



FACULTY OF
BUSINESS &
ECONOMICS

HILDA PROJECT TECHNICAL PAPER SERIES No. 1/11, September 2011

Methodology for the HILDA top-up sample

Nicole Watson

**The HILDA Project was initiated, and is funded, by the
Australian Government Department of Families, Housing,
Community Services and Indigenous Affairs**



Acknowledgements

Thanks to Athina Katiforis at Roy Morgan Research for providing comments and suggestions regarding the timing and implementation of the fieldwork.

Contents

INTRODUCTION	1
SURVEY INSTRUMENTS	1
MULTI-STAGE SAMPLE DESIGN	2
<i>STAGE 1: SELECTION OF AREAS</i>	2
<i>STAGE 2: SELECTION OF DWELLINGS</i>	3
<i>STAGE 3: SELECTION OF HOUSEHOLDS</i>	4
FIELDWORK	4
<i>RESPONDENT COMMUNICATION</i>	4
<i>INTERVIEWERS</i>	5
<i>INTERVIEWER TRAINING AND FIELDWORK SCHEDULE</i>	5
<i>OTHER FIELDWORK PROCEDURES</i>	7
<i>EXPECTED RESPONSE RATES</i>	7
DATA PREPARATION	8
REFERENCES	8

Introduction

The original sample for the Household, Income and Labour Dynamics in Australia (HILDA) Survey was selected in 2001 and since that time the population has evolved in a number of ways that the sample cannot emulate. For example, the population now includes i) immigrants permanently settling in Australia since 2001; ii) long-term visitors arriving since 2001; iii) Australians not in Australia in 2001 who have since returned from overseas; and iv) Australian-born children of the first three groups.¹ Watson (2006) estimates that the first three groups represent 6.6 per cent of the Australian population by 2011. Without a top-up sample, the HILDA sample cannot represent this significant (and growing) portion of the population.

We canvassed a number of options for a top-up to the HILDA sample (see Watson, 2006) and ultimately it was decided that a general top-up sample of 2000 responding households would be added to the sample in 2011. A general top-up sample will not only allow for the new portion of the population to be represented, but it will also increase the sample size for some analyses going forward and will also permit the study of the impact of non-response and attrition on our ongoing sample.

This paper describes the methodology employed for the top-up sample.

Survey instruments

Interviews with the top-up sample members are conducted via computer-assisted personal interviewing (CAPI), as has been the practice with the ongoing sample since 2009.² The CAPI script for the ongoing sample was amended to incorporate new households that have not been interviewed before.

The main changes to the CAPI script relate to the Household Form where the basic details of the household members are captured. In addition to the core information collected in the Household Form for the ongoing sample, two further questions are included in the initial wave of the top-up sample (these being whether English is the only language spoken at home and how well English is spoken). The following information is therefore collected about each member of the household:

- first and last name;
- sex;
- date of birth;
- whether English is the only language spoken at home;
- how well English is spoken (for those aged 5 and over);
- employment status (for those aged 15 and over);
- presence of a long term health condition; and
- relationships between household members.

¹ Also, given the wave 1 sample was restricted to people living in private dwellings in non-remote areas of Australia, there may also be a small portion of the population who have now moved out of institutions or remote areas into private dwellings in non-remote areas. In total, it is estimated that those in non-private dwellings and in remote parts of Australia constitute approximately 0.4 per cent of the 2001 Australian population and a small fraction of these may now live in private dwellings in non-remote areas of Australia.

² CAPI was introduced in the HILDA Survey in wave 9 (see Watson, 2010).

A range of questions about the fieldwork process are also completed by the interviewer to assist in future contacts with the household.

On the first approach to the household, the interviewer records a series of dwelling observations, primarily for the purposes of non-response adjustment. These questions include:

- type of dwelling;
- external condition of the dwelling;
- presence of security features, such as locked gate, security guard / doorman, security door, dangerous dog, no junk mail / hawkers sign, grab rails / ramp access, bars on windows, or roller shutters;
- likelihood that dwelling contains children under 15;
- condition of the garden; and
- type of road the dwelling is on.

The paper version of the Household Form is available on the HILDA website (see melbourneinstitute.com/downloads/hilda/Questionnaires/HouseholdFormTopUpW11M.pdf).

Once the Household Form is completed, one person in the household is asked to complete the Household Questionnaire and then each person aged 15 and older as at the 30th of June 2011 is interviewed using the New Person Questionnaire. To limit the individual interview to around 35 minutes, some sections of the Person Questionnaire are not asked of the top-up sample in 2011, these being:

- Section L, on retirement;
- Section M, on intentions and plans;
- Parts of Section G, on family formation (i.e., G27 to G67 on caring for children, grandchildren, and desires and preferences for more children); and
- Section E, on other labour market activity.

The next opportunity this sample will have to answer the fertility and retirement questions will be in 2015 (assuming a 4-year cycle), whereas the next opportunity to answer the labour market activity calendar and workplace training will be in 2012.

Following the Person Questionnaire, the respondent is also given the Self-Completion Questionnaire to fill in. This is picked up by the interviewer at a later date, or failing that the respondent is asked to return it by mail.

Multi-stage sample design

The top-up sample was selected using the same methodology as the original HILDA sample. That is, it is a three-stage area-based design that involves the selection of areas, then dwellings within those areas, and then households within those dwellings.

Stage 1: Selection of areas

At the first stage, 125 Census Collection Districts were selected with probability proportional to the number of occupied private dwellings as measured by the 2006 Census. The CDs in each state were sorted into a serpentine ordering based on the latitude and longitude of the centroid of the CD. This is essentially the same as taking a rope and starting at the top of the state or territory, laying it over the CDs first in a west to east direction, then in an east to west direction, moving down the state/territory until each CD has been passed through. The rope is

then stretched out and joined with those in the other states and territories. A systematic selection of the CDs (with probability proportional to the number of occupied private dwellings in each CD) was then taken. This method ensures that the sample provides a good coverage of the CDs across Australia.

The CDs selected in the top-up sample are distributed across the metropolitan and non-metropolitan areas of each state as shown in Table 1.

Table 1: Distribution of the clusters for the top-up sample

	Total	Metropolitan	Non-metropolitan
New South Wales	40	25	15
Victoria	31	23	8
Queensland	25	11	14
South Australia	10	8	2
Western Australia	12	9	3
Tasmania	4		
Northern Territory	1		
Australian Capital Territory	2		
Australia	125		

Within CDs that cover a very large area, a block selection stage was also employed. This involves the CD being broken up into blocks, an estimate of the number of dwellings in each block was generated (using mesh block counts from the 2006 Census and Google Maps of the areas) and a portion of the blocks selected for the dwelling listing stage. Only two CDs had a block selection stage.

Stage 2: Selection of dwellings

The second stage involved listing all of the dwellings in each of the selected areas and taking a random sample of dwellings from each.

A specifically trained team of interviewers visited the areas approximately 3 months prior to the fieldwork to record all the addresses within in the area. They followed a pre-determined path around the CD and entered the addresses into a tablet as they come across them. Particular attention was paid in recording granny flats, units above shops, battleaxe properties and residential warehouses to ensure that each dwelling has an equal probability of selection. Each address was classified as one of the following: i) private dwelling; ii) non-private dwelling (e.g., hotels, motels, nursing homes, guest houses, prisons, religious and charitable institutions, defence establishments, hospitals and other communal dwellings); or iii) non-private address (e.g., business, industrial buildings and vacant lots). Gathering an accurate count of the number of private dwellings in each area is necessary for weighting purposes. The number of private dwellings in each CD may have increased or decreased in the 5 years since the 2006 Census.

A sample of private dwellings was taken from each CD by first randomly selecting an initial dwelling and applying a skip down the list of private dwellings until the appropriate number

of dwellings was selected (non-private dwellings and non-private addresses are excluded from this list). As a general rule, a skip of five in urban areas and two in rural areas was used for cost reasons, however in high density areas (with an area of less than 0.15km²) the maximum skip was used to ensure a greater spread of the dwellings across the CD.

The actual number of dwellings selected in each CD varied depending on the expected response rates and occupancy rates. The expected response rates were based on the wave 1 experience of the HILDA Survey in the metropolitan and rural areas of each state, with some modifications given trends in response rates over the past decade based on advice from Roy Morgan Research. The expected occupancy rates were based on the 2006 Census, though additional selections for this reason are limited to 10 dwellings per CD. With a target of an average of 16 responding households in each CD, the expected number of occupied dwellings averages 24 and the average number of selected dwellings is 26.

The final sample was matched against the ongoing sample address list and any duplicates were removed from the top-up sample. This identified two such dwellings and for each of these an additional selection was taken from the CD (by applying the skip pattern from the last selected address).

Stage 3: Selection of households

The third stage of selection occurs in the field. Where there are three or fewer households in a dwelling, all are approached for interview, however if there are more than three then a random sample of three is taken. Based on the experience from wave 1, it is expected that this household-level selection will only need to occur at 2 or 3 dwellings in the top-up sample.

Fieldwork

Respondent communication

All selected dwellings were sent a pre-interview notification letter and brochure introducing the HILDA Survey and indicating that an interviewer will approach them in the next week for an interview. Several versions of the respondent communication materials were tested in two focus group sessions in April 2011, with a view to identify the style, layout, pictures, and content that is most effective in encouraging high response rates. The letter and brochure explains:

- what the survey is about;
- how the household was selected;
- who is involved in conducting the survey;
- what is involved in participating;
- the cash incentive;
- how their privacy is protected; and
- how to contact Roy Morgan Research if they have any questions about the survey.

The reverse side of the letter contained a short sentence in 8 languages introducing the survey and providing the 1800 number for sample members to call for more information.

As has been the practice with the ongoing sample in more recent years, the survey is marketed to the top-up sample under the banner of 'Living in Australia', though 'HILDA' is included in the logo and the sample members are alerted to the fact that the survey is

mentioned in the media and in reports as the ‘Household, Income and Labour Dynamics in Australia Survey’.

A copy of the respondent communication material is provided on the HILDA website (<http://www.melbourneinstitute.com/hilda/doc/ri-wave11.html>).

Interviewers

Around 75 face-to-face interviewers are undertaking the interviews for the top-up sample (see Table 2). Almost all of these interviewers worked on the HILDA Survey in wave 10 and they are also allocated many of their usual areas for the ongoing sample. To accommodate this additional workload, the fieldwork schedule was brought forward by 3 weeks (more detail about this is provided in the next section).

A further 65 or so face-to-face interviewers are working on the ongoing sample, 50 of whom have worked on the HILDA Survey before and 15 are new this wave.

This brings our face-to-face workforce to about 145 interviewers, which is only about 15 more than that used in wave 10 for the ongoing sample alone. That is, 15 new interviewers were recruited as a direct result of the top-up sample to fill in the gaps left by a small number of interviewers who need to reshuffle their work to take on the top-up sample as well as the majority of their ongoing areas.

The Team 1800 continues to provide a multi-faceted role of being telephone interviewers, providing interviewer support, tracking respondents who have moved, and answering respondent calls to the 1800 number. This team comprises of about 30 to 35 staff, some of whom conduct telephone interviews. The size of this team increased by about 5 to 10 staff members compared to previous waves in response to the addition of the top-up sample.

Table 2: Expected number of face-to-face interviewers for wave 11

Sample	Experienced HILDA interviewers	New-to-HILDA interviewers	Total interviewers
Both ongoing and top-up	70	5	75
Ongoing only	50	15	65
Top-up only	5	0	5
Total	125	20	145

Interviewer training and fieldwork schedule

To accommodate the additional work for wave 11, the start of the fieldwork was 3 weeks earlier than normal. Table 3 shows the fieldwork schedule for both the top-up sample and the ongoing sample, together with a comparison of what the fieldwork schedule would have looked like if we did not have the top-up sample this wave.

The interviewers working on the top-up sample were trained first. For the interviewers who worked on the HILDA Survey previously, the training session consisted of the standard 1-day training regarding the ongoing sample, plus a 1-day training specific to the top-up sample. The new interviewers working on the top-up sample were trained with other new interviewers (via a 3-day training session) and are then given the additional 1-day training specific to the top-up sample.

Interviewers working on both samples were given the top-up sample first and had two to three weeks to contact and interview these households.³ Once most of their households were completed, they were given their workloads relating to the ongoing sample (around the second or third week of August).⁴ This means that they can fit in the final calls to the top-up sample households amidst the calls to the ongoing sample. The interviewers only working on the ongoing sample were given their workloads from the end of July. This permits workload re-allocation of the ongoing sample should some of the interviewers working on both samples fall behind on their workload. While this shift in the start of fieldwork for the ongoing sample is expected to be beneficial for the response rates of both samples, it will result in a decrease in the proportion of sample members interviewed close to the anniversary of last wave's interview date for wave 11.

Table 3: Training and fieldwork schedule, wave 11

Month	Week	Top-up sample		Main sample		Main sample without top-up (comparison)	
		Training	Fieldwork	Training	Fieldwork	Training	Fieldwork
Jul	4	NSW, Vic, Qld	Period 1	NSW, Vic, Qld	Period 1		
Aug	1	SA/WA		NSW, Vic, Qld, SA/WA			
	2			NSW, Vic			
	3					NSW, Vic	Period 1
	4					NSW, Vic, Qld	
Sept	1					Qld, SA, WA	
	2						
	3						
	4						
	5						
Oct	1						
	2						
	3						
	4	All	Period 2	All	Period 2	All	Period 2
Nov	1						
	2						
	3						
	4						
	5						
Dec	1						
	2						
	3						
	4						
Jan	1						
	2	All	Period 3	All	Period 3	All	Period 3
	3						
	4						
Feb	1						
	2						
	3						

³ Note that the 2011 Census was on Tuesday 9th August, which was two weeks into our fieldwork period. Interviewers were advised not to make contact with the top-up sample households on this day unless prior agreement is made with household to do so.

⁴ Some interviewers may have worked their top-up sample and ongoing sample areas at the same time if it was practical in their situation to do so to minimise travel costs.

Following this initial fieldwork, the outcomes of all households will be reviewed and some will be reissued into period 2 for follow-up in late October and November. Typically the households are reissued to a different interviewer who is likely to have a different approach and more fieldwork experience. Another follow-up period occurs in January and February as appropriate to the situation of each household. This follow-up work is essential to gaining response rates that are as high as possible.

It is anticipated that this shift in the start of the fieldwork period, and the small increase in the number of face-to-face interviewers used, will be permanent changes in future waves.

Other fieldwork procedures

The other fieldwork procedures for the top-up sample are as per the ongoing sample (as documented by Watson (2002) and Summerfield (2010)). In summary:

- A cash gift of \$30 is provided at the end of each individual interview. If all adult members of the household participate, then the household receives an additional \$30.
- Interviewers make a minimum of 6 visits to a household to make contact, and then a further six visits to undertake the interviews.
- The visits are spread across a minimum of a five-day period, with a mixture of weekend and weekday as well as morning, afternoon and evening calls.
- An interviewer typically completes the fieldwork for an area within a three-week period.
- Almost all interviews with top-up sample members will be conducted face-to-face in wave 1. We would expect only a very small percentage (around 0.5 per cent) to be conducted by phone.
- Language difficulties are most often resolved with a family member or friend assisting the respondent with the interview. Otherwise a professional interpreter is arranged to visit the household with the interviewer to assist with the interview.
- Interviewers are monitored via daily fieldwork reports and reminder calls to send back Self-Completion Questionnaires. Validations of an interviewer's work are undertaken through routine contact with the respondents, such as when a household is reminded to return a Self-Completion Questionnaire, or when they are contacted to clarify some conflicting responses provided during the interview or when the household is allocated to a different interviewer later in the wave or the next wave.
- In-office coding staff and processing staff are monitored and validations are undertaken with of 10 per cent of their work.
- The response rates for the top-up sample are regularly monitored during the fieldwork period. This occurs daily within Roy Morgan Research and weekly within the Melbourne Institute.

Expected response rates

We anticipate a response rate for the top-up sample similar to that achieved with the original sample in wave 1, that is 66 per cent of all in-scope households will provide an interview. Further, within the responding households, we would expect that at least 92 per cent of the adults would provide an individual interview.

While we would prefer higher response rates, these targets are realistic. There is general evidence of declining response rates around the world (for example, de Leeuw and de Heer, 2002; Tourangeau, 2003), indicating that we have an increasingly difficult environment to gain the cooperation of sample members. Balanced against this is the strong experience of the interviewer workforce with this particular survey.

Data preparation

The data for the top-up sample will be incorporated into the household file, responding person file, enumerated person file and the combined file for wave 11 onwards. Each record will be flagged to indicate whether it is part of the ongoing sample or the top-up sample. The top-up sample members will also be added to the master file (again with an associated flag).

The household identifier for all household (whether part of the top-up sample or the ongoing sample) will be increased from five digits to six digits from wave 11 onwards. This is because the introduction of approximately 3000 households into the overall sample results in the total number of (responding and non-responding) households exceeding 10,000.⁵

A future technical paper will describe how the two samples will be combined for weighting purposes.

References

- De Leeuw, E., and de Heer, W. (2002), 'Trends in Household Survey Nonresponse: A Longitudinal and International Comparison', in Groves, R.M., Dillman, D.A., Eltinge, J.L., and Little, R.J.A. (eds), *Survey Nonresponse*, John Wiley and Sons, New York.
- Summerfield, M. (editor) (2010), 'HILDA User Manual – Release 9', Melbourne Institute of Applied Economic and Social Research, University of Melbourne.
- Tourangeau, R. (2003), 'The Challenge of Rising Nonresponse Rates', in R. Tourangeau, *Recurring Surveys: Issues and Opportunities*, Report to the National Science Foundation based on a Workshop held on March 28-29, 2003, National Science Foundation.
- 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.
- Watson, N. (2006), 'Options for a top-up sample to the HILDA Survey', *Proceedings of The ACSPRI Social Science Methodology Conference 2006*, ACSPRI, Canberra.
- Watson, N. (2010), 'The Impact of the Transition to CAPI and a New Fieldwork Provider on the HILDA Survey', *HILDA Project Discussion Paper Series*, No. 2/10, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

⁵ In the five digit identifier, the last digit of the identifier is reserved for split households that are identified during the wave and the first four digits is used to uniquely identify the household at the start of the wave. The limit on the number of households with a five digit identifier is therefore 10,000.