HILDA PROJECT TECHNICAL PAPER SERIES
NO. 4/04, July 2004

Wave 2 Weighting

Nicole Watson

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

Contents

Introduction

The second wave of the HILDA Survey is the first year in which we have had to work directly with the longitudinal nature of the survey in constructing the weights. In wave 1, we essentially had a complex cross-sectional survey. Now, in wave 2, the 'selection' of the sample is dependent on the wave 1 responding sample and the household and individual attrition between waves 1 and 2. We are interested in both cross-section estimates from wave 2 as well as longitudinal estimates across the two waves.

This paper details, primarily for the users of the data, the methodology used to construct the various wave 2 weights.

An overview of the weighting process is provided in Figure 1 below. Five weights are constructed for wave 2, these being:

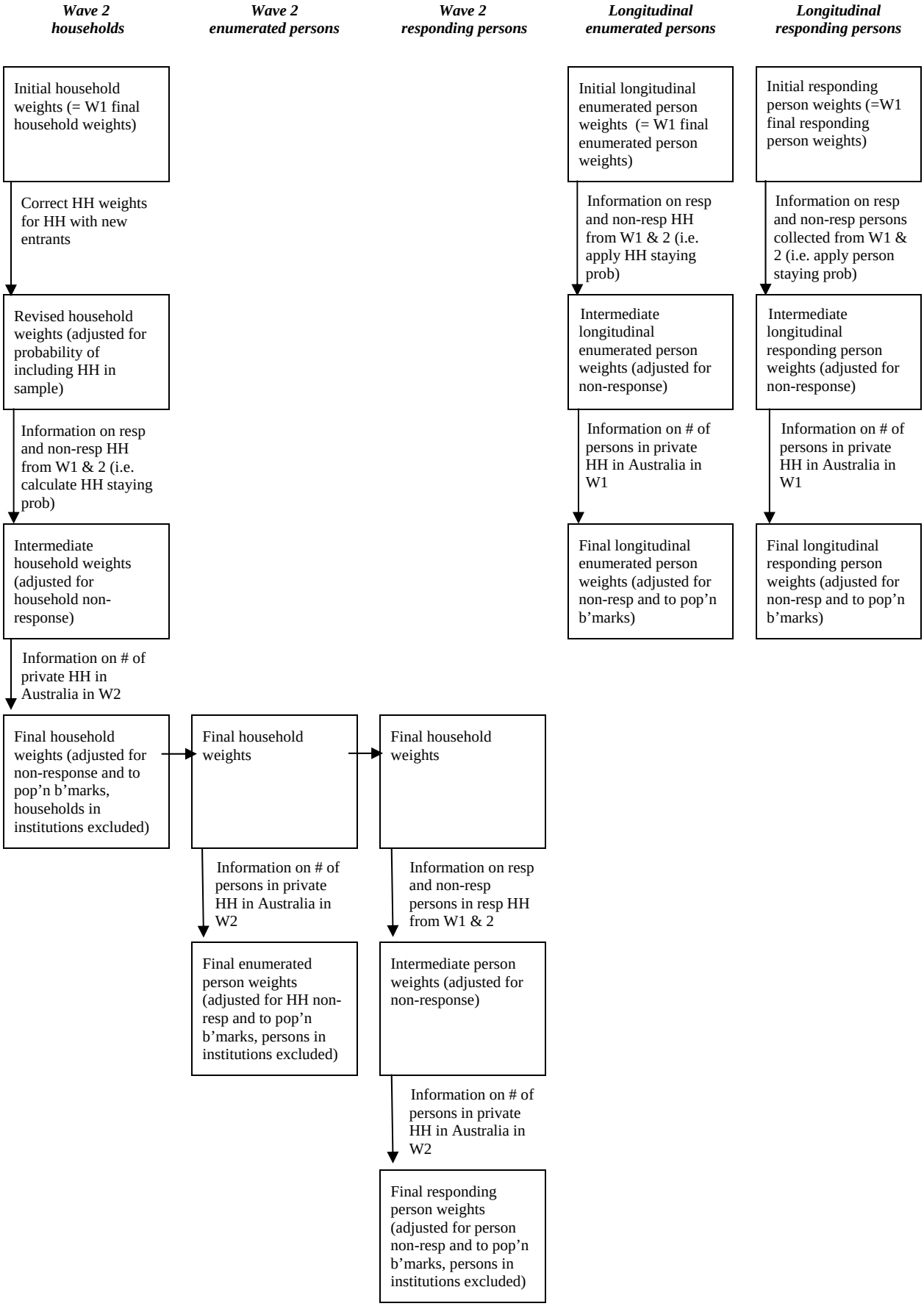
- cross-sectional household weights for wave 2 households;
- cross-sectional enumerated person weights for wave 2 individuals;
- cross-sectional responding person weights for wave 2 respondents;
- longitudinal enumerated person weights for individuals in waves 1 and 2; and
- longitudinal responding person weights for respondents in waves 1 and 2;

The cross-sectional weights for wave 2 opportunistically include temporary members into the sample (i.e., those people who are part of the sample only because they currently live with a continuing sample member). The underlying probability of selection for these households is amended to account for the various pathways into the wave 2 household. Following this, non-response adjustments are made which require within-sample modeling of non-response probabilities and benchmarking to known population estimates.

By comparison, the construction of the longitudinal weights is more straightforward and only include an adjustment for attrition and benchmarking back to wave 1 characteristics.

The weighting methodology was developed with input from the HILDA Technical Reference Group (whose membership is provided in Appendix 1). The effort each member put into understanding, discussing and helping to resolve the technical issues is greatly appreciated. Dr Martin Spiess from DIW Berlin was also of great assistance in improving our understanding of the German Socio-Economic Panel (GSOEP) weighting methodology and provided a useful sounding board for the HILDA approach.

Figure 1: Overview of wave 2 weighting



Sample Design: The Following Rules

What are the Following Rules?

As detailed in Watson and Wooden (2004b), the fully and partially responding households in wave 1 form the basis of the indefinite life panel. Members of these households are followed over time and the sample is extended to include:

- any children born to or adopted by members of the selected households; and
- new household members resulting from changes in the composition of the original households.

Continuing sample members include all members of wave 1 households (including children). Any children born to or adopted by continuing sample members are also classified as 'continuing'. Further, all new entrants to a household who have a child with a continuing sample member are converted to 'continuing' status. Continuing sample members remain in the sample indefinitely. All other people who share a household with these sample members in wave 2 or later are considered temporary sample members.

Where the household has moved, split or moved and split, the interviewers and office staff track the continuing sample members. These people (along with their new household) are then interviewed, where applicable, at their new address or by phone.[1] Temporary sample members that split from a household and are no longer part of a household with a continuing sample member are not followed. However, if the temporary sample member is converted to the 'continuing' status, then they are followed for interview as any continuing sample member would be.

Implications of the Following Rules for the Sample Composition

From wave 1, 19,914 continuing sample members were identified (being all people in fully and partially responding households). A further 233 continuing sample members were added to this number during wave 2 and were followed into wave 3:

- 212 new born babies;
- 2 adopted children; and
- 19 parents of these continuing sample members who were not previously counted as 'continuing'.

There were an additional 895 temporary sample members added to the sample for wave 2, one third of which left the sample in wave 3 as they ceased living with a continuing sample member.

[1] Note that if a child who is a continuing sample member moves without any other continuing sample member adult, they are followed to their new household and the eligible members of that household are then interviewed.

Cross-Sectional Weights

Household Weights

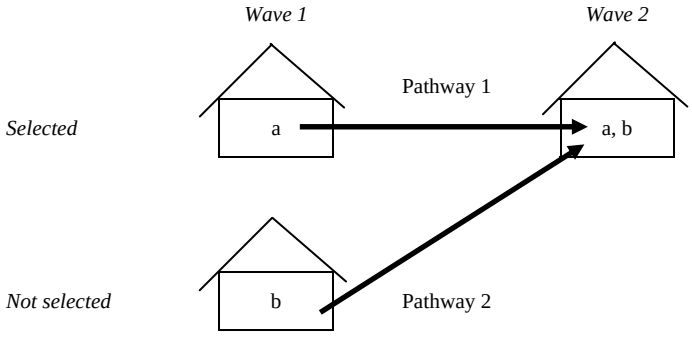
Correcting the Initial Weights for the Effect of New Entrants

As new entrants are included in the cross-sectional sample, the household and person weights need to be corrected to reflect the probability of selection into the wave 2 sample.

The motivation for this correction is illustrated with the following example. The household with person *a* was selected in wave 1. We have followed this household into wave 2 and found that new entrant *b* has moved in. Now, we could also have found this household in the wave 2 sample had we selected the household with person *b* in wave 1. The cross-section weight of the wave 2 household with person *a* and *b* needs to be down-weighted to reflect the multiple paths through which we could have selected this household: 'pathway 1' through which we *did* select the wave 2 household and 'pathway 2' which we could have followed had *b*'s household been selected in wave 1.

If we do not make this correction to the initial wave 2 household cross-sectional weights, we would overstate the number of households with new entrants compared to the population and therefore bias the results towards the activities of these households.

Figure 2: Example of pathways into a wave 2 household



The correction to the initial household weights involves the following steps:

- Step 1: Identify family groups within the new entrants joining the household. Related people are assumed to join the wave 2 household together. Unrelated people are assumed to join the household separately. Newly born babies and adoptions are considered part of the 'intact' household group (they are organic additions to the sample). From the 912 new entrants that were not organic additions to the sample, 694 new entrant family groups were identified.

- Step 2: Identify a reference person within each of these new entrant family groups. The reference person is the first within the family group to satisfy the following ordered requirements: couple, lone parent, non-dependent child,

dependent child, other related, not related. A preference for a respondent as the household reference person was taken over a non-respondent (so that as much personal information could be used as possible).[2]

- Step 3: Construct a regression model to predict a 'quasi-selection' probability for the new entrant family groups. This consists of the following steps:

  - Step 3a: Identify a reference person within the intact group from the selected wave 1 household, using similar criteria as above.

  - Step 3b: Convert the final wave 1 household weight to a 'quasi-selection' probability by taking the inverse of the weight (that is, $p_{hh,w1} = 1/w_{hh,w1}$).[3] As the 'quasi-selection' probability is bounded by 0 and 1, transform it into a new variable y which has a continuous scale, via the following:

$$y = \ln\left[\frac{p_{hh,w1}}{(1-p_{hh,w1})}\right]$$

  - Step 3c: Construct a regression model of the transformed variable y using the wave 2 person information for the reference person of the intact group and the wave 2 household information (i.e., using cases like *a* in the illustration above). The details of this model are provided in Appendix 2 (Table A2.1).

  - Step 3d: Use this model to predict a wave 1 'quasi-selection' probability ($\hat{p}_{f_{i,w1}}$) for the new entrant family groups (i.e., for cases like *b* in the above illustration). From the model of y, obtain an estimate $\hat{y}$ given the characteristics of the household and the reference person of the new entrant family group. Transform $\hat{y}$ into the probability for the i^{th} new entrant family group using:

$$\hat{p}_{f_{i,w1}} = \frac{e^{\hat{y}}}{(1+e^{\hat{y}})}$$

- Step 4: Construct the revised wave 2 household weight which adjusts for the multiple pathways into the wave 2 household. This adjustment is done via the following formula which accounts for the joint selection probabilities of these family groups:

---

[2] This preferential identification of respondents affected nine family groups. In four of these family groups, the respondent was at the same level of relationship classification as the non-respondent and in the remaining five family groups, the respondent was at a lower level than the non-respondent.

[3] The construction of the final wave 1 household weights is described in Watson and Fry (2002). The final household weight reflects the differences in selection probabilities and the response probabilities in wave 1. As we have incorporated the response probabilities, we refer to the inverse of the final wave 1 weight as a 'quasi-selection' probability.

$$w_{hhrvsd,w2} = 1 \Big/ \left[1-(1-p_{hh,w1})*(1-\hat{p}_{f_{1,w1}})*...*(1-\hat{p}_{f_{n,w1}}) \right]$$

where $p_{hh,w1}$ is the 'quasi-selection' probability for the intact family group, and $\hat{p}_{f_{i,w1}}$ is the estimated 'quasi-selection' probability for the new entrant family i. For new entrant family groups where nobody responded in wave 2, the wave 1 'quasi-selection' probability is taken to be zero as it is likely they would not have responded in wave 1 (so would not have been followed along that pathway into wave 2).[4]

We have generally followed the GSOEP approach in making this adjustment, but have included a number of enhancements.[5] These modifications include:

- identifying family groups and assuming they moved into the household together;
- using both household and person level information in the model to predict the wave 1 household selection probability for the joiners; and
- allowing for joint selection probabilities in the revised weight.

In contrast, the GSOEP method treats new entrants independently of each other, uses only person level information in the model of selection probabilities, and ignores the joint selection probabilities (i.e. treats them as zero).

An alternative method to adjust for the inclusion of the new entrants is the 'fair shares approach' that is used by the British Household Panel Study (BHPS, see Taylor et al. 2003). Under this method the sum of the weights of the wave 1 household members, after adjusting for non-response, is divided equally among the wave 2 household members. That is, the BHPS method assumes that the new entrants are like the existing household members. We considered this to be a relatively simple adjustment and the GSOEP-type approach is likely to be more accurate. However, the BHPS method does have the advantage that it is less likely to generate extreme weights.

Correcting the Initial Weights for Merged Households

For the four wave 1 households that merged with other wave 1 households in wave 2, the initial wave 2 household weight is revised via the application of step four described above. We do not need to model the wave 1 'quasi-selection' probability for these households as it is known.

Households in Non-Private Dwellings

A total of 18 wave 2 households had moved into non-private dwellings. As the wave 1 sample excluded non-private dwellings and the cross-sectional benchmarks excluded

[4] There were 177 new entrant family groups where nobody responded.

[5] The weighting methodology for GSOEP is documented in Pannenberg, et al. (2003).

non-private dwellings, these households have their cross-sectional weight set to zero.[6] (Similarly, the cross-section person weights for people in these non-private dwellings are also zero – this affects 24 enumerated persons and 19 responding persons.)

Non-Response Adjustments using Data Internal to the HILDA Survey

The adjustment to the weights for non-response makes the greatest difference to the weights. Following the correction to the initial household weights due to the effect of new entrants, the weights are adjusted for the probability that the household stayed in the responding sample for wave 2.

The probability that a household would stay in the responding sample was modelled using logistic regression. The characteristics included in the model were:

Wave 1 household characteristics
Location (State by part of State)
Remoteness area
SEIFA index of disadvantage
Dwelling type
Condition of dwelling
Number of bedrooms per person
Number of calls made to household
Whether household was partly responding
Number of person in household
Number of adults in household
Number of children in household
Household type
Housing tenure
Know whether have benefit recipient in household
Household income for last financial year
Missing household income
Time in household interviewing
Time in household unknown

Wave 1 reference person characteristics
Sex
Age
Age squared
Female aged 65 or over
Marital status
Ability in speaking English
Employment status and hours
Number of children reference person has
Country of birth
Highest level of education achieved
Relationship in household
Health status

[6] It would not be sensible to make population inferences from a sample consisting mainly of households in private dwellings together with households moving into non-private dwellings in 2002.

Likelihood of moving
Number of times moved in last 10 years
Length of PQ interview in wave 1
Length of PQ interview unknown
Whether completed SCQ in wave 1
Whether reference person provided PQ interview in wave 1

Wave 2 sample characteristics collected on all wave 2 households
Household split in wave 2
Whether moved between waves 1 and 2

The details of the model are provided in Appendix 2 (Table A2.2).

As we are interested in which wave 2 households are likely to respond, households that split into multiple parts in wave 2 were considered separately and households that merge into one were considered as one household.

The intermediate household weights are then constructed by multiplying the revised initial household weights by the inverse of the probability of the household staying in the responding sample. That is,

$$w_{hhinterim,w2} = w_{hhrvsd,w2} * 1 \Big/ \hat{p}_{hhstay,w2}$$

This means that households that are least likely to stay and actually do stay have a greater inflation factor applied to their household weight than other households (reflecting the fact that these households are less common in the responding sample). A minimum value for $\hat{p}_{hhstay,w2}$ was applied to avoid extreme weights. Households with a predicted probability of staying of less than 0.3 had their staying probability set to 0.3. This affected 21 responding households.

Non-Response Adjustments using Data External to the HILDA Survey

The final step in the creation of the household weights was to ensure the sum of the weights matched appropriate population benchmarks. The benchmarks used were:

- Household benchmark 1:- Number of households by State and part of State. For NSW, Vic, Qld, SA and WA, the part of State variable separated the metropolitan area from the rest of the State. For Tas, NT, and ACT, part of State was not used.

- Household benchmark 2:- Household type (based on number of adults and children) by broad geographic areas. There were nine household types combining one, two, and three or more adults (aged 15 and over) with zero, one and two or more children (aged under 15). The broad geographical areas included Sydney, Melbourne, Brisbane, ACT combined with rural NSW, WA combined with SA, Tas combined with rural Vic, NT combined with rural Qld.

These benchmarks were obtained from the Australian Bureau of Statistics (as a special data service) and relate to the estimated number of households in Australia as

at 30 September 2002. The benchmarks excluded households in remote areas of NT and included only those households in private dwellings.

The household weights were simultaneously calibrated to both sets of benchmarks using a SAS macro called GREGWT (developed in the Statistical Services area of the ABS).[7] Appendix 3 provides some information on how the weights were changed through the various adjustments made.

Person Weights

Following the practice adopted in wave 1, two sets of person weights have been constructed: enumerated person weights and responding person weights. This stems from the fact that not everyone who was eligible for interview actually provided an interview. Of the 7245 households participating in Wave 2, 9.7 per cent had at least one eligible person who did not complete an interview. At the person level, this translates to 7.1 per cent of all eligible people in the participating households that did not complete a person interview.

Each person who is a usual resident of a responding household has been assigned an enumerated person weight (this includes respondents, non-respondents and children). Each person providing a personal interview has been assigned a responding person weight.

Initial Weights

In line with the practice in wave 1, the initial enumerated person weight and responding person weight is equal to the final household weight (to encourage consistency between the person level and household level weights).

Non-Response Adjustment Using Data Internal to the HILDA Survey – Responding Person Weights Only

Information about the respondents and non-respondents in responding households was used to make a response adjustment to the responding person weights.

The probability that the person would provide an interview (given their household had responded) was modelled using logistic regression. This model was restricted to people aged 15 and over in responding households with two or more eligible persons. The characteristics included in the model were:

Wave 2 person characteristics
Female
Age
Age squared
Female aged 65 or over
Relationship in household

Wave 2 household characteristics

[7] The GREGWT macro performs generalized regression weighting as described in Stukel, Hidiroglou and Sarndal (1996).

Location (State by part of State)
Remoteness area
SEIFA index of disadvantage
Dwelling type
Number of bedrooms per person in household
Number of calls made to household
Number of persons in household
Three or more adults in household in wave 2
Number of children in household
Household type
Housing tenure
Household split in wave 2
Whether moved between wave 1 and 2
Whether joiners to household
Whether leavers from household
Whether both joiners and leavers

Wave 1 household characteristics
Known whether benefit recipient in household
Missing household income
Household income for last financial year
Time in household interviewing
Time in household unknown

Details of the final model are provided in Appendix 3 (Table A3.3).

The responding person weight was then multiplied by the inverse of the predicted probability of response.[8] That is:

$$w_{pers,w2} = w_{hh,w2} * 1 \Big/ \hat{p}_{persresp|hhresp,w2}$$

As a result, responding persons who were most like the non-respondents had their weights increased to a greater extent than those respondents who are least like the non-respondents. A minimum value for $\hat{p}_{persresp|hhresp,w2}$ was applied to avoid extreme weights. Respondents with a predicted probability of staying given their household responded of less than 0.5 had their response probability set to 0.5. This affected 154 respondents.

Non-Response Adjustment Using Data External to the HILDA Survey

The final step in the production of the cross-sectional person weights was to calibrate them to known benchmarks. Two sets of benchmarks were obtained from the ABS (as a special data service):

[8] The responding person weight for respondents in households with only one eligible adult were not adjusted (as they, by definition, responded if the household responded).

- Person benchmark 1:- Number of people by State, part of State, sex and age. For NSW, Vic, Qld, SA and WA, the part of State variable separated the metropolitan area from the rest of the State. For Tas, NT, and ACT, part of State was not used. The age categories used were:

    - 0-4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+ in NSW, Vic, Qld, Adelaide and Perth;

    - 0-4, 5-9, 10-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65+ in rural SA, rural WA and Tas;

    - 0-14, 15-34, 35+ in NT; and

    - 0-9, 10-14, 15-24, 25-34, 35-44, 45-54, 55+ in ACT.

- Person benchmark 2:- Number of people by labour force status and State. The labour force status included the following categories: under 15, employed, unemployed and not in the labour force. For NT and ACT, the unemployed and not in the labour force categories were collapsed.

The first set of person benchmarks related to the estimated number of residents in Australia as at 30 September 2002. The second set of person benchmarks were obtained from the Labour Force Survey, with an average calculated across four months from August to December 2001. These labour force benchmarks were proportionally adjusted so that the total number of people in each State matched the estimated residential population in the first set of person benchmarks. Both sets of benchmarks exclude people living in remote areas of NT and those living in non-private dwellings.

Only the first set of person benchmarks could be applied to the enumerated person weights. We are missing this information for non-respondents as the labour force status question was removed from the Household Form in wave 2, and therefore had to change our benchmarking practice from wave 1.[9]

Both sets of person benchmarks were used to calibrate the responding person benchmarks.

Appendix 3 provides some information on how the weights changed through the adjustment made.

[9] Note that in future waves, this question has been reintroduced to the Household Form so that it can be used in the benchmarking process.

Longitudinal Person Weights

Two longitudinal person weights have been provided and will be used in different circumstances by researchers. The most obvious longitudinal unit of analysis are persons responding in both waves 1 and 2 and a longitudinal responding person weight has been provided for this purpose. A second longitudinal weight has been provided for persons enumerated in both waves 1 and 2 (i.e. they were in responding households in both waves).

Initial Weights

The initial longitudinal weights are the corresponding person weights in wave 1. These are then adjusted for non-response and benchmarked as described below.

Non-Response Adjustment Using Data Internal to the HILDA Survey

The longitudinal responding person weight is adjusted for attrition between the two waves. A logistic model for the probability of responding in wave 2 given the person responded in wave 1 was developed. Deaths and moves overseas are treated as an acceptable 'response' along with interviews for a reason that will become apparent in the subsequent benchmarking step. The variables considered in the model include:

Wave 1 person characteristics
Female
Age
Age squared
Female aged 65 or over
Marital status
Ability in speaking English
Employment status and hours
Number of children the person has
Country of birth
Highest level of education achieved
Relationship in household
Health status
Likelihood of moving
Number of times moved in last 10 years
Length of PQ interview in wave 1
Length of PQ interview unknown
Whether completed SCQ in wave 1
Whether reference person in household

Wave 1 interview situation
Respondent's cooperation was fair, poor or very poor
Interview was assisted
English was a problem as it was a second language
Eyesight was a problem
Hearing was a problem
Other language problems occurred
Reading was a problem
Respondent was somewhat or very suspicious of interview

Respondent's understanding was fair, poor or very poor
Other adults influenced the interview

Wave 1 household characteristics
Location (State by part of State)
Remoteness area
SEIFA index of disadvantage
Dwelling type
Dwelling condition
Number of bedrooms per person in household
Number of calls made to household in wave 1
Whether household was partly responding in wave 1
Number of person in household in wave 1
Number of adults in household
Number of adults in household
Household type
Housing tenure
Known whether benefit recipient in household in wave 1
Missing household income
Household income for last financial year
Time in household interviewing in wave 1
Time in household unknown

Wave 2 household characteristics
Household split in wave 2
Whether moved between waves 1 and 2

The details of the model are provided in Appendix 2 (Table A2.3).

Readers seeking a discussion of the attrition experienced between waves 1 and 2 and the effect on the sample are directed to the technical paper on wave 2 data quality (Watson and Wooden, 2004b).

The initial longitudinal responding person weight is multiplied by the inverse of the person staying probability obtained from the above model. That is,

$$w_{resplong} = w_{resp,w1} * 1 \Big/ \hat{p}_{respstay,w2}$$

This means that people who are least likely to respond in wave 2 have their weight increased to a greater extent than those most likely to respond. A minimum value for $\hat{p}_{respstay,w2}$ was applied to avoid extreme weights. Respondents with a predicted probability of staying given their household responded of less than 0.3 had their response probability set to 0.3. This affected 38 respondents.

The longitudinal weight for enumerated persons is similarly adjusted, but, in this situation, the relevant staying probability is that of the household. Once the household responds in wave 2, all people within that household are enumerated (i.e. listed on the Household Form). The household staying probability is described earlier in this report as it was required for the construction of the cross section household weights. The longitudinal enumerated person weight is calculated as:

$$w_{enumlong} = w_{enum,w1} * 1 \Big/ \hat{p}_{hhstay,w2}$$

This means that people in households least likely to respond in wave 2 have their weight increased to a greater extent than those in households most likely to respond.

Non-Response Adjustment Using Data External to the HILDA Survey

To ensure the longitudinal weights do not diverge from expected population estimates a benchmarking step has been included in the creation of these weights. While we have not obtained population benchmarks for this longitudinal sample, we can use the wave 1 benchmarks and the wave 1 characteristics of these people with longitudinal weights in wave 2.[10]

The wave 1 person benchmarks used are the same in specification to the wave 2 person benchmarks but relate to 2001 rather than 2002. That is:

- Person benchmark 1:- Number of people by State, part of State, sex and age. For NSW, Vic, Qld, SA and WA, the part of State variable separated the metropolitan area from the rest of the State. For Tas, NT, and ACT, part of State was not used. The age categories used were:
    - 0-4, 5-9, 10-14, 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+ in NSW, Vic, Qld, Adelaide and Perth;
    - 0-4, 5-9, 10-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65+ in rural SA, rural WA and Tas;
    - 0-14, 15-34, 35+ in NT; and
    - 0-9, 10-14, 15-24, 25-34, 35-44, 45-54, 55+ in ACT.
- Person benchmark 2:- Number of people by labour force status and State. The labour force status included the following categories: under 15, employed, unemployed and not in the labour force. For NT and ACT, the unemployed and not in the labour force categories were collapsed.

The longitudinal weights for responding person and enumerated person were calibrated to these benchmarks based on their wave 1 characteristics.

To understand the impact of this benchmarking more fully, take the contrived example of 100 respondents aged 35 to 44 in wave 1, each with a weight of 1000. When we return to these respondents in wave 2, we find that 2 people had moved overseas, 88 are re-interviewed and 10 are non-respondents. They are now aged 36 to

[10] The population for which we would need longitudinal benchmarks would be all people living in Australia in 2001 (excluding remote parts of NT) in private dwellings who are still in Australia in 2002. This would exclude any new immigrants to Australia since 2001 and any Australians returning home after being overseas. Deaths and overseas moves would need to be excluded. People who were in private dwellings in 2001 who live in non-private dwellings in 2002 would be included. Obtaining benchmarks for this population would be very difficult (if not impossible).

45 with the wave 2 interview being approximately one year after the wave 1 interview. After making the response adjustments for the probability of responding in wave 2, the sum of the weights for the wave 2 respondents and people who moved out of scope turns out to be 95,000. We actually wanted the sum of the weights to be 100,000 as that is the number in the population from which we have taken a sample to track between waves 1 and 2. We therefore calibrate this longitudinal sample of 90 people to the wave 1 benchmarks which corrects the weights to sum to 100,000. We have allowed for the organic changes to the sample (such as people who have died or moved overseas) by their inclusion sample that is benchmarked. After calibration, the 88 respondents would have a weight of 1111, thus representing 97,778 people from the population of 100,000 who are still in scope. The 2 people who moved overseas would have a weight of 1111 and therefore represent 2222 people from the population that have moved out of scope.

Weights Provided in the Wave 2 Datasets

Table 1 provides a list of the weights provided on the wave 2 datasets together with a description of those weights. We have adopted the convention of adding the longitudinal weights only to the most recent wave undertaken.

Table 1: Weights provided in the wave 2 datasets

File	*Weights*	*Description*
Household File	bhhwth	The household weight is the cross-section population weight for all households responding in wave 2. Note the sum of these household weights is approximately 7.5 million.
	bhhwths	This is the cross-section household population weight rescaled to the sum of the sample size (i.e. 7245 responding households). Use this weight when the statistical package requires the weights to sum to the sample size.
	bhhwte01 to bhhwte14	The enumerated person weights are provided on both the household file and the enumerated person file. See description below.
Enumerated Person File	bhhwte	The enumerated person weight is the cross-section population weight for all people who are usual residents of the responding households in wave 2 (this includes children, non-respondents and respondents). The sum of these enumerated person weights is 19.2 million.
	blnwte	The longitudinal enumerated person weight is the longitudinal population weight for all people who were enumerated (i.e. in responding households) in both waves 1 and 2. This weight applies to children, non-respondents, intermittent respondents, and full respondents in responding households.
Responding Person File	bhhwtrp	The responding person weight is the cross-section population weight for all people who responded in wave 2 (i.e. they provided a personal interview). The sum of these responding person weights is 15.1 million.
	bhhwtrps	This is the cross-sectional responding person population weight rescaled to sum to the number of responding persons in wave 2 (i.e. 13,041). Use this weight when the statistical package requires the sum of the weights to be the sample size.
	blnwtr	The longitudinal responding person weight is the longitudinal population weight for all people responding (i.e. provided an interview) in both waves 1 and 2.

Some changes are expected to these weights with the next release. There are three reasons for this. Firstly, corrections may be made to age and sex variables when these are confirmed with individuals in subsequent wave interviews. Secondly, the benchmarks are updated from time to time. Thirdly, duplicate or excluded people in the sample may be identified after the release (this happens rarely).[11]

[11] The wave 1 weights in release 2.0 are different from those in release 1.0 for these reasons.

Advice on Using the Weights

Which Weight to Use

For some users, the array of weights on the dataset may, at first, seem confusing. This section provides examples of when it would be appropriate to use the different types of weights.

If you want to make inferences about the Australian population from frequencies or cross-tabulations of the HILDA sample then you will need to use weights. If you are only using information collected during the wave 2 interviews (either at the household level or person level) then you would use the wave 2 cross-section weights. Similarly, if you are only using wave 1 information, then you would use the wave 1 cross-section weights. If you want to infer how people have changed between waves 1 and 2, then you would use the longitudinal weights.

The following five examples show how the various weights may be used to answer questions about the population:

- *What proportion of households rent in 2002?* We would use the cross-section household weight for wave 2 and obtain a weighted estimate of proportion of households that were renting as at the time of interview.

- *How many people live in poor households in 2002?* We are interested in the number of individuals with a certain household characteristic, such as having low equivalised household incomes. We would use the cross-section enumerated person weight for wave 2 and count the number of enumerated people in households with poorest 10 per cent of equivalised household incomes. (We do not need to restrict our attention to responding persons only as total household incomes are available for all households after the imputation process. We also want to include children in this analysis and not just limit our analysis to those aged 15 year or older.)

- *What is the average salary of professionals in 2002?* This is a question that can only be answered from the responding person file using the cross-section responding person weight for wave 2. We would identify those reportedly working in professional occupations and take the weighted average of their wages and salaries.

- *How many people have moved out of the poorest 10 per cent of households between 2001 and 2002?* We might define the 'poorest' 10 per cent of households as having the lowest equivalised household incomes in each wave. We could then calculated how many people move out of the poorest decile between waves 1 and 2 by summing the longitudinal enumerated person weight for those people.

- *What proportion of people have changed their employment status between 2001 and 2002?* This question can only be answered by considering the responding persons in both waves. We would use the longitudinal responding person weight and construct a weighted cross-tabulation of the employment

status of respondents in wave 1 against the employment status of respondents in wave 2.

When constructing regression models, the researcher needs to be aware of the sample design and non-response issues underlying the data and will need to take account of this in some way.

Calculating Standard Errors

The statistical packages SAS and, until recently, SPSS, do not make it easy to appropriately treat complex survey data when constructing standard errors and confidence intervals. The HILDA survey has a complex survey design. It is:

- clustered – 488 areas were originally selected from which households were chosen and people are clustered within households;
- stratified – the 488 areas were selected from a frame of areas stratified by State and part of State; and
- unequally weighted – the households and individuals have unequal weights due to some irregularities in the selection of the sample in wave 1 and the non-random non-response in wave 1 and the non-random attrition in wave 2.

Some options available for the calculation of appropriate standard errors and confidence intervals include:

- Standard Error Tables – Based on the wave 1 data, approximate standard errors have been constructed for a range of estimates (see Horn, 2004). Similar tables for wave 2 have not as yet been produced.
- Use the recently released complex survey commands in SPSS (available in version 12).
- Use of 'svy' commands in Stata – The HILDA data can be readily transferred to the Stata package (using StatTransfer) which has a set of survey commands that deal with complex survey designs. Using the 'svyset' commands, the clustering, stratification and weights can be assigned. Various statistical procedures are available within the suite of 'svy' commands including means, proportions, tabulations, linear regression, logistic regression, probit models and a number of other commands.
- Use of GREGWT macro in SAS – Some users within FaCS and other organisations may have access to the GREGWT macro that can be used to construct various population estimates. The macro uses the jackknife method to estimate standard errors. For this procedure, replicate groups for the original sample are needed – these can be obtained from either Stephen Horn at FaCS or Nicole Watson at the Melbourne Institute.

An oversight in the production of the wave 2 files resulted in the area variable being excluded from the wave 2 files. To identify which of the 488 areas the wave 2 households are associated with, the user will need to match on the wave 1 household identifier from which the wave 2 household is derived and attach the appropriate area identifier. Any new entrants to the household should be assigned to the same area as the permanent sample member.

References

Horn, S, (2004), 'Guide to Standard Errors for Cross Section Estimates', HILDA Project Technical Paper Series No. 2/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Pannenberg, M, Pischner, R, Rendtel, U, Spiess, M, and Wagner, GG (2003) 'Sampling and Weighting', in JP Haisken-DeNew and JR Frick (eds), *Desktop Companion to the German Socio-Economic Panel Study (SOEP), Version 7.0*, DIW Berlin.

Stukel, D, Hidiroglou, MA and Särndal, CE, (1996), 'Variance estimation for calibration estimators: a comparison of Jackknifing versus Taylor linearization', *Survey Methodology*, vol. 22, no. 2, pp. 117-125.

Taylor, MF (ed), Brice, J, Buck, N, and Prentice-Lane, E, (2003) 'British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices', Colchester: University of Essex.

Watson, N, and Fry, TRL, (2002), 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 Weighting', HILDA Project Technical Paper Series No. 3/02, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N, and Wooden, M, (2004a), 'Assessing the Quality of the HILDA Survey Wave 2 Data', HILDA Project Technical Paper Series No. 5/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N, and Wooden, M, (2004b), 'Wave 2 Methodology', HILDA Project Technical Paper Series No. 1/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Appendix 1 – Technical Reference Group Membership

Mr Peter Boal
Employment Policy Group, Department of Employment and Workplace Relations

Mr James Chipperfield
Statistical Services Branch, Australian Bureau of Statistics

Dr John Henstridge
Data Analysis Australia

Mr Stephen Horn
Strategic Policy and Knowledge Branch, Department of Family and Community Services

Appendix 2 – Models for Predicting Response to the HILDA Survey

Table A2.1: Linear regression model of wave 1 quasi-selection probability (adjusted R^2=0.27)

Variable	*Estimate*	*Standard Error*	*P-value*
Intercept	-7.1303	0.0459	<.0001
Wave 2 household characteristics			
Composition and location			
2 adults 1 child in Sydney	0.1646	0.0454	0.0003
2 adults 2+ children in Sydney	0.2638	0.0403	<.0001
3+ adults 0 children in Sydney	-0.0605	0.0288	0.0356
3+ adults 1 child in Sydney	-0.0307	0.0526	0.5593
3+ adults 2+ child in Sydney	0.2667	0.0537	<.0001
1 adults 0 children in Sydney	-0.0479	0.1758	0.7854
1 adults 1 child in Sydney	0.1166	0.0683	0.0880
1 adults 2+ children in Sydney	-0.0793	0.0665	0.2332
2 adults 0 children in Melbourne	0.1614	0.0185	<.0001
2 adults 1 child in Melbourne	0.2623	0.0456	<.0001
2 adults 2+ children in Melbourne	0.2458	0.0405	<.0001
3+ adults 0 children in Melbourne	0.1195	0.0284	<.0001
3+ adults 1 child in Melbourne	0.1460	0.0471	0.0019
3+ adults 2+ child in Melbourne	0.3099	0.0566	<.0001
1 adults 0 children in Melbourne	-0.0551	0.1759	0.7540
1 adults 1 child in Melbourne	0.2030	0.0656	0.0020
1 adults 2+ children in Melbourne	0.3076	0.0643	<.0001
2 adults 0 children in Brisbane	0.1295	0.0221	<.0001
2 adults 1 child in Brisbane	0.1915	0.0499	0.0001
2 adults 2+ children in Brisbane	0.4480	0.0441	<.0001
3+ adults 0 children in Brisbane	0.0138	0.0360	0.7012
3+ adults 1 child in Brisbane	0.2999	0.0585	<.0001
3+ adults 2+ child in Brisbane	0.1190	0.0641	0.0634
1 adults 0 children in Brisbane	-0.1462	0.1764	0.4072
1 adults 1 child in Brisbane	0.4602	0.0806	<.0001
1 adults 2+ children in Brisbane	0.2105	0.0806	0.0091
2 adults 0 children in ACT & Rural NSW	0.0882	0.0285	0.0020
2 adults 1 child in ACT & Rural NSW	0.1400	0.0529	0.0081
2 adults 2+ children in ACT & Rural NSW	0.3383	0.0461	<.0001
3+ adults 0 children in ACT & Rural NSW	0.0121	0.0391	0.7582
3+ adults 1 child in ACT & Rural NSW	0.1398	0.0553	0.0115
3+ adults 2+ child in ACT & Rural NSW	0.2121	0.0580	0.0003
1 adults 0 children in ACT & Rural NSW	-0.1099	0.1773	0.5354
1 adults 1 child in ACT & Rural NSW	0.2723	0.0642	<.0001
1 adults 2+ children in ACT & Rural NSW	0.0937	0.0687	0.1729
2 adults 0 children in WA & SA	0.1890	0.0198	<.0001
2 adults 1 child in WA & SA	0.3013	0.0465	<.0001
2 adults 2+ children in WA & SA	0.2719	0.0407	<.0001
3+ adults 0 children in WA & SA	0.0517	0.0302	0.0862
3+ adults 1 child in WA & SA	0.0648	0.0511	0.2051
3+ adults 2+ child in WA & SA	0.1564	0.0556	0.0050
1 adults 0 children in WA & SA	-0.0578	0.1759	0.7426
1 adults 1 child in WA & SA	0.3755	0.0553	<.0001

(Table A2.1 c'td)

Variable	*Estimate*	*Standard Error*	*P-value*
1 adults 2+ children in WA & SA	0.1286	0.0618	0.0375
2 adults 0 children in Tas & Rural Vic	0.3403	0.0256	<.0001
2 adults 1 child in Tas & Rural Vic	0.2449	0.0550	<.0001
2 adults 2+ children in Tas & Rural Vic	0.5057	0.0438	<.0001
3+ adults 0 children in Tas & Rural Vic	0.1527	0.0369	<.0001
3+ adults 1 child in Tas & Rural Vic	0.4753	0.0644	<.0001
3+ adults 2+ child in Tas & Rural Vic	0.3620	0.0674	<.0001
1 adults 0 children in Tas & Rural Vic	0.0178	0.1748	0.9189
1 adults 1 child in Tas & Rural Vic	0.2121	0.0760	0.0053
1 adults 2+ children in Tas & Rural Vic	0.5619	0.0701	<.0001
2 adults 0 children in NT & Rural Qld	-0.1573	0.0441	0.0004
2 adults 1 child in NT & Rural Qld	0.0411	0.0624	0.5102
2 adults 2+ children in NT & Rural Qld	0.1006	0.0583	0.0844
3+ adults 0 children in NT & Rural Qld	-0.1554	0.0531	0.0034
3+ adults 1 child in NT & Rural Qld	-0.0720	0.0699	0.3025
3+ adults 2+ child in NT & Rural Qld	-0.0773	0.0707	0.2744
1 adults 0 children in NT & Rural Qld	-0.2842	0.1804	0.1151
1 adults 1 child in NT & Rural Qld	0.1783	0.0760	0.0189
1 adults 2+ children in NT & Rural Qld	0.0758	0.0756	0.3159
Rural NSW	0.1287	0.0246	<.0001
Rural Vic	-0.0415	0.0197	0.0351
Rural Qld	0.3088	0.0408	<.0001
Adelaide	0.0238	0.0157	0.1288
Rural SA	0.3284	0.0202	<.0001
Rural WA	0.0573	0.0205	0.0051
Dwelling type (base category separate house)			
Semi-detached	-0.0009	0.0110	0.9366
Apartment less than 3 storeys	-0.0182	0.0114	0.1118
Apartment 3 storeys or more	-0.1612	0.0165	<.0001
Other dwelling	0.0069	0.0405	0.8650
Type unknown	-0.0652	0.0367	0.0757
Dwelling condition			
Good	0.0342	0.0072	<.0001
Average	0.0552	0.0081	<.0001
Poor	0.0538	0.0145	0.0002
Very poor/almost derelict	0.0601	0.0399	0.1322
Condition unknown	0.0699	0.0321	0.0294
New baby born to HH	-0.0383	0.0184	0.0370
Other joiner to HH	0.0245	0.0129	0.0569
W1 HH member died or moved overseas	-0.0580	0.0255	0.0227
W1 HH member left	-0.1541	0.1731	0.3734
HH split in wave 2	0.1066	0.1732	0.5383
HH merged in wave 2 (with other wave 1 HH)	-0.1061	0.0870	0.2226
HH income for last financial year	-0.00000006	0.00000006	0.3275
Missing HH income	0.0053	0.0080	0.5082

(Table A2.1 c'td)

Variable	*Estimate*	*Standard Error*	*P-value*
Wave 2 reference person characteristics			
Female	0.0213	0.0073	0.0036
Female aged 65 or over	-0.0189	0.0151	0.2117
Age group (base category 15-19)			
20-24	0.0087	0.0263	0.7396
25-34	0.0266	0.0253	0.2947
35-44	0.0627	0.0259	0.0154
45-54	0.0753	0.0263	0.0042
55-64	0.0734	0.0271	0.0069
65+	0.0890	0.0288	0.0020
Marital status (base category married)			
De facto	0.0172	0.0116	0.1387
Separated	0.0449	0.0288	0.1194
Divorced	0.0177	0.0281	0.5283
Widowed	-0.0018	0.0287	0.9504
Never married	0.0297	0.0279	0.2867
Relationship in household (base category couple with child under 15)			
Couple with dependent student	-0.0164	0.0360	0.6479
Couple with non-dependent child	-0.0353	0.0352	0.3168
Couple without children	-0.0128	0.0340	0.7070
Lone parent with child under 15	-0.0487	0.0335	0.1461
Lone parent with dependent child	-0.0214	0.0495	0.6652
Lone parent with non-dependent child	-0.0370	0.0452	0.4127
Dependent student	-0.1608	0.1076	0.1351
Non-dependent child	0.1212	0.0539	0.0247
Other family member	-0.0028	0.0448	0.9500
Lone person	0.2412	0.1789	0.1775
Unrelated to all HH members	0.0051	0.0464	0.9120
Country of birth (base category Australia)			
Main English speaking country	-0.0091	0.0094	0.3283
Main non-English speaking country	-0.0298	0.0124	0.0164
Ability in speaking English (base category English only language spoken)			
Speaks English well or very well	-0.0092	0.0138	0.5070
Speaks English not well	-0.0143	0.0282	0.6114
Speaks English not at all	-0.1157	0.0691	0.0942
Highest level of education achieved (base category yr12 or below)			
Certificate or diploma	0.0091	0.0068	0.1821
Bachelor or post-graduate	-0.0138	0.0086	0.1086
Number of children respondent has	0.0040	0.0025	0.1081
Employment status (base category employed)			
Unemployed	-0.0155	0.0186	0.4039
Not in the labour force	0.0042	0.0120	0.7228
Usual hours worked	0.0001	0.0002	0.7429

Table A2.2: Logistic regression model of household responding in wave 2

Variable	*Estimate*	*Standard Error*	*P-value*
Intercept	2.9594	0.8276	0.0003
Wave 1 household characteristics			
Location (base category Sydney)			
Rural NSW	0.0207	0.1641	0.8997
Melbourne	-0.2267	0.1207	0.0603
Rural Vic	-0.0798	0.1961	0.6842
Brisbane	-0.0867	0.1556	0.5775
Rural Qld	0.3781	0.1777	0.0334
Adelaide	0.0712	0.1797	0.6921
Rural SA	-0.0283	0.2739	0.9178
Perth	-0.0652	0.1604	0.6844
Rural WA	-0.0131	0.2938	0.9643
Tas	-0.0334	0.2629	0.8990
NT	0.8588	0.5739	0.1346
ACT	0.3491	0.3728	0.3491
Remoteness area (base category major cities)			
Inner regional	-0.1286	0.1259	0.3069
Outer regional	-0.1726	0.1719	0.3153
Remote	-0.4844	0.3558	0.1734
SEIFA index of disadvantage (base category is lowest decile – most disadvantaged)			
Second decile	-0.3202	0.1674	0.0557
Third decile	-0.4753	0.1689	0.0049
Fourth decile	-0.3731	0.1732	0.0312
Fifth decile	-0.5048	0.1733	0.0036
Sixth decile	-0.0288	0.1843	0.8757
Seventh decile	-0.5100	0.1787	0.0043
Eighth decile	-0.1314	0.1782	0.4610
Ninth decile	-0.3348	0.1845	0.0695
Tenth decile (least disadvantaged)	-0.2695	0.1935	0.1636
Dwelling type			
Semi-detached	-0.1234	0.1309	0.3457
Apartment less than 3 storeys	-0.2511	0.1487	0.0913
Apartment 3 storeys or more	-0.5044	0.1971	0.0105
Other dwelling	0.1561	0.5133	0.7610
Dwelling type unknown	-0.5378	1.1092	0.6278
Dwelling type (base category separate house)			
Good	-0.1264	0.0921	0.1701
Average	0.0174	0.1032	0.8660
Poor	-0.1402	0.1675	0.4025
Very poor/almost derelict	-0.2527	0.5002	0.6135
Condition unknown	-0.2932	1.2265	0.8111
Number of bedrooms per person	-0.2079	0.0841	0.0135
Number of calls made to HH	-0.0483	0.0140	0.0006
Whether HH was partly responding	-1.1931	0.1547	<.0001
Number of person in HH	-0.2218	0.0671	0.0009

(Table A2.2 c'td)

Variable	*Estimate*	*Standard Error*	*P-value*
Number of adults in HH (base category two adults)			
One adult in HH	-0.2737	0.2844	0.3359
Three or more adults in HH	-0.3654	0.1539	0.0176
Number of children in HH (base category no children)			
One child in HH	-0.7265	0.3872	0.0606
Two or more children in HH	-0.2831	0.4014	0.4806
Household type (base category couple without children)			
Couple with children under 15	0.3203	0.6107	0.5999
Couple with dependent student	-0.1082	1.2694	0.9320
Couple with non-dependent child	0.2254	0.8089	0.7805
Lone parent with children under 15	0.7096	0.8443	0.4007
Lone parent with dependent child	1.2759	1.2760	0.3173
Lone parent with non-dependent child	0.2430	0.9562	0.7994
Multifamily HH	0.5123	0.4760	0.2818
Housing tenure (base category own)			
Rent	-0.2180	0.1042	0.0364
Rent-buy	0.3760	0.7952	0.6363
Rent free	0.0467	0.2527	0.8534
Know whether have benefit recipient in HH	-0.1499	0.1004	0.1354
HH income for last financial year	0.0000005	0.000001	0.7065
Missing HH income	-0.0861	0.1225	0.4820
Time in HH interviewing	0.00491	0.00217	0.0234
Time in HH unknown	0.8938	0.2674	0.0008
Wave 2 household characteristics			
HH split in wave 2	-0.4016	0.1024	<.0001
Whether moved between w1 and w2	-0.7051	0.0990	<.0001
Wave 1 reference person characteristics			
Female	0.1266	0.0893	0.1563
Age	0.0631	0.0153	<.0001
Age squared	-0.00060	0.000152	<.0001
Female aged 65 or over	0.0142	0.2043	0.9446
Marital status (base category married)			
De facto	-0.1747	0.1286	0.1742
Separated	0.3702	0.3106	0.2333
Divorced	0.5199	0.2978	0.0808
Widowed	0.6549	0.3211	0.0414
Never married	0.4019	0.2992	0.1792
Ability in speaking English (base category English only language spoken)			
Speaks English well or very well	-0.2717	0.1545	0.0786
Speaks English not well	-0.6056	0.2551	0.0176
Speaks English not at all	0.3231	0.7533	0.6679

(Table A2.2 c'td)

Variable	*Estimate*	*Standard Error*	*P-value*
Employment status and hours (base category not in labour force)			
Unemployed	-0.3743	0.1716	0.0292
Employed less than 25 hrs pw	0.1031	0.1432	0.4716
Employed 25 to 34 hrs pw	-0.1062	0.1692	0.5301
Employed 35 to 44 hrs pw	-0.2201	0.1256	0.0797
Employed 45 to 54 hrs pw	0.1140	0.1562	0.4655
Employed 55 or more hrs pw	-0.0986	0.1665	0.5539
Number of children have	0.0199	0.0329	0.5445
Country of birth (base category Australia)			
Main English speaking country	-0.2039	0.1191	0.0869
Main non-English speaking country	-0.1608	0.1434	0.2622
Highest level of education achieved (base category yr12 or below)			
Certificate or diploma	0.0878	0.0815	0.2814
Bachelor or post-graduate	0.7063	0.1215	<.0001
Relationship in household			
Couple with dependent student	-0.1555	1.2186	0.8985
Couple with non-dependent child	-0.3726	0.7369	0.6131
Couple without children	-0.3946	0.6146	0.5208
Lone parent with child under 15	-0.8747	0.8023	0.2756
Lone parent with dependent child	-2.0050	1.2944	0.1214
Lone parent with non-dependent child	-1.1780	0.9607	0.2201
Other family member	-0.3614	0.7237	0.6175
Lone person	-0.2104	0.7258	0.7719
Unrelated to all HH members	-0.7214	0.6908	0.2964
Health status (base category excellent)			
Very Good	0.0473	0.1046	0.6515
Good	0.0935	0.1103	0.3969
Fair	0.2060	0.1393	0.1392
Poor	-0.1533	0.1866	0.4113
Likelihood of moving (base category not likely to move)			
Not sure if moving	-0.0280	0.1227	0.8195
Likely or very likely to move	0.1960	0.1105	0.0759
Number of times moved in last 10 yrs (base category no moves)			
1 to 2 times	-0.2058	0.1119	0.0658
3 to 4 times	-0.1496	0.1200	0.2126
5 to 9 times	0.00870	0.1368	0.9493
10 or more	-0.3202	0.1757	0.0684
Unknown number	-0.5764	1.2101	0.6338
Length of PQ ivw in w1	-0.00827	0.00403	0.0400
Length of PQ ivw unknown	-0.1710	0.2939	0.5607
Whether completed SCQ in W1	-0.7242	0.1166	<.0001
Whether reference person provided PQ ivw in w1	-1.0169	0.4563	0.0258

Table A2.3: Logistic regression model of person responding in wave 2, given household responded

Variable	*Estimate*	*Standard Error*	*P-value*
Intercept	2.5100	0.6020	<.0001
Wave 2 person characteristics			
Female	0.5894	0.0847	<.0001
Age	0.0315	0.0143	0.0283
Age squared	-0.00048	0.000146	0.0011
Female aged 65 or over	-0.5032	0.2097	0.0164
Relationship in household (base category couple with children under 15)			
Couple with dependent student	-0.6077	0.3516	0.0840
Couple with non-dependent child	0.9445	0.2457	0.0001
Couple without children	0.4944	0.3269	0.1305
Lone parent with child under 15	0.3974	0.4487	0.3759
Lone parent with dependent child	1.2843	1.1057	0.2454
Lone parent with non-dependent child	0.8752	0.3665	0.0169
Non-dependent child	0.0772	0.2322	0.7395
Other family member	-0.6518	0.3135	0.0376
Unrelated to all HH members	-0.3697	0.4285	0.3884
Wave 2 household characteristics			
Location (base category Sydney)			
Rural NSW	0.4608	0.1802	0.0106
Melbourne	0.1737	0.1213	0.1522
Rural Vic	-0.0583	0.2256	0.7959
Brisbane	0.6120	0.1674	0.0003
Rural Qld	0.2664	0.2001	0.1832
Adelaide	0.1827	0.1955	0.3502
Rural SA	0.7027	0.3396	0.0385
Perth	-0.1008	0.1626	0.5356
Rural WA	-0.1967	0.2950	0.5049
Tas	0.1409	0.2993	0.6379
NT	0.0128	0.6637	0.9846
ACT	0.3575	0.2943	0.2246
Remoteness Area (base category major cities)			
Inner regional	0.2650	0.1649	0.1079
Outer regional	-0.0148	0.1944	0.9391
Remote	-0.0596	0.2883	0.8361
SEIFA index of disadvantage (base category is lowest decile – most disadvantaged)			
Second decile	-0.1016	0.1882	0.5894
Third decile	0.0556	0.1958	0.7766
Fourth decile	0.2055	0.1972	0.2973
Fifth decile	0.4403	0.2054	0.0320
Sixth decile	0.1290	0.1978	0.5142
Seventh decile	-0.0468	0.1907	0.8062
Eighth decile	0.1313	0.1882	0.4852
Ninth decile	-0.0329	0.1937	0.8650
Tenth decile (least disadvantaged)	0.1508	0.1932	0.4353

(**Table A2.3 c'td)**

Variable	*Estimate*	*Standard Error*	*P-value*
Dwelling type (base category separate house)			
Semi-detached	-0.0259	0.1779	0.8842
Apartment less than 3 storeys	0.0385	0.2461	0.8757
Apartment 3 storeys or more	-0.0656	0.2563	0.7980
Dwelling unknown	-0.5081	0.3909	0.1937
Non-private dwelling	0.3256	0.4465	0.4658
Dwelling condition (base category excellent)			
Good	-0.0753	0.0945	0.4252
Average	-0.0276	0.1137	0.8081
Poor	-0.00185	0.2024	0.9927
Very poor/almost derelict	-0.7631	0.4317	0.0771
Condition unknown	-0.3359	0.3521	0.3401
Number of bedrooms per person in HH	0.1348	0.1467	0.3582
Number of calls made to HH	-0.1891	0.0121	<.0001
Number of person in HH	-0.2193	0.0585	0.0002
Three or more adults in HH in w2	-0.3991	0.1425	0.0051
Number of children in HH (base category zero)			
One	-0.1456	0.3397	0.6681
Two or more	0.0910	0.3433	0.7910
Household type (base category couple without children)			
Couple with children under 15	0.7604	0.4186	0.0693
Couple with dependent student	1.2584	0.3882	0.0012
Couple with non-dependent child	-1.0908	0.3281	0.0009
Lone parent with children under 15	0.5478	0.4774	0.2511
Lone parent with dependent child	0.4220	0.5238	0.4204
Lone parent with non-dependent child	-0.8754	0.3568	0.0141
Other related	0.6504	0.4111	0.1136
Group household	0.3756	0.4335	0.3863
Multifamily HH	-0.2018	0.3499	0.5641
Housing tenure (base category own/rent-buy)			
Rent	0.1144	0.1188	0.3354
Rent-free	-0.1016	0.2476	0.6815
HH split in wave 2	-1.4185	0.4657	0.0023
Whether moved between w1 and w2	0.1157	0.1250	0.3547
Joiners to HH	-0.2160	0.1248	0.0835
Leavers from HH	1.0888	0.4593	0.0178
Both joiners and leavers	0.4534	0.4818	0.3467
Wave 1 household characteristics			
Known whether benefit recipient in HH	0.1830	0.0957	0.0558
Missing hh income	-1.1054	0.1236	<.0001
HH income for last financial year	0.000002	0.000001	0.2401
Time in HH interviewing	0.0160	0.00143	<.0001
Time in HH unknown	1.0206	0.1694	<.0001

Table A2.4: Logistic regression model of wave 2 response for previous wave 1 respondents

Variable	*Estimate*	*Standard Error*	*P-value*
Intercept	2.0709	0.4337	<.0001
Wave 1 person characteristics			
Female	0.0427	0.0632	0.4988
Age	0.0835	0.0112	<.0001
Age squared	-0.00078	0.000115	<.0001
Female aged 65 or over	-0.0315	0.1488	0.8324
Marital status (base category married)			
De facto	-0.1713	0.0999	0.0865
Separated	0.3155	0.2243	0.1595
Divorced	0.2344	0.1970	0.2342
Widowed	0.6829	0.2412	0.0046
Never married	0.2820	0.1669	0.0912
Ability in speaking English (base category English only language spoken)			
Speaks English well or very well	-0.1630	0.1098	0.1378
Speaks English not well	-0.2339	0.2144	0.2752
Speaks English not at all	-0.4922	0.4579	0.2824
Employment status and hours (base category not in labour force)			
Unemployed	-0.1346	0.1291	0.2970
Employed less than 25 hrs pw	0.0584	0.0969	0.5466
Employed 25 to 34 hrs pw	-0.2424	0.1267	0.0558
Employed 35 to 44 hrs pw	-0.3080	0.0904	0.0007
Employed 45 to 54 hrs pw	-0.1255	0.1143	0.2726
Employed 55 or more hrs pw	-0.3366	0.1207	0.0053
Number of children have	-0.0140	0.0257	0.5846
Country of birth (base category Australia)			
Main English speaking country	-0.2574	0.0893	0.0040
Main non-English speaking country	-0.2141	0.1055	0.0424
Highest level of education achieved (base category yr12 or below)			
Certificate or diploma	0.1320	0.0630	0.0361
Bachelor or post-graduate	0.5462	0.0907	<.0001
Relationship in household (base category couple with child under 15)			
Couple with dependent student	-0.2954	0.2204	0.1801
Couple with non-dependent child	-0.1092	0.2211	0.6214
Couple without children	-0.1730	0.2437	0.4779
Lone parent with child under 15	-0.3477	0.2959	0.2400
Lone parent with dependent child	-0.4480	0.4254	0.2923
Lone parent with non-dependent child	0.0644	0.3167	0.8389
Other family member	-0.7738	0.2300	0.0008
Lone person	-0.1936	0.3973	0.6261
Unrelated to all HH members	-0.7692	0.2534	0.0024

(**Table A2.4 c'td)**

Variable	*Estimate*	*Standard Error*	*P-value*
Health status (base category excellent)			
Very Good	0.1145	0.0741	0.1220
Good	0.0448	0.0779	0.5653
Fair	0.0842	0.1021	0.4096
Poor	-0.1847	0.1458	0.2053
Likelihood of moving (base category not likely to move)			
Not sure if moving	-0.0294	0.0903	0.7447
Likely or very likely to move	0.0774	0.0801	0.3334
Number of times moved in last 10 yrs (base category no moves)			
Moved 1 to 2 times in last 10 yrs	-0.1118	0.0783	0.1535
Moved 3 to 4 times in last 10 yrs	0.0277	0.0854	0.7455
Moved 5 to 9 times in last 10 yrs	0.0214	0.0966	0.8248
Moved 10 or more times in last 10 yrs	-0.2501	0.1345	0.0630
Moved unknown number of times in last 10 yrs	-0.0173	0.9642	0.9856
Length of PQ ivw in w1	-0.00656	0.00309	0.0339
Length of PQ ivw unknown	-0.1872	0.2253	0.4061
Whether completed SCQ in W1	-0.7718	0.0885	<.0001
Whether reference person in HH	0.0198	0.0682	0.7715
Wave 1 interview situation			
Respondent's cooperation was fair, poor or very poor	-0.7281	0.1523	<.0001
Interview was assisted	-0.0114	0.0786	0.8843
English was a problem as it was a second language	-0.1386	0.1468	0.3451
Eyesight was a problem	0.2706	0.3218	0.4004
Hearing was a problem	-0.0167	0.2289	0.9420
Other language problems occurred	-0.2989	0.2842	0.2929
Reading was a problem	-0.0281	0.1961	0.8860
Respondent was somewhat or very suspicious of interview	-0.4683	0.1129	<.0001
Respondent's understanding was fair, poor or very poor	-0.0913	0.1220	0.4544
Other adults influenced the interview	-0.0743	0.0560	0.1846
Wave 1 household characteristics			
Location (base category Sydney)			
Rural NSW	0.1434	0.1206	0.2343
Melbourne	0.0819	0.0878	0.3513
Rural Vic	-0.0369	0.1455	0.7997
Brisbane	0.2333	0.1179	0.0479
Rural Qld	0.4310	0.1355	0.0015
Adelaide	0.3538	0.1349	0.0087
Rural SA	0.2153	0.2066	0.2973
Perth	0.2333	0.1207	0.0532
Rural WA	0.1156	0.2207	0.6004
Tas	-0.2090	0.1868	0.2632
NT	1.5739	0.5681	0.0056
ACT	0.2429	0.2389	0.3093
Remoteness Area (base category major cities)			
Inner regional	0.0366	0.0957	0.7018
Outer regional	-0.0193	0.1316	0.8835
Remote	-0.5992	0.2577	0.0201

(**Table A2.4 c'td)**

Variable	*Estimate*	*Standard Error*	*P-value*
SEIFA index of disadvantage (base category is lowest decile – most disadvantaged)			
Second decile	-0.3343	0.1317	0.0112
Third decile	-0.5572	0.1330	<.0001
Fourth decile	-0.3326	0.1354	0.0140
Fifth decile	-0.4322	0.1345	0.0013
Sixth decile	-0.1301	0.1418	0.3590
Seventh decile	-0.5971	0.1364	<.0001
Eighth decile	-0.1787	0.1369	0.1917
Ninth decile	-0.4247	0.1397	0.0024
Tenth decile (least disadvantaged)	-0.1921	0.1466	0.1900
Dwelling type (base category separate house)			
Semi-detached	0.00339	0.1068	0.9747
Apartment less than 3 storeys	-0.3266	0.1184	0.0058
Apartment 3 storeys or more	-0.4155	0.1575	0.0083
Other dwelling	0.0981	0.4317	0.8202
Dwelling unknown	-0.0103	1.1126	0.9926
Dwelling condition (base category excellent)			
Good	-0.0740	0.0663	0.2642
Average	0.0834	0.0761	0.2733
Poor	0.0388	0.1381	0.7786
Very poor/almost derelict	-0.1081	0.4155	0.7948
Condition unknown	-1.5467	0.7462	0.0382
Number of bedrooms per person in HH	-0.2419	0.0716	0.0007
Number of calls made to HH in w1	-0.0622	0.0105	<.0001
Whether HH was partly responding in W1	-0.8166	0.1142	<.0001
Number of person in HH in w1	-0.1452	0.0491	0.0031
Number of adults in HH (base category two adults)			
One adult in HH in w1	0.1727	0.2914	0.5533
Three or more adults in HH in w1	-0.3369	0.1120	0.0026
Number of adults in HH (base category zero children)			
One child in HH in w1	-0.4663	0.3406	0.1711
Two or more children in HH in w1	-0.1630	0.3607	0.6512
Household type (base category couple without children)			
Couple with children under 15	0.1075	0.3689	0.7708
Couple with dependent student	-0.1324	0.2698	0.6236
Couple with non-dependent child	-0.2312	0.2729	0.3969
Lone parent with children under 15	0.0436	0.3813	0.9089
Lone parent with dependent child	-0.0899	0.3316	0.7864
Lone parent with non-dependent child	-0.5745	0.2895	0.0472
Multifamily HH	-0.1054	0.3201	0.7419
Housing tenure (base category own/rent-buy)			
Rent	-0.2016	0.0794	0.0111
Rent-buy	1.0652	0.7567	0.1592
Rent free	0.2003	0.2051	0.3287
Known whether benefit recipient in HH in w1	-0.1961	0.0699	0.0051
Missing HH income	-0.0802	0.0845	0.3422
HH income for last financial year	-0.0000001	0.0000008	0.8636

(**Table A2.4 c'td)**

Variable	*Estimate*	*Standard Error*	*P-value*
Time in HH interviewing in W1	0.00329	0.00143	0.0212
Time in HH unknown	0.8717	0.1989	<.0001
Wave 2 household characteristics			
HH split in wave 2	0.5032	0.0924	<.0001
Whether moved between w1 and w2	-0.6536	0.0784	<.0001

Appendix 3 – Effect of Adjustment on Weights

Table A3.1: Distribution of the Weights

	Mean	Min	Q1	Median	Q3	Max
Cross-section household-level						
Initial household weight	1043	176	872	991	1140	4101
Revised for new entrants	1043	183	857	1013	1167	4260
Adjusted for probability of response	1043	165	818	963	1146	8300
Adjusted to benchmarks	1043	157	796	958	1168	8360
Cross-section enumerated person-level						
Initial enumerated person weight	1047	155	779	956	1196	8273
Adjusted to benchmarks	1047	130	758	957	1209	7711
Cross-section responding person-level						
Initial responding person weight	1179	175	894	1096	1347	9336
Adjusted for probability of response	1179	168	856	1053	1317	13413
Adjusted to benchmarks	1179	112	854	1079	1361	8991
Longitudinal enumerated person-level						
Initial longitudinal enumerated weight	1100	151	869	1046	1249	4517
Adjusted for probability of enumeration	1100	135	819	997	1242	7179
Adjusted to benchmarks	1100	136	816	998	1251	7410
Longitudinal responding person-level						
Initial longitudinal responding weight	1250	156	969	1162	1419	5906
Adjusted for probability of response	1250	136	914	1109	1409	9995
Adjusted to benchmarks	1250	128	911	1111	1413	10355

Table A3.1: Distribution of Percentage change in the Weights

	Mean	Min	Q1	Median	Q3	Max
Cross-section household-level						
Revised for new entrants	0.3	-81.6	3.9	3.9	3.9	3.9
Adjusted for probability of response	-0.5	-79.9	-5.9	-3.3	2.0	199.9
Adjusted to benchmarks	0.5	-78.8	-9.4	-3.4	5.5	232.4
Cross-section enumerated person-level						
Adjusted to benchmarks	0.1	-35.8	-7.5	-0.2	6.4	64.5
Cross-section responding person-level						
Adjusted for probability of response	-1.1	-7.5	-6.8	-5.5	-2.1	85.0
Adjusted to benchmarks	-0.1	-58.9	-10.7	-1.5	9.6	134.5
Longitudinal enumerated person-level						
Adjusted for probability of enumeration	-0.8	-13.5	-9.7	-7.0	-1.1	185.8
Adjusted to benchmarks	-0.9	-33.6	-10.9	-6.3	1.2	257.8
Longitudinal responding person-level						
Adjusted for probability of response	-1.3	-13.8	-9.3	-6.2	-0.3	184.6
Adjusted to benchmarks	-1.3	-28.9	-10.0	-5.6	1.4	222.3

Notes: The percentage change is calculated as 100 * (new weight – initial weight)/ initial weight. The initial weight has been adjusted to sum to the total of the relevant population.