working less than 25 hours per week), are more likely than non-workers to refuse to participate.

4. Effect of Attrition on Population Estimates

In this section we briefly consider the impact that attrition has on population estimates. Table 7 provides two population estimates for selected sample characteristics measured at wave 2. The first population estimate is simply based on population weights carried forward from wave 1, whereas the second population estimate uses weights that have made adjustments for the non-random attrition (further details of the construction of the weights are provided in the footnotes of table 7). For example, if the attrition between waves 1 and 2 was ignored, the estimate for the proportion of people living in Sydney in wave 2 would be 19.9 per cent. However, when the differential attrition is factored into the weighting process, the proportion of people living in Sydney is revised upwards to 21.2 per cent.

As can be seen from this table, for a number of the characteristics the two estimates are quite different. The largest differences occur when the attrition has been high for a particular group and the population size is reasonably large. For example, we would understate the population estimates for the proportion of people who live in Sydney, are single, are relatively young, were born in a mainly non-English speaking country, are currently renting, or who have moved since wave 1. Conversely, we would overstate the proportion of people who are married, born in Australia, have high education levels, or have high levels of life satisfaction.

In areas where the attrition has been reasonably differential but the proportion of the population affected is relatively small, the effect on the population estimates is less apparent. This can be seen in the estimates for indigenous status, the proportion of people living in the Northern Territory or ACT, those that are widowed, the unemployed and those reporting low levels of life satisfaction.

In general though it is clear that the estimates most affected by attrition are demographic characteristics. Consistent with our earlier multivariate analysis, weighting for attrition between wave 1 and wave 2 appears to make very little difference to the estimation of our selected outcome variables.

Table 7 Wave 2 Characteristics Using Unadjusted Weights and Weights Adjusted for Attrition (Per cent)

| Wave 2 Characteristic | Unadjusted Weights ^a | Adjusted Weights ^b | Wave 2 Characteristic | Unadjusted Weights ^a | Adjusted Weights ^b |
|--|---------------------------------|-------------------------------|----------------------------------|---------------------------------|-------------------------------|
| Area | | | Indigenous status | | |
| Sydney | 19.9 | 21.2 | Indigenous | 1.6 | 1.7 |
| Rest of NSW | 12.8 | 12.5 | Non-indigenous | 98.4 | 98.3 |
| Melbourne | 18.2 | 18.4 | Education attainment | | |
| Rest of Victoria | 7.1 | 7.0 | Year 11 or below | 33.1 | 33.4 |
| Brisbane | 8.8 | 8.7 | Year 12 | 12.3 | 13.0 |
| Rest of Qld | 10.1 | 9.8 | Certificate | 26.7 | 26.6 |
| Adelaide | 6.1 | 5.8 | Diploma | 8.5 | 8.4 |
| Rest of SA | 2.1 | 2.1 | Degree of higher | 19.3 | 18.5 |
| Perth | 7.5 | 7.2 | Dwelling type^d | | |
| Rest of WA | 2.6 | 2.6 | House | 80.1 | 79.3 |
| Tasmania | 2.5 | 2.5 | Semi-detached | 7.3 | 7.4 |
| Northern Territory | 0.7 | 0.7 | Flat, unit, apartment | 9.8 | 10.4 |
| ACT | 1.7 | 1.6 | Other | 1.7 | 1.7 |
| Sex | | | Index of disadvantage | | |
| Male | 49.2 | 49.5 | Lowest quintile | 18.2 | 18.8 |
| Female | 50.8 | 50.5 | 2 nd lowest quintile | 22.0 | 21.9 |
| Age group (years) | | | Middle quintile | | |
| 15-19 | 7.0 | 7.6 | 2 nd highest quintile | 18.2 | 18.2 |
| 20-24 | 7.7 | 8.6 | Highest quintile | 23.0 | 22.7 |
| 25-34 | 18.6 | 19.0 | Labour force status | | |
| 35-44 | 19.7 | 19.3 | Employed full-time | 43.2 | 43.3 |
| 45-54 | 18.2 | 17.8 | Employed part-time | 18.7 | 18.6 |
| 55-64 | 13.1 | 12.5 | Unemployed | 3.6 | 3.9 |
| 65-74 | 9.1 | 8.6 | Not in labour force | 34.4 | 34.1 |
| 75+ | 6.6 | 6.5 | Life satisfaction | | |
| Marital status | | | Low (0-4) | | |
| Married | 54.7 | 53.5 | Medium (5-7) | 30.9 | 31.3 |
| De facto | 9.2 | 9.6 | High (8-10) | 65.9 | 65.4 |
| Separated | 3.2 | 3.1 | Household tenure | | |
| Divorced | 5.3 | 5.0 | Own/purchasing | 73.0 | 71.5 |
| Widowed | 5.5 | 5.3 | Rent | 24.2 | 25.7 |
| Single | 22.2 | 23.5 | Other | 2.7 | 2.8 |
| Relationship in h'hold | | | Benefit recipient status | | |
| Married couple | 54.3 | 53.1 | Benefit recipient | 34.3 | 34.4 |
| De facto couple | 9.0 | 9.4 | Not benefit recipient | 65.6 | 65.6 |
| Lone parent | 5.4 | 5.4 | Annual household income | | |
| Child/relative | 13.9 | 14.9 | <\$20,000 | 15.6 | 15.4 |
| Lone person | 15.3 | 14.8 | \$20,000 - \$39,999 | 19.6 | 19.6 |
| Unrelated | 2.1 | 2.3 | \$40,000 - \$59,999 | 18.3 | 18.4 |
| No. of adults in h'hold^c | | | \$60,000 - \$79,999 | | |
| One adult | 17.6 | 17.1 | \$80,000 - \$119,999 | 15.6 | 15.7 |
| Two adults | 52.6 | 51.3 | \$120,000+ | 19.4 | 19.4 |
| Three adults | 17.4 | 18.1 | Annual personal income | | |
| Four or more adults | 12.5 | 13.5 | Negative/nil | 3.0 | 3.3 |
| No. of children in h'hold | | | \$1 - \$9,999 | | |
| No children | 69.8 | 69.8 | \$10,000 - \$14,999 | 22.4 | 22.0 |
| One child | 12.3 | 12.6 | \$15,000 - \$29,999 | 13.5 | 13.3 |
| Two children | 11.8 | 11.5 | \$30,000 - \$59,000 | 20.7 | 20.9 |
| Three or more | 6.1 | 6.1 | \$60,000+ | 28.5 | 28.2 |
| Country of birth | | | Did not move | | |
| Australia | 73.8 | 72.2 | Moved b/w wave 1 and 2 | 16.3 | 18.0 |
| O/S: Main English-spkg | 10.9 | 10.9 | Did not move | 83.7 | 82.0 |
| O/S: Other | 15.2 | 16.9 | | | |

Notes: a The unadjusted weights are the wave 1 weights for responding wave 2 individuals.

b The adjusted (for attrition) weights are calculated by: i) multiplying the wave 1 responding person weights by the inverse of the modeled probability of response to wave 2 (given an interview provided in wave 1); and ii) benchmarking these weights for the wave 2 respondents and out-of-scopes to wave 1 characteristics such as age, sex, State, part of State, and labour force status.

c An adult is defined here as anyone aged 15 years or over. A child is therefore anyone under the age of 15 years.

d Not reported are a small proportion of cases where the dwelling type was not reported.

5. Conclusions

The key conclusion from this analysis is very clear – the characteristics of those who attrit and those who do not are quite different. It is thus difficult to avoid the conclusion that the potential for attrition bias exists and, more importantly, is likely to become more serious as the panel develops. In particular, it seems likely that without any correction, the data are likely to overstate population stability. Whether any correction is needed, and how to make that correction, is a question that will be specific to the type of model being estimated. Our own judgment, however, is that for the first few waves at least, any bias is likely to be relatively small. The relatively low explanatory power of our estimated models and the general insignificance of most outcome variables in these models suggests that attrition will not have much impact on most analyses of these data. The analysis of attrition within the PSID sample by Fitzgerald, Gottschalk and Moffitt (1998) reached the conclusion that in a regression context, attrition was likely to affect intercept terms but have relatively little effect on the slopes of key coefficients. We suspect a similar conclusion is applicable to the HILDA sample. A productive avenue for future research would be to confirm this result in the HILDA context and to test the effectiveness of the interview situation variables as instruments in dealing with the non-random attrition.¹³

Researchers, however, will eventually be forced to deal with the potential for bias that attrition gives rise to. One approach often favoured by economists is the use of sample selection models. The main problem with this approach, however, is that it is often difficult to identify instruments that are not also correlated with the dependent variable of interest. As we have seen, the HILDA Survey does collect data on the interview situation which, in many situations, may provide acceptable instruments.

The other main alternative is through the use of weights. The survey management team has greatly assisted the ability of data users to implement weighting approaches by providing a set of population weights that adjust for panel attrition within the data set. Nevertheless, it still needs to be borne in mind that the weights are only as good as the model used to predict attrition. There are almost certainly many unobserved influences on attrition, and if these influences are correlated with the variables of interest, then it is not obvious that weighted estimates will completely eliminate the effect of attrition bias.

References

- Buck, N., Burton, J., Laurie, H. and Lynn, P. (2002), *Quality Profile – British Household Panel Survey Waves 1 to 10: 1991-2000*, Institute for Social and Economic Research, University of Essex, Colchester.
- Collins, M., Sykes, W., Wilson, P. and Blackshaw, N. (1988), 'Nonresponse: The UK Experience', in Groves, R.M., Biemer, P., Lyberg, L.E., Massey, J.T., Nicholls II, W.L., and Waksberg, J. (eds), *Telephone Survey Methodology*, John Wiley and Sons, New York.
- Duncan, G. (2000), 'Using Panel Studies to Understand Household Behaviour and Well-being', in Rose, D. (ed.), *Researching Social and Economic Change: The Uses of Household Panel Studies*, Routledge, London.

¹³ This suggestion of further research was made by a referee.

- 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*, 33, 251-99.
- Frankel, J. and Sharp, L.M. (1981), 'Measurement of Respondent Burden', *Statistical Reporter*, 81, 105-111.
- Haisken-DeNew, J. (2001), 'A Hitchhiker's Guide to the World's Household Panel Data Sets', *The Australian Economic Review*, 34, 356-66.
- Laurie, H., Smith, R. and Scott, L. (1999), 'Strategies for Reducing Nonresponse in a Longitudinal Panel Survey', *Journal of Official Statistics*, 15, 269-282.
- Rose, D. (2000), 'Household Panel Studies: An Overview', in Rose, D. (ed.), *Researching Social and Economic Change: The Uses of Household Panel Studies*, Routledge, London.
- Watson, N. and Wooden, M. (2002), 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 Survey Methodology', HILDA Project Technical Paper Series no. 1/02, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.
- Wooden, M. and Watson, N. (2002), 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: An Introduction', *Australian Social Policy*, 2001-02, 79-100.