HILDA User Manual – Release 6

Last modified 4/04/08

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

Acknowledgements

The *HILDA User Manual-Release 6* was produced by Nicole Watson. Contributions by Mark Wooden, Nicole Watson, Simon Freidin, Clinton Hayes, Alison Goode, Paul Agius, Bruce Headey and Rosslyn Starick are included in this manual.

Simon Freidin, in consultation with Mark Wooden, Nicole Watson and Bruce Headey, prepared the HILDA Data Files and developed the associated data processing system with the support of Paul Agius and Diana Warren. In addition, Nicole Watson, Clinton Hayes, Peter Boal, Tim Fry, John Henstridge, Stephen Horn, Rosslyn Starick and Frank Yu contributed to the development of the weighting and imputation strategies.

Citation

Readers wishing to cite this document should use the following form of words:

> N. Watson (ed) (2008) *HILDA User Manual – Release 6*, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Updates to This Manual

Date	*Update*
04/04/08	Correction to Figure 20 and 21 (Income model) as foreign pensions are included in calculation of _tife variables.
13/03/08	Added note on setting SPSS missing values back on
22/02/08	Added note to Appendix 2 regarding changes to _edhists and _edhigh.
07/02/08	Added mention of User's Guide for calculating HILDA standard errors.
31/01/08	Updated Manual for Release 6.
23/04/07	Updated Survey Instrument Development and Sources (Appendix 1B). Updated Changes to the Data (Appendix 2A and 2B).
31/01/07	Updated Manual for Release 5.
22/05/06	Amended programs to create long longitudinal files with correct weights.
22/03/06	Added Table on Self Completion Questionnaire response rates.
07/02/06	Added new section on calculating hourly wage rates.
01/02/06	Added Lists of Figures and Tables. Renumbered tables and updated household sample figures for waves 2 to 4.
16/01/06	Updated Manual for Release 4.
14/04/05	Renumbered Appendix 1 to 1A. Added Appendix 1B on data item sources.
25/01/05	Added Appendix 3 on Little and Su method.
13/01/05	Original version of manual for Release 3.

Contents

LIST OF FIGURES ..... **VII**

LIST OF TABLES ..... **VIII**

USING THIS MANUAL **1**

OVERVIEW OF THE HILDA SURVEY **2**

THE HILDA SAMPLE AND FOLLOWING RULES: A SUMMARY 2
QUESTIONNAIRES 3
Household Form *4*
Household Questionnaire *4*
Person Questionnaires *4*
Self-Completion Questionnaire *5*

THE HILDA DATA **6**

ORDERING THE DATA 6
CROSS-NATIONAL EQUIVALENT FILE (CNEF) 7
A REMINDER OF THE SECURITY REQUIREMENTS FOR THE DATA 8
HOW THE DATA FILES ARE PROVIDED 9
STRUCTURE OF THE DATA FILES 9
IDENTIFIERS AND USEFUL VARIABLES 10
MATCHING FILES 12
MATCHING PARTNER INFORMATION TO FILES 13
CREATING LONGITUDINAL FILES 15
PANELWHIZ 27
VARIABLE NAME CONVENTIONS 27
MISSING VALUE CONVENTIONS 32
Numeric Variables *32*
Text Variables *33*
DATA WITH NEGATIVE VALUES 33
CONFIDENTIALISATION 34

DERIVED VARIABLES **35**

AGE AND SEX 35
HISTORY VARIABLES 36
SPATIAL VARIABLES 41
CURRENT EDUCATION VARIABLES 42
CURRENT MARRIAGE AND DEFACTO RELATIONSHIP VARIABLES 45
CHILDREN VARIABLES 45
CHILD CARE VARIABLES 46
EMPLOYMENT VARIABLES 56
CALCULATING HOURLY WAGE RATES 58
CALENDAR VARIABLES 59
FAMILY VARIABLES 59
HEALTH VARIABLES 62
HOUSING VARIABLES 63
TIME USE VARIABLES 64
HOUSEHOLD EXPENDITURE VARIABLES 64
PERSONALITY VARIABLES 67
RELIGION 67
INCOME VARIABLES AND INCOME IMPUTATION 68
Income, Tax and Family Benefits Model *68*
Imputation Method *75*
Imputed Income Variables *76*
WEALTH VARIABLES AND WEALTH IMPUTATION (WAVE 2 AND 6 SPECIAL TOPIC) 78
Wealth Model *78*
Imputation Method *81*
Imputed Wealth Variables *82*
WEIGHTS 84

Cross-Sectional Weights ... *84*
Longitudinal Weights ... *86*
Replicate Weights ... *87*
Weights Provided on the Data Files ... *87*
Advice on Using Weights ... *89*

DOCUMENTATION ... **92**

DOCUMENTATION CHOICES ... 92
MARKED-UP QUESTIONNAIRES ... 92
VARIABLE LISTINGS ... 93
Derived Variable Listing ... *93*
File-Based Listing ... *94*
Subject Listing ... *94*
Cross-Wave Variable Listing ... *94*
FREQUENCIES ... 95
ON-LINE DATA DICTIONARY ... 95

DATA QUALITY ISSUES ... **97**

SUMMARY OF DATA QUALITY ISSUES ... 97
MISSING INCOME DATA AND THE EXTENT OF INCOME IMPUTATION ... 103
MISSING WEALTH DATA AND THE EXTENT OF WEALTH IMPUTATION ... 106

THE HILDA SAMPLE ... **109**

SAMPLE DESIGN ... 109
Overview ... *109*
Reference Population ... *109*
Sampling Units ... *110*
Sample Selection ... *110*
FOLLOWING RULES ... 111

DATA COLLECTION ... **112**

PILOT TESTING ... 112
QUESTIONNAIRE LENGTH ... 112
INTERVIEWERS ... 113
FIELDWORK PROCESS ... 113
Data Collection Mode ... *113*
Timeline ... *114*
Survey Notification Material ... *115*
Respondent Incentives ... *115*
Call routine, Follow-Up and Refusal Aversion ... *116*
Foreign Language Interviews ... *116*
Interviewer Monitoring ... *116*
RESPONSE RATES ... 117
ATTRITION BIAS ... 123
DATA PROCESSING ... 125
Data Entry ... *125*
Coding Responses ... *125*

HILDA USER TRAINING ... **127**

GETTING MORE INFORMATION ... **128**

REFERENCES ... **129**

APPENDIX 1A: SUMMARY OF HILDA SURVEY CONTENT, WAVES 1 – 6 ... **131**

APPENDIX 1B: SURVEY INSTRUMENT DEVELOPMENT AND SOURCES ... **140**

HOUSEHOLD FORM (HF) ... 140
HOUSEHOLD QUESTIONNAIRE (HQ) ... 141
PERSON QUESTIONNAIRE (PQ) ... 143
SELF-COMPLETION QUESTIONNAIRE (SCQ) ... 151

APPENDIX 2: CHANGES TO THE DATA FROM RELEASE 5.1 156

CHANGES TO MASTER FILE BETWEEN RELEASES 5.1 (MARCH 2007) AND 6.0 (JANUARY 2008) 156
CHANGES TO WAVE 5 FILES BETWEEN RELEASES 5.1 (MARCH 2007) AND 6.0 (JANUARY 2008) 156
CHANGES TO WAVE 4 FILES BETWEEN RELEASES 5.1 (MARCH 2007) AND 6.0 (JANUARY 2008) 157
CHANGES TO WAVE 3 FILES BETWEEN RELEASES 5.1 (MARCH 2007) AND 6.0 (JANUARY 2008) 158
CHANGES TO WAVE 2 FILES BETWEEN RELEASES 5.1 (MARCH 2007) AND 6.0 (JANUARY 2008) 159
CHANGES TO WAVE 1 FILES BETWEEN RELEASES 5.1 (MARCH 2007) AND 6.0 (JANUARY 2008) 160

APPENDIX 3: LITTLE AND SU METHOD 162

FORMULAE 162
EXAMPLE 163

List of Figures

Figure 1: The Evolution of the HILDA Survey Sample Across the First Three Waves 3

Figure 2: SAS program to match wave 1 household and responding person files 12

Figure 3: SPSS program to match wave 1 household and responding person files 12

Figure 4: STATA program to match wave 1 household, enumerated and responding person files 12

Figure 5: SAS program to add partner variables 13

Figure 6: SPSS program to add partner variables 14

Figure 7: STATA program to add partner variables 14

Figure 8: SAS program to create long longitudinal files 16

Figure 9: SPSS program to create long longitudinal files 19

Figure 10: SAS macro to strip the first letter from the variable name 22

Figure 11: STATA program to create long longitudinal files 22

Figure 12: SAS program to create wide longitudinal file 24

Figure 13: SPSS program to create wide longitudinal file 25

Figure 14: STATA program to create wide longitudinal file 26

Figure 15: Family where new defacto relationship is formed 60

Figure 16: Family where new child is born 60

Figure 17: Family with child under 15 and non-dependent children 61

Figure 18: Income units in a family with child under 15 and non-dependent children 61

Figure 19: Financial Year Income Model: Household 70

Figure 20: Financial Year Income Model: Enumerated Person 71

Figure 21: Financial Year Income Model: Responding Person 72

Figure 22: Wealth Model Diagram, Household-level 79

Figure 23: Example of the marked-up questionnaires 93

Figure 24: Example of the derived variable listing 93

Figure 25: Example of the file-based listing 94

Figure 26: Example of the subject listing 94

Figure 27: Example of the cross-wave variable listing 95

Figure 28: Example of the frequencies 95

List of Tables

Table 1: Sections of the Person Questionnaires 5

Table 2: Sections of the Self-Completion Questionnaire 5

Table 3: HILDA Survey total number of data users: Release 1 to 5 6

Table 4: HILDA Survey data users by type: Release 1 to 5 6

Table 5: List of Useful Variables 11

Table 6: Broad subject area naming conventions, characters 2 and 3 (sorted by code) 28

Table 7: Unchanged variable name and differing codes for Household Response Status 29

Table 8: Unchanged variable name and differing codes for Person Response Status 30

Table 9: Unchanged variable name and differing codes for SCQ Field Response Status 31

Table 10: Unchanged variable name and differing codes for Household Membership 31

Table 11: Unchanged variable name and differing codes for New Location of mover 31

Table 12: Missing value conventions for numeric variables 32

Table 13: Missing value conventions for text variables 33

Table 14: History Variables 36

Table 15: Derived spatial variables 41

Table 16: Derived current education variables 43

Table 17: Derived current marriage and defacto relationship variables 45

Table 18: Derived children variables 45

Table 19: Derived child care variables 47

Table 20: Derived employment variables 57

Table 21: Derived calendar variables 59

Table 22: Derived family variables 62

Table 23: Derived health variables 63

Table 24: Derived housing variables 63

Table 25: Derived time use variables 64

Table 26: Derived household expenditure variables 64

Table 27: Derived personality variables 67

Table 28: Australian Resident Income Tax Rates, Waves 1- 6 69

Table 29: Other derived income variables 74

Table 30: Proportion of missing cases imputed by imputation method 76

Table 31: Imputed income variables 77

Table 32: Other derived wealth variables at household-level 80

Table 33: Proportion of missing cases imputed by imputation method (wealth), waves 2 and 6 82

Table 34: Proportion of missing cases imputed by imputation method (home value), waves 1 - 6 82

Table 35: Imputed wealth variables 82

Table 36: Weights 87

Table 37: Sample design variables 91

Table 38: Summary of the data quality issues in the HILDA data 97

Table 39: Proportion of cases with missing income data, waves 1 – 6 104

Table 40: Mean financial year income ($) (including imputed values) and proportion of mean income imputed, waves 1 – 6 (weighted) 105

Table 41: Proportion of cases with missing wealth data including and excluding wealth band responses, waves 2 and 6 107

Table 42: Proportion of households with missing home value data, waves 1 – 6 107

Table 43: Mean wealth value ($) (including imputed values) and proportion of mean value imputed, waves 2 and 6 (weighted) 108

Table 44: Mean home value ($) (including imputed values) and proportion of mean value imputed, waves 1 – 6 (weighted) 108

Table 45: Average time (minutes) taken to complete questionnaires, waves 1 to 6 112

Table 46: Number of interviewers and percentage of new interviewers each wave 113

Table 47: Proportion of respondents interviewed by telephone (%) 113

Table 48: Fieldwork dates and proportion of interviews post December 114

Table 49: Distribution of individual interviews by month, waves 2 to 6 114

Table 50: Time between interview dates and anniversary of previous wave interview, waves 2 to 6 115

Table 51: Wave 1 Household Outcomes 117

Table 52: Wave 1 Person Outcomes 117

Table 53: Wave 2 Household Outcomes 119

Table 54: Wave 3 Household Outcomes 119

Table 55: Wave 4 Household Outcomes 120

Table 56: Wave 5 Household Outcomes 120

Table 57: Wave 6 Household Outcomes 121

Table 58: Wave 2 to 6 Person Outcomes Against Wave 1 Person Outcomes..... 121

Table 59: Individual response rates (%) for the HILDA Survey, waves 2 to 6 compared 122

Table 60: Self Completion Questionnaire response rate by wave 122

Table 61: Percentage of Wave 1 Respondents Re-interviewed by Selected Sample Characteristics 124

USING THIS MANUAL

The HILDA Survey is a household-based longitudinal survey named the Household, Income and Labour Dynamics in Australia Survey. This manual has been designed for the users of the HILDA data.

The manual aims to cover the things that you need to know to use the HILDA data – such as missing data conventions, an introduction to the derived variables, how to put the data files together, income imputation, how to find your way around the documentation, and how to deal with things such as weights.

The best way to use this manual is as a reference tool. It is unlikely that you will sit down and read it cover to cover and take away everything you need to know about the data. More realistically, you will start to work with the data and will need some information about certain aspects of the data or the survey – and hopefully you will be able to find it within this manual fairly easily.

This is the fourth version of the manual and we welcome any feedback you have. It is an evolving manual and will be updated as successive waves are made available to researchers. If there is something that you expected to find in the manual and didn't, or if you had difficulty finding or understanding any section, please let us know (email hilda-inquiries@unimelb.edu.au).

OVERVIEW OF THE HILDA SURVEY

The HILDA Survey is a broad social and economic longitudinal survey, with particular attention paid to family and household formation, income and work. As the HILDA Survey has a longitudinal design, most questions are repeated each year. Nevertheless, within each survey wave, scope exists for asking questions on topics that will not be covered every year. The main additional topics to date are as follows:

- Wave 1 – Family background and personal history variables (subsequently included in every New Person Questionnaire);
- Wave 2 & 6 – Household wealth;
- Wave 3 – Retirement and plans for retirement;
- Wave 4 – Private health insurance, and youth;
- Wave 5 – Fertility and partnering, intentions and plans.

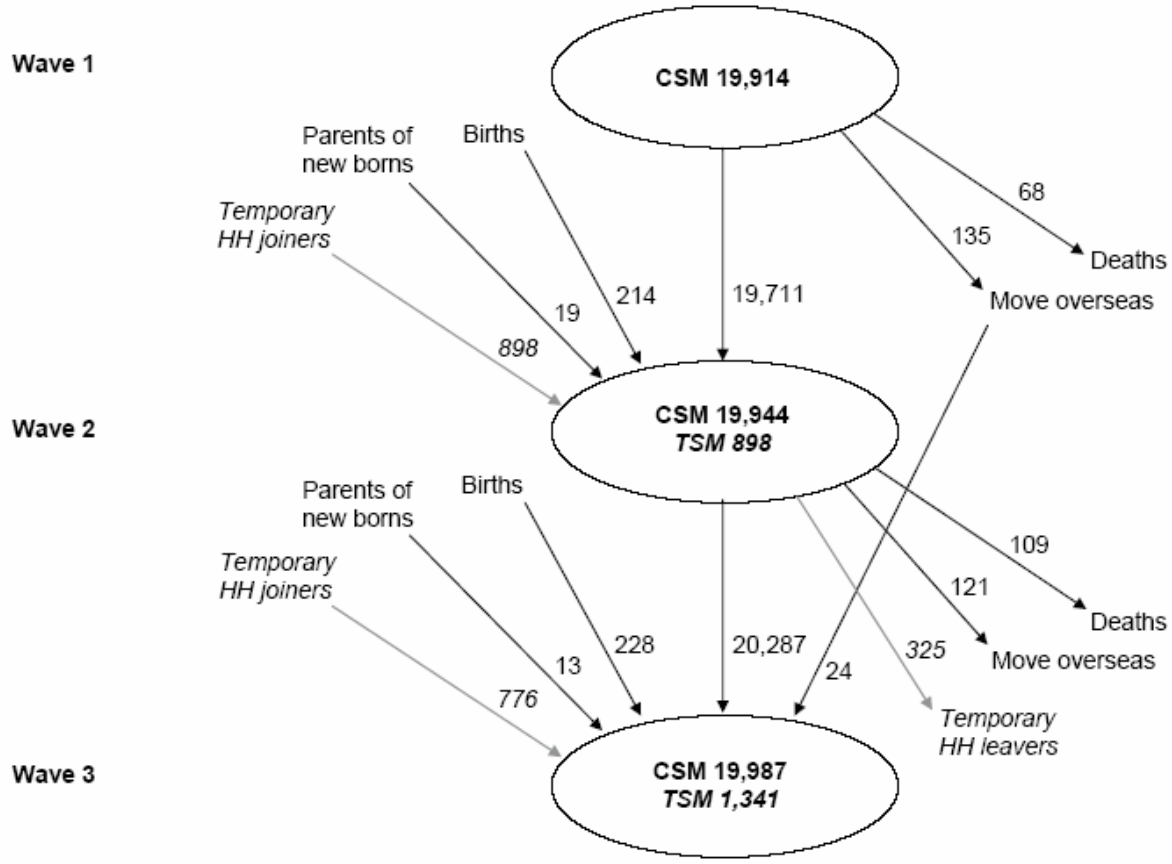
The HILDA Sample and Following Rules: A Summary

The HILDA Survey began with a large national probability sample of Australian households occupying private dwellings. All members of the households providing at least one interview in wave 1 form the basis of the panel to be pursued in each subsequent wave. The sample has been gradually extended to include any new household members resulting from changes in the composition of the original households.

Continuing Sample Members (CSMs) are defined to include all members of wave 1 households. Any children subsequently born to or adopted by CSMs are also classified as CSMs. Further, all new entrants to a household who have a child with a CSM are converted to CSM status. CSMs remain in the sample indefinitely. All other people who share a household with a CSM in wave 2 or later are considered Temporary Sample Members (TSMs). TSMs are followed for as long as they share a household with a CSM. The variable *hhsm* on the masterfile identifies TSMs while the CSMs are split into two groups: OSMs (original sample members from wave 1) and OPMs (other permanent sample members, ie 'new' CSMs).

Figure 1 shows an example of how the sample evolved across the first three waves. In wave 1, the sample consisted of 19,914 people. A further 442 births and 54 parents of newborns who were not originally CSMs have been added to the sample in waves 2 and 3. A total of 177 deaths have been identified across the two follow-up waves and 256 people have moved overseas, though 24 returned after being away for one wave. Of the TSMs joining the sample in wave 2, a third had moved out by wave 3.

Figure 1: The Evolution of the HILDA Survey Sample Across the First Three Waves



Questionnaires

In wave 1, the HILDA survey comprised four different instruments. These were:

- the Household Form (HF);
- the Household Questionnaire (HQ);
- the Person Questionnaire (PQ); and
- the Self-Completion Questionnaire (SCQ).

In subsequent waves, the PQ was replaced with two instruments:

- the Continuing Person Questionnaire (CPQ), for people who have been interviewed in a previous wave, and
- the New Person Questionnaire (NPQ), for people who have never been interviewed before (which collects family background and personal history information).

Appendix 1a provides a guide to topics covered in the HILDA Survey across the first five waves. Appendix 1b provides a list of sources used in constructing survey questions.

Household Form

The HF is designed to record basic information about the composition of the household immediately after making contact. The HF is the 'master document' used by interviewers to decide who to interview, how to treat joiners and leavers of the household, and to record call information and non-interview reasons. The date the HF is completed is provided in *_hhcomps*.

Household Questionnaire

The HQ collects information about the household rather than about individual household members per se, and is only administered to one member of the household. In practice, however, interviewers are encouraged to be flexible. If more than one household member wishes to be present at the interview this is perfectly acceptable. Further, interviewers are given the flexibility to deliver part of this interview to one household member and part to another. Indeed, this was often required, with questions on childcare needing to be asked of the primary care giver. The date the HQ is completed is provided in *_hhqivw*.

The HQ mainly covers childcare arrangements, housing, household spending (until wave 5) and, in waves 2 and 6, household wealth.

Person Questionnaires

The CPQ is administered to every member of the household aged 15 years and over who has previously completed a person questionnaire. The NPQ is administered to every member of the household aged 15 years and over who has not previously completed a person questionnaire. Parental consent is sought before interviewing persons aged under 18 years. The date the PQ is completed is provided in *_hhidate*.

The sections of the person questionnaires are shown in Table 1 together with the letter used to identify the section. These will help you locate questions on the questionnaires (for example, if you wanted to find questions on education, look in section C of the wave 1 Person Questionnaire and section A of the Continuing Person Questionnaire and New Person Questionnaire from wave 2 onwards).

The PQ in Wave 1 is distinctive from that used in the later waves by collecting biographical data that only needs to be asked once. These questions are spread throughout the survey and include questions about country of birth and language, family background, educational attainment, employment history, and marital history. In addition, at later waves further biographical information about visa category for immigrants (wave 4) and parents' education (wave 5) was collected.

The NPQ differs from the CPQ in the inclusion of these additional biographical history questions.

Table 1: Sections of the Person Questionnaires

Topics	*Section*	
	Wave 2 onwards	*Wave 1*
Country of birth	AA (NPQ only, except in wave 4*)	A
Family background	BB (NPQ only)	B
Education	A	C
Employment status	B	D
Current employment	C	E
Persons not in paid employment	D	D, F
Annual activity calendar	E	FG
Income	F	G
Family formation	G	H
Partnering/Relationships	H	J
Health, life satisfaction, moving	K	K
Tracking information	T	T
Interviewer observations	Z	Z
Special Topics		
Wealth (wave 2 and 6)	J	
Retirement (wave 3 only)	L	
Private health insurance (wave 4 only)	J	
Youth issues (wave 4 only)	L	
Fertility and partnering (wave 5)	G, H	
Intentions and Plans (wave 5)	L	

* Immigration Status asked in wave 4 in section AA

Self-Completion Questionnaire

Finally, all persons completing a person questionnaire are asked to complete a Self-Completion Questionnaire which the interviewer collects at a later date, or failing that, is returned by mail. This questionnaire comprises mainly attitudinal questions, many of which cover topics which respondents may feel slightly uncomfortable answering in a face-to-face interview. The date that the SCQ is completed is not collected.

Table 2 shows the sections of the SCQ together with the letter used to identify the section.

Table 2: Sections of the Self-Completion Questionnaire

Topics	*Wave 2 onwards*	*Wave 1*
General health and well-being (SF-36)	A	A
Lifestyle and living situation	B	B
Personal and household finances	C	C
Attitudes and values	-	D
Job and workplace issues	D	E
Parenting	E	F
Sex and age	F	-

THE HILDA DATA

The HILDA Survey has already developed a sizeable community of users. Table 3 and Table 4 show the total number of individuals who have been approved access to at least one of the last five data releases and the composition of our user community. There are also 28 users of the HILDA-Cross-National Equivalent File (HILDA-CNEF).

Table 3: HILDA Survey total number of data users: Release 1 to 5

Release	*Total data orders*	*Orders by new users*	*Cumulative no. of users*
Release 1	204	204	204
Release 2	265	167	371
Release 3	280	154	525
Release 4	329	171	696
Release 5	369	165	861

Table 4: HILDA Survey data users by type: Release 1 to 5

	Release				
Type of user	1	2	3	4	5
Academic – Australia	84	105	118	142	168
Academic – Overseas	5	13	20	20	24
Students – Honours year	5	14	18	15	12
Students – Postgraduate	9	23	24	33	43
Government – Australian	87	89	85	100	104
Government – State/Local	7	13	5	14	7
Other	7	8	10	5	11
Total	204	265	280	329	369

Ordering the Data

As of Release 6, the access to the data can be gained by an Organisational Licence or an Individual Licence. Organisations that are likely to have more than four individuals who wish to use the HILDA data should consider signing up to an Organisational Licence as this would provide quicker access to the data (and at a lower cost) once the Organisational Licence is signed.

Researchers who wish to use the HILDA data must:

- Read the *Manual for the Access to and Use of FaHCSIA Longitudinal Survey Datasets* which provides details of the licencing arrangements. The manual is available on the 'data and documentation' page of the HILDA website: www.melbourneinstitute.com/hilda/data.html.

- If your organisation has signed an Organisational Deed of Licence, then you should:
    - Contact the Data Manager for your organisation for the Deed of Confidentiality;
    - Sign a Deed of Confidentiality;
    - Be given access to the data by your Data Manager (your organisation would have paid a flat fee of $330 for administration costs for each Release).
- If your organisation has not signed an Organisational Deed of Licence, then you should:
    - Complete the appropriate Order Form and Individual Deed of Licence, available from the HILDA website: www.melbourneinstitute.com/hilda/data.html. To avoid lengthy delays in processing your application, make sure you have completed EVERY part that you need to (see the notes about this on the Order Form).
    - Send these completed documents to the address indicated on the Order Form, together with your payment of $77 for administration costs ($121 for overseas users).
    - Receive an email from the Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) confirming that your Order Form and Individual Deed of Licence has arrived. If you do not receive this, then email: hilda@fahcsia.gov.au as soon as possible to avoid delays in your application.
    - Wait approximately two to three weeks for the delivery of the data on DVD.
- If you work for FaHCSIA, then you have different access arrangements and you should contact hilda@fahcsia.gov.au.

Up until Release 4 the HILDA data files have been referred to as the "confidentialised" and "unconfidentialised" files. From Release 5 onwards these files will be referred to as the "General Release" files (the confidentialised files) and the "In-confidence Release" files (the unconfidentialised files).

Cross-National Equivalent File (CNEF)

There is now a cross-national equivalent file available for HILDA. This equivalent file provides a set of constructed variables that are not available on the original surveys. For more details on the CNEF and how to order both the HILDA-CNEF and CNEF data for other countries please go to: http://www.melbourneinstitute.com/hilda/cnef/cnef.htm

A Reminder of the Security Requirements for the Data

The Deed of Licence stipulates numerous security requirements for the data, some of which are outlined below:

- If you plan to change employers and you have an Individual Deed of Licence, you MUST contact FaHCSIA prior to doing so to discuss suitable arrangements for the data. Under certain conditions you may be able to take the data with you. Otherwise, you will need to delete any data files and destroy the CD/DVD.

- If you plan to change employers and your organisation has an Organisational Deed of Licence, you MUST contact your organisation's Data Manager to resolve your user obligations to the security of the dataset.

- If you change your research project you MUST seek permission for the new project from FaHCSIA.

- The HILDA CD/DVD MUST be kept secure in a locked filing cabinet or other secure container when not in use.

- The keys or combinations for the filing cabinet or other secure container must be kept secure and not given to any unauthorised user.

- The HILDA data (and any derivatives of the HILDA data) must be stored on a password protected computer or network.

- Your password MUST include a mixture of upper and lowercase characters, be at least 7 characters long, and include some non-alphanumeric characters such as # ; * or !.

- Any printed unit record output MUST be stored in a locked filing cabinet or other secure container when not in use. Any printed unit record output MUST be shredded if no longer required.

- You CANNOT provide the data to any unauthorised individual (to be authorised, you must have an Individual Deed of Licence countersigned by the FaHCSIA delegate or have a Deed of Confidentiality countersigned by your organisation's approved Data Manager).

- There MUST be a means of limiting access to the work area where the data is kept and tamper evident barriers to access (i.e., if there were a break-in, it would be obvious from broken glass, damaged lock, etc).

How the Data Files are provided

All data are provided in SAS, SPSS[1] and STATA (Version 8) formats[2].

The DVD also includes extensive documentation of the data, including coding frameworks, marked-up questionnaires and variable frequencies. The files and the documentation are discussed in detail in later sections. Changes to the data files between Release 5.1 and Release 6 are provided in Appendix 2. Changes to the data files, including those made for earlier releases, can be found on the HILDA website: http://www.melbourneinstitute.com/hilda/doc/doc_hildamanual.htm.

The data files can be transferred to other statistical packages using StatTransfer, DBMS Copy, or any other data conversion package of your choice.[3] You may need to restrict the number of variables to be included in your transferred datasets due to the limitations on the number of variables imposed by some other packages.

Structure of the Data Files

For each wave, there are four files:

- Household File – containing information from the HF and HQ.
- Enumerated Person File – listing all persons in all ***responding*** households and contains limited information from the HF (includes respondents, non-respondents and children).
- Responding Person File – containing all persons who provided an interview and contains CPQ/NPQ and SCQ information.
- Combined file – this is a combined file of the three files above. The household information and responding person information is matched to each enumerated person.

In addition, a master file and a longitudinal weights file are provided with the files for the most recent wave (for Release 6, this is wave 6). The master file contains all persons enumerated at any wave, their interview status in each wave and limited information about the individual. The longitudinal weights file contains weights for all sequential balanced panel combinations and all balanced pairs of waves.

[1] SPSS portable files provided on previous releases are no longer provided on the DVD but can be obtained by special request.
[2] You will need to use STATA SE as there are more than 2047 variables in the datasets. Suitable *memory* and *maxvar* values are provided in "Readme 60.pdf" on the DVD.
[3] A trial copy of StatTransfer Version 9 can be downloaded from www.stattransfer.com or purchased online at www.stattransfer.com/html/store.html. DBMS/Copy Version 8 can be purchased online at www.dataflux.com/product-services/products/dbms.asp.

Identifiers and Useful Variables

Household and person level files within a wave can be merged by using *_hhrhid* (i.e. ahhrhid for wave 1, bhhrhid for wave 2, etc).[4] Note that where we use the underscore '_' in the variable name, you will need to replace it with the appropriate letter for the wave, 'a' for wave 1, 'b' for wave 2, etc. Enumerated and responding person files within a wave can be merged by using the cross-wave identifier *xwaveid* or the wave specific person identifier *_hhrpid*. In wave 1, the first six characters of *_hhrpid* is the household identifier and the last two characters of the person identifier is the person number within the household. In wave 2 onwards, the first five characters are the household identifier and the next two are the person number.

Information from enumerated or responding person files can be linked across waves by using either:

- the cross-wave identifier *xwaveid*; or
- the master file which shows the identifiers for each person each wave.

Partners within the household are identified by their cross-wave identifier (*_hhpxid*) or by their two digit person number for the household (*_hhprtrid*). These variables are provided on both the enumerated and responding person files and are derived using the HF relationship grid. Partners are either married or de-facto and include same sex couples. *_hhprtid* is the person number for the household (for example, if person 02's partner is person 05, the partner identifier for person 02 will contain '05' and for person 05 it will contain '02'). You will need to concatenate the household identifier with the partner identifier before you can match on partner characteristics to the person file. Using the partner's cross-wave identifier (_hhpxid) will be much easier.

Parents within the household are similarly identified in *_hhfxid* and *_hhmxid* (father's and mother's crosswave identifiers) or *_hhfid* (father's person number) and *_hhmid* (mother's person number). A parent may be natural, adopted, step or foster (a parent's de facto partner also counts as a parent).

Note that while *xwaveid* is the unique identifier to match each person across all waves, *_hhrhid* and *_hhrpid* are specific identifiers to match each person within a wave. As *_hhrhid* and *_hhrpid* are randomly assigned each wave, the same person will have a different *_hhrhid* and *_hhrpid* from wave to wave. Persons in the same household in each wave will share the same *_hhrhid* and the same first 5 digits in *_hhrpid* (or the same first 6 digits in *ahhrpid* in the case of wave 1).

Listed below in Table 5 are some useful socio-demographic variables. These are provided to help new users get started with using the HILDA data.

[4] Users of the In-confidence Release (unconfidentialised) files can alternatively use *_hhid* to match the household and person files, and *_hhpid* to match the person files. In wave 1, the household identifier is six digits long, corresponding to area (three digits), dwelling number (two digits) and household number (one digit). The person identifier in wave 1 is then eight digits long – the first six are the household identifier, followed by two digits for the person number. In subsequent waves, the household identifier is five digits long, and the person identifier is seven digits long.

Table 5: List of Useful Variables

Variable	*Description*	*Variable*	*Description*
_xwaveid	Cross wave person identifier	*_hhfty*	Family type
_hhrhid	Random household identifier	*_hhiu*	Income unit
_hhrpid	Random person identifier	*_hhpxid*	Partner's cross-wave identifier
hhsm	Sample member type	*_hhfxid*	Father's cross-wave identifier
_hhresp	Household response status	*_hhmxid*	Mother's cross-wave identifier
_fstatus	Person response status (master file)	*_hhmsr*	Major statistical region
_hhpers	Number of persons in household	*_hhstate*	State
_hhtype	Household type	*_hhsos*	Section of state
_hhyng	Age of youngest person in HH	*_ancob*	Country of birth
_hhold	Age of oldest person in HH	*_hgage*	Age
_hhrih	Relationship in household	*_hgsex*	Gender
_hhfam	Family number	*_mrcurr*	Marital status
_esbrd, _esdtl	Employment status (broad, detail)	*_losat*	Life satisfaction
_jbhruc	Combined per week usually worked in all jobs	*_edfts*	Full-time student
_jbmocc2	Occupation code 2-digit ASCO	*_edagels*	Age left school
_wscei	Imputed current weekly gross wages and salary – all jobs	*_edhists*	Highest year of school completed/currently attending
_wsfei	Imputed financial year gross wages and salary	*_edtypes*	Type of school attended/attending
_hifdip, _hifdin	Household disposable income (positive and negative)	*_edhigh*	Highest education level achieved
_hhda10	SEIFA decile of socio-economic disadvantage	*_fhelth*	Long term health condition/disability/impairment from PQ

Matching Files

Two sample programs (Figure 2 and Figure 3) are provided below in SAS and SPSS for matching the household and responding person files together. Figure 4 provides a STATA program for matching the household, responding person and enumerated person files together.

Figure 2: SAS program to match wave 1 household and responding person files

```
/* ************************************************************ */
/* Created by: Simon Freidin (Updated by: Nicole Watson) */
/* ************************************************************ */
libname w1 'e:\release6.0\a60c';
data w1p;
  set w1.rperson_a60c;
data w1h;
  set w1.household_a60c (keep=ahhrhid ahifefn ahifefp ahifdip ahifdin);
proc sort data = w1p; by ahhrhid;
proc sort data = w1h; by ahhrhid;
data w1p;
  merge w1p w1h;
  by ahhrhid;
```

Figure 3: SPSS program to match wave 1 household and responding person files

```
/* ************************************************************ */
/* Created by: Simon Freidin (Updated by: Nicole Watson) */
/* ************************************************************ */

file handle w1p /name='e:\release6.0\a60c\Rperson a60c.sav'.
file handle w1h /name='e:\release6.0\a60c\Household a60c.sav'.
file handle w1psort /name='e:\release6.0\a60c\Rperson sorted a60c.sav'.
file handle w1hsort /name='e:\release6.0\a60c\Household sorted a60c.sav'.

get file=w1h / keep= ahhrhid ahifefn ahifefp ahifdip ahifdin.
sort cases by ahhrhid.
save outfile=w1hsort.
get file=w1p.
sort cases by ahhrhid.
save outfile=w1psort.
match files file=w1psort / table=w1hsort / by ahhrhid.
exe.
```

Figure 4: STATA program to match wave 1 household, enumerated and responding person files

```
/* ************************************************************ */
/* Created by: Alison Goode (Updated by: Nicole Watson) */
/* ************************************************************ */

*Enumerated Person File
use "C:\eperson_a60c.dta",clear
sort ahhrpid
save "C:\eperson_2001.dta", replace
clear

*Responding Person File
use "C:\rperson_a60c.dta",clear
sort ahhrpid
save "C:\rperson_2001.dta", replace
clear
```

Figure 4: (c'td)

```
*Household File
use "C:\household_a60c.dta",clear
sort ahhrhid
save "C:\household_2001.dta", replace
clear

use "C:\eperson_2001.dta",clear                    /*1: merge enumerated and responding person files*/
merge ahhrpid using "C:\rperson_2001.dta"
drop _merge
sort ahhrhid

merge ahhrhid using "C:\household_2001.dta"        /*2: merge new file with household file*/
drop _merge
sort xwaveid

save "C:\Wave1.dta", replace                       /*3: now save cross-section of combined files*/

erase "C:\eperson_2001.dta"
erase "C:\rperson_2001.dta"
erase "C:\household_2001.dta"
clear
```

Matching Partner Information to Files

Some users may want to include variables for a respondent's partner in their analyses. The following sample programs in SAS, SPSS and Stata show how to utilise the partner's cross-wave identifier *_hhpxid* to add partner variables onto the responding person file.

Figure 5: SAS program to add partner variables

```
/* ************************************************************ */
/* Created by: Clinton Hayes (Updated by: Nicole Watson) */
/* ************************************************************ */
* Merge partner details on to the responding person file;
libname w5 "X:\HILDA\Release 60\files\SAS e60c" access=readonly;

*extract and sort the responding person file;
proc sort data=w5.rperson_e60c out=rperson; by xwaveid; run;

* Select hours worked, age and sex;
* rename and merge back with the main file;
data partners;
  set rperson (keep = ehhpxid ehgage ehgsex ejbhruc
            rename=(ehhpxid=xwaveid
                  ejbhruc = parthour
                  ehgage  = partage
                  ehgsex  = partsex ));
  where xwaveid ne '';
run;

proc sort data=partners; by xwaveid; run;

data rperson;
  merge partners rperson (in=a);
  by xwaveid;
  if a;
run;
```

Figure 6: SPSS program to add partner variables

```
/* ************************************************************ */
/* Created by: Simon Freidin (Updated by: Nicole Watson) */
/* ************************************************************ */

/* Sort the responding person file by xwaveid and save the sorted file */
get file='X:\HILDA\Release 60\files\SPSS e60c\Rperson e60c.sav'.
sort cases by xwaveid.
save outfile='E:\working\Rperson e60c.sav'.

get file='X:\HILDA\Release 60\files\SPSS e60c\Rperson e60c.sav'
  /keep= ehhpxid ehgage ehgsex ejbhruc .
sel if ehhpxid ne ''.
rename vars
  (ehhpxid   = xwaveid)
  (ejbhruc   = parthour)
  (ehgage    = partage)
  (ehgsex    = partsex).

sort cases by xwaveid.
match file table=*/file='E:\working\Rperson e60c.sav'/by xwaveid.
exe.
```

Figure 7: STATA program to add partner variables

```
/* ************************************************************ */
/* Created by: Clinton Hayes (Updated by: Nicole Watson) */
/* ************************************************************ */
use "X:\HILDA\Release 60\files\STATA e60c\Rperson_e60c.dta", clear
sort xwaveid
save "E:\working\rperson.dta", replace

use "E:\working\rperson.dta", clear
keep ehhpxid ehgage ehgsex ejbhruc
drop if ehhpxid ==""
rename ehhpxid  xwaveid
rename ejbhruc  parthour
rename ehgage   partage
rename ehgsex   partsex
sort xwaveid
save "E:\working\partner.dta", replace

use "E:\working\rperson.dta"
sort xwaveid
merge xwaveid using "E:\working\partner.dta"
drop if _merge==2
drop _merge
save "E:\working\rperson.dta", replace
```

Creating Longitudinal Files

There are a number of ways users might want to create a balanced longitudinal file:

- Wide file of responding persons – this is where we keep only people responding in all waves and put the variables for each wave next to each other.
- Wide file of enumerated persons – this is where we keep only those people who were in responding households in all waves and the variables for each wave are put next to each other.
- Long file of responding persons – this is where we keep only people responding in all waves and the information for each wave is stacked together (that is, there is a separate row of data for each wave of information for each person).
- Long file of enumerated persons – this is where we keep only those people who were in responding households in all waves and the information for each wave is stacked together (that is, there is a separate row of data for each wave of information for each person).

Most users will probably want to restrict the files to only include respondents or people from responding households. A few users may also want to add people who have died or moved out of scope (depending on the research question they are answering).

Example SAS and SPSS programs to create balanced long files of responding persons are provided in Figure 8 and Figure 9 below. The wide files are created by matching the responding or enumerated files for each wave together using *xwaveid*. An alternative way to strip off the first letter of the variable names in SAS is provided in Figure 10.

Some users may want to create an unbalanced panel – where you take all respondents or enumerated persons available at each wave (not just those that consistently respond or are consistently in responding households). An example STATA program to create an unbalanced panel and balanced panel is provided in Figure 11 below.

Example SAS, SPSS and STATA programs to create wide files are provided in Figure 12, Figure 13 and Figure 14, respectively.

The longitudinal weights on the enumerated person file and the responding person file are for the full balanced panel of respondents and enumerated persons from wave 1 (i.e., across the first two, three, … six waves). If you are constructing a balanced panel with different specifications, you should find a suitable weight in the longitudinal weights file. Out of scopes are treated as acceptable outcomes, so these people have weights applied as well.

Figure 8: SAS program to create long longitudinal files

```
/* Example SAS program to create balanced enumerated person */
/* (combined) long longitudinal file. */
/* Created by: Simon Freidin (Updated by: Nicole Watson) */
/* ************************************************************************ */
/*  Section 1: locate datasets, get balanced cases, load renaming macro  */
/* ************************************************************************ */
* (1.1) Location of datasets (edit to reflect your paths);
  libname wave1 'e:\a\written datasets\a60\SAS a60c';
  libname wave2 'e:\b\written datasets\b60\SAS b60c';
  libname wave3 'e:\c\written datasets\c60\SAS c60c';
  libname wave4 'e:\d\written datasets\d60\SAS d60c';
  libname wave5 'e:\e\written datasets\e60\SAS e60c';
  libname wave6 'e:\f\written datasets\f60\SAS f60c';

  libname long  'e:\a\data\long';                    * output datasets;

* Master file;
  data master;
  set wave6.master_f60c (keep=xwaveid ahhrpid afstatus bhhrpid bfstatus chhrpid cfstatus
                             dhhrpid dfstatus ehhrpid efstatus fhhrpid ffstatus);
  * Wave 1 enumerated persons;
  data w1c;
    set wave1.Combined_a60c;
  * Wave 2 enumerated persons;
  data w2c;
    set wave2.Combined_b60c;
  * Wave 3 enumerated persons;
  data w3c;
    set wave3.Combined_c60c;
  * Wave 4 enumerated persons;
  data w4c;
    set wave4.Combined_d60c;
  * Wave 5 enumerated persons;
  data w5c;
    set wave5.Combined_e60c;
  * Wave 6 enumerated persons;
  data w6c;
    set wave6.Combined_f60c;
  run;

* (1.2) Identify cases responding in all waves from the master file;
*       status codes: 1=face to face interview, 2=telephone interview;
*       Release 6 balanced panel, n = 8864 ;
data masters (keep=xwaveid ahhrpid bhhrpid chhrpid dhhrpid ehhrpid fhhrpid);
  set master;
  if  afstatus in (1,2) and bfstatus in (1,2) and cfstatus in (1,2)
  and dfstatus in (1,2) and efstatus in (1,2) and ffstatus in (1,2);
run;
* (1.3) Macro to rename variables, dropping first character of variable name,;
*       ID's are not renamed;
*       Writes a sas program to 'c:\temp', alter path if no c:\temp, also alter paths in %include statements;
  %macro rename (ds);
  data ren;
    set &ds;
    if _n_=1;
  proc transpose data=ren out=ren2;var _all_;
  data ren3;
    file 'c:\temp\rencmds.sas';
    set ren2 end=eof;
```

Figure 8: (c'td)

```
    if _n_=1 then put " rename";
    if _name_ not in ("XWAVEID","xwaveid",
    "AHHRHID","AHHRPID","BHHRPID","BHHRHID","CHHRPID","CHHRHID","DHHRPID","DHHRHID",
    "EHHRHID","EHHRPID","FHHRHID","FHHRPID",
    "ahhrhid","ahhrpid","bhhrpid","bhhrhid","chhrpid","chhrhid","dhhrpid","dhhrhid","ehhrhid","ehhrpid",
    "fhhrhid","fhhrpid")
    then do;
      _name2_=_name_ || '=' || substr(_name_,2);
      put " " _name2_;
    end;
    if eof then put " ;";
  run;
  %mend;
run;

/* ************************************************************** */
/* Section 2: create "long" responding person longitudinal file     */
/* ************************************************************** */

* (2.1) Rename waves. Run macro and then include generated rename;
*       command file. Keep only those records in selected master file.;
%rename(w1c);run;
data longcw1;
  set w1c;
  %include 'c:\temp\rencmds.sas';
proc sort data=masters; by xwaveid;
proc sort data=longcw1; by xwaveid;
data longcw1;
  merge masters (in=inc) longcw1;
  by xwaveid;
  if inc;
run;

%rename(w2c); run;
data longcw2;
  set w2c;
  %include 'c:\temp\rencmds.sas';
proc sort data=masters; by xwaveid;
proc sort data=longcw2; by xwaveid;
data longcw2;
  merge masters (in=inc) longcw2;
  by xwaveid;
  if inc;
run;

%rename(w3c); run;
data longcw3;
  set w3c;
  %include 'c:\temp\rencmds.sas';
proc sort data=masters; by xwaveid;
proc sort data=longcw3; by xwaveid;
data longcw3;
  merge masters (in=inc) longcw3;
  by xwaveid;
  if inc;
run;
%rename(w4c); run;
data longcw4;
  set w4c;
```

Figure 8: (c'td)

```
%include 'c:\temp\rencmds.sas';
proc sort data=masters; by xwaveid;
proc sort data=longcw4; by xwaveid;
data longcw4;
  merge masters (in=inc) longcw4;
  by xwaveid;
  if inc;
run;
%rename(w5c); run;
data longcw5;
  set w5c;
  %include 'c:\temp\rencmds.sas';
proc sort data=masters; by xwaveid;
proc sort data=longcw5; by xwaveid;
data longcw5;
  merge masters (in=inc) longcw5;
  by xwaveid;
  if inc;
run;

%rename(w6c); run;
data longcw6;
  set w6c;
  %include 'c:\temp\rencmds.sas';
proc sort data=masters; by xwaveid;
proc sort data=longcw6; by xwaveid;
data longcw6;
  merge masters (in=inc) longcw6;
  by xwaveid;
  if inc;
run;

* (2.2) Put datasets together to create the long longitudinal responding;
*       person file, and add a year variable.;
data longc;
  set longcw1 (in=w1) longcw2 (in=w2) longcw3 (in=w3) longcw4 (in=w4) longcw5 (in=w5) longcw6 (in=w6);
  * Create year;
  year=0;
  if (w1=1) then year=2001;
  if (w2=1) then year=2002;
  if (w3=1) then year=2003;
  if (w4=1) then year=2004;
  if (w5=1) then year=2005;
  if (w6=1) then year=2006;
run;

proc freq data = longc;
  table year;
run;

* (2.3) Save permanent dataset;
data long.longc;
  set longc;
*proc contents varnum data=long.longc;
run;
```

Figure 8: (c'td)

```
/* Supplementary programme to add longitudinal weights variables for Responding and Enumerated Persons
specifically to be used in a balanced file of 6 waves. */

data weights;
  set wave6.Combined_f60c;
  keep xwaveid flnwtrp flnwte;
  rename flnwtrp=lnwtrp flnwte=lnwte;
run;
data longc;
  set long.longc;
  drop lnwte lnwtrp;
run;

proc sort data=weights; by xwaveid;
proc sort data=longc; by xwaveid;

data long.longfinal;
   merge longc (in=inlong) weights;
   by xwaveid;
   if inlong;
run;

data test;
  set long.longfinal;
  keep xwaveid year lnwte lnwtrp;
run;

proc sort data=test; by year;
proc means; var lnwtrp; by year;
run;
```

Figure 9: SPSS program to create long longitudinal files

```
/*  Example SPSS program to create balanced responding person */
/*  long longitudinal file. */
/*  Created by: Simon Freidin 02/12/05 (Updated by: Nicole Watson) */

/* ************************************************************** */
/*   Section 1: locate datasets, get balanced cases, load renaming macro */
/* ************************************************************************** */
set printback=listing.
* (1.1) Files
* (1.1.1) Input: combined files.
* You MUST edit the name paths in ALL of Sections 1.1.1 to Section 1.1.4
file handle w1c     /name='e:\a\written datasets\a60\SPSS a60c\Combined a60c.sav'.
file handle w2c     /name='e:\b\written datasets\b60\SPSS b60c\Combined b60c.sav'.
file handle w3c     /name='e:\c\written datasets\c60\SPSS c60c\Combined c60c.sav'.
file handle w4c     /name='e:\d\written datasets\d60\SPSS d60c\Combined d60c.sav'.
file handle w5c     /name='e:\e\written datasets\e60\SPSS e60c\Combined e60c.sav'.
file handle w6c     /name='e:\f\written datasets\f60\SPSS f60c\Combined f60c.sav'.
file handle master  /name='e:\f\written datasets\f60\SPSS f60c\Master f60c.sav' .
* (1.1.2) Output: generated renaming program, balanced master file.
file handle renprog  /name='e:\xwave\rename.sps'.
file handle balanced /name='e:\xwave\balanced.sav'.
* (1.1.3) Output: intermediate longitudinal files.
file handle w1l     /name='e:\xwave\w1l.sav'.
file handle w2l     /name='e:\xwave\w2l.sav'.
file handle w3l     /name='e:\xwave\w3l.sav'.
file handle w4l     /name='e:\xwave\w4l.sav'.
```

Figure 9: (c'td)

```
file handle w5l      /name='e:\xwave\w5l.sav'.
file handle w6l      /name='e:\xwave\w6l.sav'.
* (1.1.4) Output: final balanced long responding person file.

file handle w1w6long /name='e:\xwave\w1w6long.sav'.

* (1.2) Identify cases responding in all waves from the master file.
*       status codes: 1=face to face interview, 2=telephone interview.

*       Release 6 balanced panel, master n = 8864.
get file=master.
select if any(afstatus,1,2) and any(bfstatus,1,2) and any(cfstatus,1,2) and any(dfstatus,1,2) and
any(efstatus,1,2) and any(ffstatus,1,2).
sort cases by xwaveid.
save outfile=balanced/keep=xwaveid ahhrpid bhhrpid chhrpid dhhrpid ehhrpid fhhrpid.

* (1.3) Load macro to write a program to rename variables, dropping first
*       character (wave identifier) of variable name. ID's are not renamed.
define renwave (!pos=!cmdend).
get file=!1.
n of cases 1.
set printback=off /results=no.
flip.
set printback=on /results=yes.
compute x=1.
match files file=*/last=last/by x.
string out (a60).
do if $casenum=1.
write outfile=renprog/"rename variables".
end if.
/* exclude ids from renaming */
do if (index(lower(case_lbl),'waveid')>0) or
     (index(lower(case_lbl),'hhrhid')>0) or
     (index(lower(case_lbl),'hhraid')>0) or
     (index(lower(case_lbl),'hhrpid')>0) .
compute out=concat(" (",case_lbl," = ",lower(case_lbl),")").
else.
compute out=concat(" (",lower(case_lbl)," = ",substr(lower(case_lbl),2),")").
end if.
write outfile=renprog/out.
do if (last).
write outfile=renprog/" .".
end if.
exe.
!enddefine.
/* end of macro */

/* ************************************************************************** */
/*   Section 2: create "long" balanced responding person longitudinal file   */
/* ************************************************************************** */
* (2.1) Rename each wave: run macro "renwave" and then include the generated rename program.
*       command file. Keep only those records from balanced master file.

renwave w1c .
get file=w1c.
include file=renprog.
sort cases by xwaveid.
match files file=balanced/in=inm1/file=*/by xwaveid.
select if (inm1).
```

Figure 9: (c'td)

```
save outfile=w1l.

renwave w2c .
get file=w2c.
include file=renprog.
sort cases by xwaveid
match files file=balanced/in=inm2/file=*/by xwaveid.
select if (inm2).
save outfile=w2l.

renwave w3c .
get file=w3c.
include file=renprog.
sort cases by xwaveid.
match files file=balanced/in=inm3/file=*/by xwaveid.
select if (inm3).
save outfile=w3l.

renwave w4c .
get file=w4c.
include file=renprog.
sort cases by xwaveid.
match files file=balanced/in=inm4/file=*/by xwaveid.
select if (inm4).
save outfile=w4l.

renwave w5c .
get file=w5c.
include file=renprog.
sort cases by xwaveid.
match files file=balanced/in=inm5/file=*/by xwaveid.
select if (inm5).
save outfile=w5l.

renwave w6c .
get file=w6c.
include file=renprog.
sort cases by xwaveid.
match files file=balanced/in=inm6/file=*/by xwaveid.
select if (inm6).
save outfile=w6l.

* (2.3) Add datasets together to create the balanced long longitudinal
*       responding person file. Add a year variable. Balanced panel long
add files file=w1l/in=w1/file=w2l/in=w2/file=w3l/in=w3/file=w4l/in=w4/file=w5l/in=w5/file=w6l/in=w6.
if (w1=1) year=2001.
if (w2=1) year=2002.
if (w3=1) year=2003.
if (w4=1) year=2004.
if (w5=1) year=2005.
if (w6=1) year=2006.

save outfile=w1w6long.

freq vars=year.
```

Figure 9: (c'td)

```
/*(2.4) Supplementary programme to create longitudinal weights variables for Responding and Enumerated
Persons specifically to be used in a balanced file of 6 waves. */

get file=w1w6long/drop=lnwtrp lnwte.
sort cases by xwaveid.
match files file=*table=w6c/rename=(flnwtrp flnwte= lnwtrp lnwte)/keep=xwaveid to year lnwtrp lnwte/by
xwaveid.
means vars=lnwtrp by year.
save outfile=w1w6long_1.
```

Figure 10: SAS macro to strip the first letter from the variable name

```
%macro HILDARenameStripFirst(dataset);
/* ---------------------------------------------------------------------------
   Generates a rename statement of the form
   Xvar1=var1 Xvar2=var2 etc
   where XvarN includes all variables (except Xwaveid).

   Usage example when setting up a long file:
   data long2;
     set hilda.combined_a (rename=(%HILDARenameStripFirst(hilda.combined_a)) in=in1)
         hilda.combined_b (rename=(%HILDARenameStripFirst(hilda.combined_b)) in=in2);
     if in1 then Wave=1;
     else if in2 then Wave=2;
     keep xwaveid hhrhid wave;
   run;
   proc sort data=long2;by xwaveid wave;run;

   Bruce Bradbury Aug07.
   Tested on SAS version 9.1.3
   I used SAS paper 107-28 by Derek Morgan as inspiration for the sysfunc calls.
--------------------------------------------------------------------------- */
%let DS = %sysfunc(open(&dataset,i)); /* open dataset to get variable names */
%if (&DS = 0) %then %put %sysfunc(sysmsg()); /* if can't open */
%else
  %do i=1 %to %sysfunc(attrn(&DS,NVARS));
    %let varname = %sysfunc(varname(&DS,&i)); /* get the ith variable name */
    %let newvarname = %substr(&varname,2);    /* strip off the first character */
       %if %upcase(&varname)^=XWAVEID %then %do;
      %* output code here;
      &varname=&newvarname
        %end;
  %end;
%let CloseCode = %sysfunc(close(&DS));
%mend HILDARenameStripFirst;
```

Figure 11: STATA program to create long longitudinal files

```
*/*************************************************************************************************
*  Example STATA program to create Unbalanced & Balanced responding person
*  long longitudinal files.
*  Created by: Alison Goode  04/01/06  (Updated by: Nicole Watson and Clinton Hayes)
*  Note: Users will need to restrict the variables used as memory allocation is not sufficient for 6 waves.
*************************************************************************************************/
/*Section 1; create a long UNBALANCED dataset – if not all variables required, select and save as a
separate file, then proceed.*/
#delimit cr
set mem 800000
set maxvar 10000
clear
```

Figure 11: (c'td)

```
use "C:\RELEASE 6.0\Master_f60c.dta"
/* Identify cases responding in all waves from the master file using status codes (_fstatus):
   1=face to face interview, 2=telephone interview and create a variable identifying them*/
sort xwaveid
#delimit ;
by xwaveid:gen totwave=6 if (afstatus==1|afstatus==2)&(bfstatus==1|bfstatus==2)&
(cfstatus==1|cfstatus==2)&(dfstatus==1|dfstatus==2)&(efstatus==1|efstatus==2)&(ffstatus==1|ffstatus==2);
label variable totwave "present-6waves";
keep xwaveid totwave ahhrpid bhhrpid chhrpid dhhrpid ehhrpid fhhrpid afstatus bfstatus cfstatus dfstatus
efstatus ffstatus;
#delimit cr
sort xwaveid
save "C:\RELEASE 6.0\My_Masterfile_f60c.dta", replace
clear
/*Remove prefix from each of the combined files and generate new variable (wave) to prepare to merge
 files with the masterfile*/
use "C:\RELEASE 6.0\Combined_a60c.dta"
renpfix a
gen wave=1
sort xwaveid
save "C:\RELEASE 6.0\Combined_wave1.dta",replace
clear
use "C:\RELEASE 6.0\Combined_b60c.dta"
renpfix b
gen wave=2
sort xwaveid
save "C:\RELEASE 6.0\Combined_wave2.dta", replace
clear
use "C:\RELEASE 6.0\Combined_c60c.dta"
renpfix c
gen wave=3
sort xwaveid
save "C:\RELEASE 6.0\Combined_wave3.dta", replace
clear
use "C:\RELEASE 6.0\Combined_d60c.dta"
renpfix d
gen wave=4
sort xwaveid
save "C:\RELEASE 6.0\Combined_wave4.dta", replace
clear
use "C:\RELEASE 6.0\Combined_e60c.dta"
renpfix e
gen wave=5
sort xwaveid
save "C:\RELEASE 6.0\Combined_wave5.dta", replace
clear
use "C:\RELEASE 6.0\Combined_f60c.dta"
renpfix f
gen wave=5
sort xwaveid
save "C:\RELEASE 6.0\Combined_wave6.dta", replace
clear

*Now create a single long file
use "C:\RELEASE 6.0\My_Masterfile_f60c.dta"
merge xwaveid using "C:\RELEASE 6.0\combined_wave1.dta"
sort xwaveid
drop _merge
append using "C:\RELEASE 6.0\combined_wave2.dta"
```

Figure 11: (c'td)

```
sort xwaveid
append using "C:\RELEASE 6.0\combined_wave3.dta"
sort xwaveid
append using "C:\RELEASE 6.0\combined_wave4.dta"
sort xwaveid wave
append using "C:\RELEASE 6.0\combined_wave5.dta"
sort xwaveid wave
append using "C:\RELEASE 6.0\combined_wave6.dta"
sort xwaveid wave
tab wave
tab totwave
by xwaveid:egen totwave1=sum(totwave)

save "C:\RELEASE 6.0\Masterfile_Unbalanced.dta", replace
clear

*/Section 2: To convert to a balanced panel
* Release 6 balanced panel, n = 8864/*
use "C:\RELEASE 6.0\Masterfile_Unbalanced.dta"
sort xwaveid wave
by xwaveid:keep if totwave1==6
tab wave

*/* Supplementary programme to create longitudinal weights variables for Responding and Enumerated
Persons specifically to be used in a balanced file of all 6 waves. */*

destring xwaveid, gen(id)
sort id wave
by id:egen flnwte_1=max(lnwte) if wave==6
by id:egen flnwte_2=max(flnwte_1)
drop flnwte_1 lnwte
rename flnwte_2 lnwte

sort id wave
by id:egen flnwtrp_1=max(lnwtrp) if wave==6
by id:egen flnwtrp_2=max(flnwtrp_1)
drop flnwtrp_1 lnwtrp
rename flnwtrp_2 lnwtrp
sort wave
save "C:\RELEASE 6.0\Masterfile_Balanced.dta", replace
clear
```

Figure 12: SAS program to create wide longitudinal file

```
/*Example SAS program to create unbalanced and balanced wide files*/
/*Created by Paul Agius 2/11/2006 (Updated by: Nicole Watson)*/

/*Setting libraries*/
libname wave1 'E:\Release 6.0\a60c'; run;
libname wave2 'E:\Release 6.0\b60c'; run;
libname wave3 'E:\Release 6.0\c60c'; run;
libname wave4 'E:\Release 6.0\d60c'; run;
libname wave5 'E:\Release 6.0\e60c'; run;
libname wave6 'E:\Release 6.0\f60c'; run;

/*Load the library of formats required for the datasets*/
libname library 'E:\Release 6.0\formats'; run;

/*Sorting files by xwaveid for merging*/
proc sort data = wave1.combined_a60c out=w1;
```

Figure 12: (c'td)

```
by xwaveid;
proc sort data = wave2.combined_b60c out=w2;
by xwaveid;
proc sort data = wave3.combined_c60c out=w3;
by xwaveid;
proc sort data = wave4.combined_d60c out=w4;
by xwaveid;
proc sort data = wave5.combined_e60c out=w5;
by xwaveid;
proc sort data = wave6.combined_f60c out=w6;
by xwaveid;
proc sort data = wave6.master_f60c out=master;
by xwaveid;
run;

/*Creating unbalanced data file*/
data wide_UNBALANCED;
  merge w1 w2 w3 w4 w5 w6 master;
  by xwaveid;
run;

/*Creating balanced data file for interviewed persons*/
data wide_BALANCED;
  set wide_UNBALANCED;
  if        (afstatus = 1 or afstatus = 2) and
                        (bfstatus = 1 or bfstatus = 2) and
                        (cfstatus = 1 or cfstatus = 2) and
                        (dfstatus = 1 or dfstatus = 2) and
                        (efstatus = 1 or efstatus = 2) and
                        (ffstatus = 1 or ffstatus = 2);
run;
```

Figure 13: SPSS program to create wide longitudinal file

```
/* ********************************************************************** */
/* Create wide longitudinal panel */
/* Files taken directly from CD (presorted by xwaveid) */
/* Created by: Simon Freidin 2006 (Updated by: Nicole Watson) */
/* ********************************************************************** */

match files
  file='E:\f\written datasets\f60\SPSS f60c\Master f60c.sav'  /in=inmaster
 /file='E:\a\written datasets\a60\SPSS a60c\Combined a60c.sav'/in=inwave1
 /file='E:\b\written datasets\b60\SPSS b60c\Combined b60c.sav'/in=inwave2
 /file='E:\c\written datasets\c60\SPSS c60c\Combined c60c.sav'/in=inwave3
 /file='E:\d\written datasets\d60\SPSS d60c\Combined d60c.sav'/in=inwave4
 /file='E:\e\written datasets\e60\SPSS e60c\Combined e60c.sav'/in=inwave5
/file='E:\f\written datasets\f60\SPSS f60c\Combined f60c.sav'/in=inwave6
 /by xwaveid.

/* choose one selection option */
/* no 'select if' gives an unbalanced (enumerated) panel: 24856 cases */

/*  interviewed all 6 waves: 8864 cases */
*select if (ahgint) and (bhgint) and (chgint) and (dhgint) and (ehgint) and (fhgint).

/* enumerated all 6 waves: 12798 cases */
*select if (ahgenum) and (bhgenum) and (chgenum) and (dhgenum) and (ehgenum) and (fhgenum).

save outfile='e:\xwave\w1w6c.sav'.
```

Figure 14: STATA program to create wide longitudinal file

```
*/*Example STATA program to create Unbalanced and Balanced Wide File
Created by: Alison Goode 27/09/06 (Updated by: Nicole Watson)*/
*/*********************************************************************************************************/
*/*Please note - uses all variables. If not all variables required, select and save as a separate file,
   then proceed.
  Please note: same variables for each consecutive wave NOT next to one another. To organise variables
this way.  If required, use the AORDER or ORDER command.
**********************************************************************************************************/
set maxvar 15000

*/Before merging the master file and combined files, sort xwaveid in each file and save each of them*/
use "E:\Release 6.0\Masterfile_f60c.dta"
sort xwaveid
merge xwaveid using "E:\Release 6.0\Combined_a60c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 6.0\Combined_b60c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 6.0\Combined_c60c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 6.0\Combined_d60c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 6.0\Combined_e60c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 6.0\Combined_f60c.dta"
sort xwaveid
drop _merge
save "E:\Release 6.0\Wide-UNBALANCED_6waves.dta"
*/*************************************************************************
*Section 2: To convert to a balanced wide file
 n==8864
**************************************************************************/
use "E:\Release 6.0\Wide-UNBALANCED_6waves.dta"
sum afstatus bfstatus cfstatus dfstatus efstatus ffstatus
sort xwaveid
#delimit ;
by xwaveid:keep if
(afstatus==1|afstatus==2)&(bfstatus==1|bfstatus==2)&(cfstatus==1|cfstatus==2)&(dfstatus==1|dfstatus==2)
&(efstatus==1|efstatus==2)&(ffstatus==1|ffstatus==2);
#delimit cr
sum afstatus bfstatus cfstatus dfstatus efstatus ffstatus
save "E:\Release 6.0\Wide-BALANCED_6waves.dta"
drop _all
```

PanelWhiz

PanelWhiz is a collection of STATA/SE Add-On scripts to make using panel datasets easier. The package allows the user to easily select vectors of variables across waves. Matching and merging of the files is done automatically to create a long file (with one record per person per wave). More information and how to order PanelWhiz can be found at: http://www.melbourneinstitute.com/hilda/doc/doc_panelwhiz.htm

Variable Name Conventions

Variable names have been limited to eight characters (so that the files can be read in older versions of SPSS and SAS). The variable name is divided into three parts and attempts to provide information on the content of the variables:

- First character – wave identifier, with ‘a’ being used for wave 1, ‘b’ for wave 2, ‘c’ for wave 3, etc.
- Second and third character – general subject area (see Table 6) for the conventions).
- Fourth to eighth character – specific subject of data item.

Excluding the first character, variable names are the same across waves if the question and response options are the same. If the question or response options have significantly changed, the variable name will also be modified. There are, however, a few variables where we have decided to vary from this convention:

- Household response status;
- Person response status;
- SCQ in-field response flag;
- Household membership; and
- New location of mover.

For these variables, it was thought more important to keep the same variable names. These variables are used for survey administration purposes by the HILDA Survey team at the Melbourne Institute. Many users will not use the detail in these variables. Table 7 to Table 11 show how the response categories differ for these variables across the first six waves.

Table 6: Broad subject area naming conventions, characters 2 and 3 (sorted by code)

Code	Broad Subject Area	Code	Broad Subject Area	Code	Broad Subject Area
AL	Leave	HH	Household information, identifiers and cross-sectional weights	OP	Other property
AN	Ancestry			OR	Other relationships
AT	Attitudes and values			PA	Parenting
BA	Bank accounts	HI	Household income	PH	Private health insurance
BI	Business income	HS	Housing		
BF	Business	HW	Household wealth	PJ	Previous job
BM	Body mass index	HX	Household expenditure	PI	Personal information
BN	Benefits	IC	Intentions to have children	PN	Personality
CA	Calendar			PR	Partnering / relationships
CC	Childcare general	IO	Interviewer obs		
CH	Childcare during school holidays	IP	Intentions and plans	PW	Personal wealth
		JB	Job characteristics of employed	RC	Resident children
CN	Non employment related childcare			RE	Religion
		JO	Opinions about job	RG	Relationship grid
CP	Childcare for children not yet at school	JS	Job search of those not employed	RP	Residential property
				RT	Retirement intentions
CR	Credit cards	LE	Major life events	RW	Replicate weight
CS	Child care during school terms	LN	Longitudinal weights	SA	Superannuation
		LO	Life opinions	TC	Total children
DB	Debt	LS	Lifestyle	TI	Total income
DO	Dwelling observations	MH	Moving house	TS	Time stamps
DT	Personal debt	MO	Mutual obligations	TX	Taxes
ED	Education	MR	Marital relationships	UJ	Job history of those not in paid employment
EH	Employment history	MV	Motor vehicles		
ES	Employment status	NC	Non-resident children	WC	Workers compensation
FA	Financial assets	NL	Not in labour force	WS	Wage and salaries
FI	Attitudes to finances	NP	Non-employment related childcare for children not yet at school	XP	Expenditure
FM	Family background			YE	Youth - employment
GH	General health and well-being			YH	Youth - education
		NS	Non-employment related childcare for children at school	YI	Youth - importance
HB	Household bills	OA	Other assets	YP	Youth - property
HE	Health	OI	Other income	YS	Youth - life satisfaction
HG	H'h enumeration grid				

Table 7: Unchanged variable name and differing codes for Household Response Status

Description (applies to final _hhresp, initial _hhrespi[1], follow-up _hhrespf[1])	*Codes used*			
	Wave 1	*Wave 2*	*Wave 3*	*Wave 4 onwards*
Full Response				
Every eligible member of current HH interviewed	62	62	62	62
Part Response				
Part refused	63	63	63	63
Part non-contact	64	64	64	64
Part contact made with all non-response	65	65	65	65
Part away for workload period	66	66	66	66
Part language problem	67	67	67	67
Part incapable/death/illness	68	68	68	68
Non-Response				
Refusal	69			
Refusal - PSMs still live there	N/A	69	69	69
Refusal - Dont know if PSMs still live there	N/A	70	70	70
Address occupied - no contact with a sample member	70	71	71	71
Contact made and all calls made	71	72	72	72
All residents away for workload period	72	73	73	73
HH does not speak English	73	74	74	74
HH incapable/illness	74	75	75	75
Refusal to Nielsen via 1800 number	75	76	76	76
Terminate (no PQs)	76	77	77	77
HH deceased	N/A	78	78	78
HH moved out of scope	N/A	79	79	79
All PSMs moved in with another PSM	N/A	80	80	80
All PSMs non respondents in last 2 waves	N/A	N/A	81	81
Not in Area/No phone number				82
Untraceable	N/A	99	99	99*
Not issued this wave	N/A	N/A	100	100
Deceased at previous wave	N/A	N/A	101	101
TSM no longer living with PSM at previous wave				102
Dwelling out of scope				
Dwelling vacant for workload period	77	N/A	N/A	N/A
Non-private dwelling - place of business	78	N/A	N/A	N/A
Used for temporary accommodation only	79	N/A	N/A	N/A
Institution with no private HH usually resident	80	N/A	N/A	N/A
Not a main residence (eg holiday home)	81	N/A	N/A	N/A
All people in household out of scope	82	N/A	N/A	N/A
Derelict dwelling/demolished/to be demolished	83	N/A	N/A	N/A
Dwelling under construction/unliveable renovations	84	N/A	N/A	N/A
Listing error	85	N/A	N/A	N/A

1. *_hhrespi* and *_hhrespf* are only on the In-confidence Release (unconfidentialised) files. For initial response status *_hhrespi*, subtract 60 from all codes except 98 and 99. For followup response status *_hhrespf*, substract 30 from all codes except 98 and 99.
* for *_hhrespi* only: Untraceable is coded 89.

Table 8: Unchanged variable name and differing codes for Person Response Status

Description (applies to _fstatus, initial_hgri and _hgri1 to hgri14; follow-up_hgrf and _hgrf1 to _hgrf14; final_hgivw and _hgivw1 to _hgivw14[1])	*Codes used*		
	Wave 1	*Wave 2*	*Wave 3 onwards*
Interview in person	1	1	1
Interviewed by telephone	2	2	2
Ineligible for interview			
Less than 15 years old at 30th of June	3	3	3
Overseas for more than 6 months		4	4
In prison		5	5
TSM no longer living with PSM	N/A	N/A	6
Not part of the household NFI	4		
Refusal			
Too busy	12	6	7
Too invasive	11	7	8
Other reasons	13	8	9
Refusal via 1800 number/email	14	9	10
Interview terminated	15	10	11
Other non-interview			
Deceased	N/A	11	12
Moved to another HF	N/A	12	13
Language problem	6	13	14
Incapable/illness/infirmity	5	14	15
Home but unable to contact	9	15	16
Away for workload period	8	16	17
Away at boarding school/university	7		
Other reasons	10		
Household non-contact	N/A	17	18
Household contact made no interviews	N/A	18	19
Household not issued to field – persistent non-respondent	N/A	N/A	20
Overseas permanently	N/A		21
Household all PSMs non-responding in last 2 waves	N/A	N/A	22
Permanently incapable from previous wave	N/A		23
Household out of scope NFI	N/A	19	
Overseas and aged < 15	N/A	20	28
Untraceable from prior waves	N/A	N/A	29
Untraceable determined this wave	N/A	99	99

1. *_hgri, _hgri1 to _hgri14, _hgrf and _hgrf1 to _hgrf14* are only on the In-confidence Release (unconfidentialised) files.

Table 9: Unchanged variable name and differing codes for SCQ Field Response Status

Description (applies to _hgsi, _hgsi1 to _hgsi14, _hgsf, _hgsf1 to _hgsf14, _hgscq, _hgscq1 to _hgscq14[1])	*Codes used*	
	Wave 1	*Wave 2 onwards*
Picked up	1	1
To be sent	3	2
Refused	2	3
Not given	4	4

1. *_hgsi*, *_hgsi1* to *_hgsi14*, *_hgsf*, and *_hgsf1* to *_hgsf14* are only on the In-confidence Release (unconfidentialised) files.

Table 10: Unchanged variable name and differing codes for Household Membership

Description (applies to _hghhm, _hghhm1 to _hghhm14[1])	*Codes used*		
	Wave 1	*Wave 2*	*Wave 3 onwards*
Listed			
Resident	N/A	1	1
Absent for workload	N/A	2	2
No longer member of household	N/A	3	3
Deceased	N/A	4	4
Not listed			
Re-joiner/merger	N/A		5
New resident	N/A	5	6
Absent for workload new resident	N/A	6	7

1. For *_hghhm*, the value labels are quite different, but the meaning of many of the codes are the same. Wave 3 value labels are listed in this table.

Table 11: Unchanged variable name and differing codes for New Location of mover

Description (applies to _hgnlc1 to _hgnlc14)	*Codes Used*		
	Wave 1	*Wave 2*	*Wave 3 onwards*
Overseas	N/A	1	
Within Australia – new local address	N/A	2	1
Within Australia – new non-local address	N/A	3	2
Address unknown	N/A	4	3
Deceased	N/A	5	4
Overseas permanently	N/A		5
Overseas but not permanently	N/A		6

Missing Value Conventions

Global codes are used throughout the dataset to identify missing data. These codes are not restated for each variable in the coding framework.

Numeric Variables

All missing numeric data are coded into the following set of negative values shown in Table 12.

Table 12: Missing value conventions for numeric variables

Code	*Description*
-1	Not asked: question skipped due to answer to a preceding question
-2	Not applicable
-3	Don't know
-4	Refused or not answered
-5	Invalid multiple response (SCQ only)
-6	Value implausible (as determined after intensive checking)
-7	Unable to determine value
-8	No Self-Completion Questionnaire returned and matched to individual record
-9	Non-responding household
-10	Non-responding person (Combined File only)

Note that the SPSS files have these global missing values (-10 to -1) set to SPSS user-defined missing. To turn off this setting for an individual variable use "MISSING VALUES varname1 ()." To turn off this setting for all variables (for example, if you need to include those who are coded as -1 'Not asked') use the following code:

```
set errors=none.
do repeat x=all.
missing values x ().
end repeat.
set errors=listing.
execute.
```

Text Variables

Text variables with missing values will typically contain the following text (as shown in Table 13).

Table 13: Missing value conventions for text variables

Text	*Description*
[blank]	Missing information (no reason specified)
-1	Not asked
-2	Not applicable
-3	Don't know
-4	Refused
-7	Unable to determine value
-9	Non-responding household
99	Not given (new person)

Data With Negative Values

Data items that can have both negative and positive values, such as business income, total household income, etc, are provided as two variables:

- the variable for positive amounts; and
- the variable for negative amounts.

If the overall value is not missing and is positive, then the negative variable will be zero and the positive variable will hold the actual value. If the overall value is not missing and is negative, then the positive variable will be zero and the negative variable will hold the absolute value of the amount. For example, if we have a person with a business income loss of $20,000 in the last financial year, then the positive variable of business income will be zero and the negative variable will be $20,000.

Missing data information will be provided in both variables following the negative conventions described above.

Therefore, after handling the missing data, you can create your own variable by subtracting the negative variable from the positive variable. For example, you might set the missing values of business income to system missing and then create a new business income variable as follows:

```
abifp-abifn.
```

Confidentialisation

The HILDA datasets released have been confidentialised to reduce the risk that individual sample members can be identified. This has involved:

- withholding some variables (such as postcode);
- aggregating some variables (for example, occupation has been provided at the two digit level while it is collected at the four digit level); and
- top-coding some variables (such as age, income and wealth variables).

Top-coding substitutes an average value for all the cases which are equal to or exceed a given threshold. The substituted value is calculated as the weighted average of the cases subject to top-coding. As a result, the cross-sectionally weighted means of the top-coded variable will be the same as the original variable. (In earlier releases, the cut-off value was used which failed to preserve the weighted means.)

Take, for example, the top-coding of *_wscg* (current gross wages per week in main job). All cases whose wages are equal or exceed $4800 have had their value replaced by the weighted average of all those cases whose income is equal to or exceeds $4800. Let us say that the weighted average of the 22 cases earning $4800 or more is $8450. $8450 is then substituted as the wages for those 22 cases. This maximizes confidentiality and preserves the weighted distribution means. If the distribution of wages had been simply cut off at $4800, when the relevant weights are applied, the value would be too low.

The top-coding thresholds are adjusted over time to overcome the tendency of income and wealth measures to inflate. Without adjustment, increasing numbers of cases would exceed the threshold and be topcoded. If you need to know the threshold values that have been used at a particular release, please contact hilda-inquiries@unimelb.edu.au.

DERIVED VARIABLES

Derived variables are created from the data in the following circumstances.

- When questions are asked in an easy-to-answer form which requires recombination to a common metric.
- When some 'other, specify' answers are coded (notably sources of other income).
- When a complex combination of data occurs (for example, family type).
- When open-ended answers are converted to standard codeframes (industry; occupation; post-school qualifications).
- When missing data are imputed.
- When external data are matched to derive applicable measures (for example, weighting; socio-economic indicators for areas; remoteness area).

Derived variables are created at both the household and person levels. Most derived variables are available in each wave. A description of how the variable was derived is supplied in the coding framework and additional information is provided in this user manual as necessary.

All derived variables have the prefix 'DV:' in the variable label. Missing values have the same codes as collected data. Derived variables are not annotated on the marked up questionnaires, but are included in the various coding frameworks.

Age and Sex

For each person interviewed, two ages have been provided:

- *_hgage* which is the age at last birthday as of 30 June immediately preceding the fieldwork for that wave (for wave 1, *ahgage* is the respondents age at 30 June 2001); and
- *_hhiage* which is the age at last birthday as of the date of interview for that wave (the interview dates for each wave spread over 6 to 8 months).

For non-interviewed people in responding households, *_hgage* is provided on the enumerated file.

On the household file, the age of each person as of the 30th of June is provided in the variables *_hgage1* to *_hgage16,* where *_hgage1* is for person 01, etc.[5]

For the small number of cases where age was not provided, it has been imputed via a hotdeck method[6] and *_hgagef, _hgagf1 to _hgag16* flags which cases have been imputed.

[5] *_hgage15* and *_hgage16* are only included in wave 6.
[6] The hotdeck method seeks to find a donor with a similar set of characteristics to the non-respondent.

Note that if the respondent provides a correction to the date of birth printed on the Household Form each wave, this correction is applied back through the previous waves. As a result the above calculated ages may change from one release to another (hopefully not by much!). This is why you may find some 14 year olds interviewed in a previous wave.

Similarly, if the respondent provides a correction to the sex printed on the Household Form each wave, this correction is applied back through the previous waves.

History Variables

History variables contain data accumulated across successive waves. Some history variables contain background information that does not change and is only asked in the first interview (e.g. country of birth). Others contain accumulated statuses (e.g. number of qualifications; current marriage duration). The variables are provided in the responding person file each wave from wave 2 onwards, and show the status at the completion of each wave.

History variables first have data in the year the respondent entered the survey, and are updated the next time the respondent is interviewed. Someone who was a new entrant at say wave 2, did not respond in wave 3 and was interviewed, again, in wave 4, will not have history data for waves 1 and 3, even for invariant information such as Country of Birth. Those using unbalanced panels will be particularly affected and may need to write a program to 'fill-in' the missing years.

History variables have the prefix ‘History:’ in the variable label. History variables are not annotated on the marked up questionnaires, but are included in the various coding frameworks. Table 14 provides a list of the history variables included on the datasets. Notes about the construction of the variables are included in the coding framework (and are not duplicated here).

Note that the job history variables (relating to the previous job for those that are employed or the last job for those that are not employed) that were first provided in Release 5 have been removed from the datasets as they require some more work.[7]

Table 14: History Variables

Variable	Description
Ancestry	
_ANCOB	History: Country of birth
_ANBCOB	History: Country of birth - brief
_ANYOA	History: Year first came to Australia to live
_ANENGF	History: Is English the first language you learned to speak as a child
_ANATSI	History: Aboriginal or Torres Strait Islander origin

[7] These variables were: _JHTSJHA, _JHWKU, _JHLHRU, _JHLHRUW, _JHLJIND, _JHLJTYP, _JHLJCNT, _JHLHTHA, _JHLJTWK, _JHLJTYR, _JHLJOCC, _JHLJREA, _JHLJOCS, _JHLJI88, _JHLJ182, _JHLJOC2, _JHLJII2.

Table 14: (c'td)

Variable	Description
Family background	
_FMLWOP	History: Were you living with both your own mother and father around the time you were 14 years old
_FMNPREA	History: Why were you not living with both your parents at age 14
_FMPDIV	History: Did your mother and father ever get divorced or separate
_FMPJOIN	History: Did your mother and father ever get back together again
_FMAGEPS	History: How old were you at the time your parents separated
_FMAGELH	History: How old were you when you first moved out of home as a young person
_FMHSIB	History: Ever had any siblings
_FMNSIB	History: How many siblings
_FMELDST	History: Were you the oldest child
_FMFCOB	History: Fathers country of birth
_FMMCOB	History: Mothers country of birth
_FMFEMP	History: Was father in paid employment when you were 14
_FMFOCC[1]	History: Fathers occupation - 4 digit ASCO
_FMFOCC2	History: Fathers occupation - 2 digit ASCO
_FMFI88[1]	History: ISCO-88, fathers occupation
_FMFI82	History: ISCO-88 2-digit, fathers occupation
_FMFOCCS	History: ANU4 occupational status scale, fathers occupation
_FMFUEMP	History: Was father unemployed for 6 months or more while you were growing up
_FMMEMP	History: Was mother in paid employment when you were 14
_FMMOCC[1]	History: Mothers occupation - 4 digit ASCO
_FMMOCC2	History: Mothers occupation - 2 digit ASCO
_FMMI88[1]	History: ISCO-88, mothers occupation
_FMMI82	History: ISCO-88 2-digit, mothers occupation
_FMMOCCS	History: ANU4 occupational status scale, mothers occupation
Education	
_EDAGELS	History: Age left school
_EDHISTS	History: Highest level of school completed/currently attending
_EDTYPES	History: Type of school attended/attending
_EDCLY	History: Country of last school year
_EDQENR	History: Ever enrolled in a course of study to obtain a qualification
_EDCOQ	History: Country completed highest qualification in
_EDQ100	History: ASCED 100 Postgraduate - Number of qualifications obtained since leaving school
_EDQ110	History: ASCED 110 Doctoral degree - Number of qualifications obtained since leaving school

Table 14: (c'td)

Variable	Description
_EDQ120	History: ASCED 120 Master degree - Number of qualifications obtained since leaving school
_EDQ200	History: ASCED 200 Grad Dip and Grad Cert - Number of qualifications obtained since leaving school
_EDQ221	History: ASCED 221 Graduate certificate - Number of qualifications obtained since leaving school
_EDQ211	History: ASCED 211 Graduate diploma - Number of qualifications obtained since leaving school
_EDQ310	History: ASCED 310 Bachelor degree - Number of qualifications obtained since leaving school
_EDQ311	History: ASCED 311 Bachelor (Honours) degree - Number of qualifications obtained since leaving school
_EDQ312	History: ASCED 312 Bachelor (Pass) degree - Number of qualifications obtained since leaving school
_EDQ400	History: ASCED 400 Advanced diploma and diploma - Number of qualifications obtained since leaving school
_EDQ413	History: ASCED 413 Associate degree - Number of qualifications obtained since leaving school
_EDQ421	History: ASCED 421 Diploma - Number of qualifications obtained since leaving school
_EDQ411	History: ASCED 411 Advanced diploma - Number of qualifications obtained since leaving school
_EDQ500	History: ASCED 500 Certificate - dont know level - Number of qualifications obtained since leaving school
_EDQ524	History: ASCED 524 Certificate level I - Number of qualifications obtained since leaving school
_EDQ521	History: ASCED 521 Certificate level II - Number of qualifications obtained since leaving school
_EDQ514	History: ASCED 514 Certificate level III - Number of qualifications obtained since leaving school
_EDQ511	History: ASCED 511 Certificate level IV - Number of qualifications obtained since leaving school
_EDQ600	History: ASCED 600 Secondary education - Number of qualifications obtained since leaving school
_EDQ611	History: ASCED 611 Year 12 - Number of qualifications obtained since leaving school
_EDQUNK	History: Unknown - not enough information - Number of qualifications obtained since leaving school
_EDHIGH	History: DV: Highest education level achieved
Marriage and Defacto Relationships	
_MRN	History: How many times have you been legally married
_MRPMTH	History: Month - present or most recent marriage
_MRPYR	History: Year - present or most recent marriage
_MR1YR	History: Year - First marriage if married more than once
_MR2YR	History: Year - Second marriage if married more than twice

Table 14: (c'td)

Variable	Description
_MR3YR	History: Year - Third marriage if married more than 3 times
_MR4YR	History: Year - Fourth marriage if married more than 4 times
_MRPLV	History: Live together before marriage - Present or most recent marriage
_MR1LV	History: Live together before marriage - First marriage if married more than once
_MR2LV	History: Live together before marriage - Second marriage if married more than twice
_MR3LV	History: Live together before marriage - Third marriage if married more than 3 times
_MR4LV	History: Live together before marriage - Fourth marriage if married more than 4 times
_MRPEND	History: How did the marriage end - present or most recent marriage
_MR1END	History: How did the marriage end - First marriage if married more than once
_MR2END	History: How did the marriage end - Second marriage if married more than twice
_MR3END	History: How did the marriage end - Third marriage if married more than 3 times
_MR4END	History: How did the marriage end - Fourth marriage if married more than 4 times
_MRPWIDW	History: Year widowed - Present or most recent marriage
_MR1WIDW	History: Year widowed - First marriage if married more than once
_MR2WIDW	History: Year widowed - Second marriage if married more than twice
_MR3WIDW	History: Year widowed - Third marriage if married more than 3 times
_MR4WIDW	History: Year widowed - Fourth marriage if married more than 4 times
_MRPSEP	History: Year separated - Present or most recent marriage
_MR1SEP	History: Year separated - First marriage if married more than once
_MR2SEP	History: Year separated - Second marriage if married more than twice
_MR3SEP	History: Year separated - Third marriage if married more than 3 times
_MR4SEP	History: Year separated - Fourth marriage if married more than 4 times
_MRPDIV	History: Year divorced - Present or most recent marriage
_MR1DIV	History: Year divorced - First marriage if married more than once
_MR2DIV	History: Year divorced - Second marriage if married more than twice
_MR3DIV	History: Year divorced - Third marriage if married more than 3 times
_MR4DIV	History: Year divorced - Fourth marriage if married more than 4 times
_ORDFPST	History: Ever lived with someone for at least one month without marrying
_ORDFN	History: Number of times lived in a defacto relationship
_MRPLVT	History: Years living together before present marriage
_MR1LVT	History: Years living together before first marriage
_MR2LVT	History: Years living together before second marriage
_MR3LVT	History: Years living together before third marriage
_MR4LVT	History: Years living together before fourth marriage
_ORCDUR	History: Current defacto duration - years
_MRCDUR	History: Current marriage duration - years
_MRWDUR	History: Current widow duration - years
_MRSDUR	History: Current separated or divorced from date of separation - years

Table 14: (c'td)

Variable	Description
Children	
_TCHAD[2]	History: Total children ever had
_TCDIED[2]	History: Total children since died
Employment	
_RTAGE	History: Age retired/intends to retire
_EHTSE	History: Time since FT education - years
_EHTJB	History: Time in paid work - years
_EHTUJ	History: Time unemployed and looking for work - years
_EHTO	History: Time not working and not looking for work - years
Health[3]	
_HESPNCY	History: Year condition first developed – Sight problems not corrected by glasses/lenses
_HEHEARY	History: Year condition first developed – Hearing problems
_HESPCHY	History: Year condition first developed – Speech problems
_HEBFLCY	History: Year condition first developed – Blackouts, fits or loss of consciousness
_HESLUY	History: Year condition first developed – Difficulty learning or understanding things
_HELUAFY	History: Year condition first developed – Limited use of arms or fingers
_HEDGTY	History: Year condition first developed – Difficulty gripping things
_HELUFLY	History: Year condition first developed – Limited use of feet or legs
_HENECY	History: Year condition first developed – A nervous or emotional condition which requires treatment
_HECRYPAY	History: Year condition first developed – Any condition which restricts physical activity or physical work (e.g. back problems, migraines)
_HEDISFY	History: Year condition first developed – Any disfigurement or deformity
_HEMIRHY	History: Year condition first developed – Any mental illness requiring help or supervision
_HESBDBY	History: Year condition first developed – Shortness of breath or difficulty breathing
_HECRPY	History: Year condition first developed – Chronic or recurring pain
_HEHIBDY	History: Year condition first developed – Long term effects as a result of a head injury, stroke or other brain damage
_HEMEDY	History: Year condition first developed – Long term condition or ailment which is still restrictive even though it is being treated or medication is being taken for it
_HEOTHY	History: Year condition first developed – Other long term condition such as arthritis, asthma, heart disease, Alzheimers disease, dementia etc)
Housing	
_HSYRCAD	History: Years at current address

1. Variables are only on the In-confidence Release (unconfidentialised) files.
2. For these variables, 'children' refers to the respondent's natural and adopted children.
3. Wave 3 onwards.

Spatial Variables

From Release 5, the household addresses from each wave have been geocoded and assigned a 2001 Census Collection District (CD). Where the address details were not sufficient to geocode exactly, the nearest cross section or street segment was used. Further, some fuzzy matching and manual look-up of maps were employed where the street name or suburb did not provide a reasonable match. We are able to build up from CD level to the following geographic regions:

- Statistical Local Area (SLA);
- Local Government Area (LGA);
- Statistical Sub-Division (SSD);
- Statistical Division (SD);
- Section of State (SOS); and
- Major Statistical Region (MSR).

The General Release (confidentialised) HILDA files do not include geographical descriptors for CD, postcode, SLA, LGA, SSD and SD. These files only include State, Section of State and MSR. The In-confidence Release (unconfidentialised) files include all geographical descriptors mentioned above.

Table 15 lists the derived spatial variables. Aside from the area identifiers, several other spatial variables are included on the file such as:

- Remoteness area – this is derived based on the assigned SLA;
- Socio-Economic Indexes for Areas (SEIFA) – deciles are assigned for four types of SEIFA scores based on the assigned SLA;
- The distance moved from the last wave – this is calculated from the geocoded addresses. Where the geocoding had to be approximated and the household moves close by, there may be some households who have moved but the distance moved is calculated as zero.

Other related spatial variables which are not derived that you should be aware of are State (*_hhstate*) and whether the household has moved from the last wave (*_hhmove*).

Table 15: Derived spatial variables

Variable	Description
_HHSLA[1]	DV: Statistical Local Area
_HHSLA9[1]	DV: 9 Digit Statistical Local Area
_HHLGA[1]	DV: Local Government Area
_HHSSD[1]	DV: Statistical Subdivision
_HHSD[1]	DV: Statistical Division
_HHMSR	DV: Major Statistical Region
_HHSOS	DV: Section of State

Table 15 (c'td)

Variable	Description
_HHRA	DV: Remoteness area
_HHDA[1]	DV: SEIFA 2001 Index of relative socio-economic disadvantage
_HHAD[1]	DV: SEIFA 2001 Index of relative socio-economic advantage/disadvantage
_HHEC[1]	DV: SEIFA 2001 Index of economic resources
_HHED[1]	DV: SEIFA 2001 Index of education and occupation
_HHDA10	DV: SEIFA 2001 Decile of Index of relative socio-economic disadvantage
_HHAD10	DV: SEIFA 2001 Decile of Index of relative socio-economic advantage/disadvantage
_HHEC10	DV: SEIFA 2001 Decile of Index of economic resources
_HHED10	DV: SEIFA 2001 Decile of Index of education and occupation
_HHMOVEK	DV: Distance person moved since last wave (kilometers), available from wave 2 onwards
_HHMOVEM	DV: Distance person moved since last wave (miles), available from wave 2 onwards
_HHMVEHK	DV: Distance household moved since last wave (kilometers), available from wave 2 onwards
_HHMVEHM	DV: Distance household moved since last wave (miles), available from wave 2 onwards
AHHCD96[1]	DV: ABS 1996 Census Collection District
_HHCD01[1]	DV: ABS 2001 Census Collection District

1. Variables are only on the In-confidence Release (unconfidentialised) files

Current Education Variables

The education questions have been used to derive variables (listed in Table 16) based on the Australian Standard Classification of Education (ASCED). There are a series of variables at the 3-digit ASCED level which contain information about:

- the number of qualifications completed (for new respondents only);
- which qualifications the respondent is currently studying for; and
- which qualifications have been obtained since the last interview (for continuing respondents only).

Where a qualification cannot be categorised to the detailed level (for example, 211 Graduate Diploma or 221 Graduate Certificate), the broader category has been used (for example, 200 Graduate Diploma and Graduate Certificate).

Unless you are specifically interested in what qualifications the respondent has completed since the last interview, you should use the history variables described earlier (which combines the answers provided in the current and previous wave interviews).

Note that a flag has been created to identify full time students (_*edfts*).

Table 16: Derived current education variables

Variable	Description
Qualifications Obtained (People Interviewed For the First Time)	
_EDQ100N	DV: ASCED 100 Postgraduate - Number of qualifications
_EDQ110N	DV: ASCED 110 Doctoral degree - Number of qualifications
_EDQ120N	DV: ASCED 120 Master degree - Number of qualifications
_EDQ200N	DV: ASCED 200 Grad Dip and Grad Cert - Number of qualifications
_EDQ211N	DV: ASCED 211 Graduate Diploma - Number of qualifications
_EDQ221N	DV: ASCED 221 Graduate certificate - Number of qualifications
_EDQ310N	DV: ASCED 310 Bachelor degree - Number of qualifications
_EDQ311N	DV: ASCED 311 Bachelor (Honours) Degree - Number of qualifications
_EDQ312N	DV: ASCED 312 Bachelor (Pass) Degree - Number of qualifications
_EDQ400N	DV: ASCED 400 Advanced Diploma and Diploma - Number of qualifications
_EDQ411N	DV: ASCED 411 Advanced Diploma - Number of qualifications
_EDQ413N	DV: ASCED 413 Associate Degree - Number of qualifications
_EDQ421N	DV: ASCED 421 Diploma - Number of qualifications
_EDQ500N	DV: ASCED 500 Certificate level - Number of qualifications
_EDQ511N	DV: ASCED 511 Certificate IV - Number of qualifications
_EDQ514N	DV: ASCED 514 Certificate III - Number of qualifications
_EDQ521N	DV: ASCED 521 Certificate II - Number of qualifications
_EDQ524N	DV: ASCED 524 Certificate I - Number of qualifications
_EDQ600N	DV: ASCED 600 Secondary education - Number of qualifications
_EDQ611N	DV: ASCED 611 Year 12 - Number of qualifications
_EDQUNKN	DV: ASCED Unknown - Not enough information - Number of qualifications
Qualifications Currently Studying For	
_EDCQ100	DV: ASCED 100 Postgraduate - Currently studying
_EDCQ110	DV: ASCED 110 Doctoral degree - Currently studying
_EDCQ120	DV: ASCED 120 Master degree - Currently studying
_EDCQ200	DV: ASCED 200 Grad Dip and Grad Cert - Currently studying
_EDCQ211	DV: ASCED 211 Graduate Diploma - Currently studying
_EDCQ221	DV: ASCED 221 Graduate certificate - Currently studying
_EDCQ310	DV: ASCED 310 Bachelor degree - Currently studying
_EDCQ311	DV: ASCED 311 Bachelor (Honours) Degree - Currently studying
_EDCQ312	DV: ASCED 312 Bachelor (Pass) Degree - Currently studying

Table 16: (c'td)

Variable	Description
_EDCQ400	DV: ASCED 400 Advanced Diploma and Diploma - Currently studying
_EDCQ411	DV: ASCED 411 Advanced Diploma - Currently studying
_EDCQ413	DV: ASCED 413 Associate Degree - Currently studying
_EDCQ421	DV: ASCED 421 Diploma - Currently studying
_EDCQ500	DV: ASCED 500 Certificate level - Currently studying
_EDCQ511	DV: ASCED 511 Certificate IV - Currently studying
_EDCQ514	DV: ASCED 514 Certificate III - Currently studying
_EDCQ521	DV: ASCED 521 Certificate II - Currently studying
_EDCQ524	DV: ASCED 524 Certificate I - Currently studying
_EDCQ600	DV: ASCED 600 Secondary education - Currently studying
_EDCQ611	DV: ASCED 611 Year 12 - Currently studying
_EDCQUNK	DV: ASCED Unknown - Not enough information - Currently studying
_EDFTS	DV: Full-time student
Qualifications Obtained Since Last Interview	
_EDRQ100	DV: ASCED 100 Postgraduate - Recent qualifications
_EDRQ110	DV: ASCED 110 Doctoral degree - Recent qualifications
_EDRQ120	DV: ASCED 120 Master degree - Recent qualifications
_EDRQ200	DV: ASCED 200 Grad Dip and Grad Cert - Recent qualifications
_EDRQ211	DV: ASCED 211 Graduate Diploma - Recent qualifications
_EDRQ221	DV: ASCED 221 Graduate certificate - Recent qualifications
_EDRQ310	DV: ASCED 310 Bachelor degree - Recent qualifications
_EDRQ311	DV: ASCED 311 Bachelor (Honours) Degree - Recent qualifications
_EDRQ312	DV: ASCED 312 Bachelor (Pass) Degree - Recent qualifications
_EDRQ400	DV: ASCED 400 Advanced Diploma and Diploma - Recent qualifications
_EDRQ411	DV: ASCED 411 Advanced Diploma - Recent qualifications
_EDRQ413	DV: ASCED 413 Associate Degree - Recent qualifications
_EDRQ421	DV: ASCED 421 Diploma - Recent qualifications
_EDRQ500	DV: ASCED 500 Certificate level - Recent qualifications
_EDRQ511	DV: ASCED 511 Certificate IV - Recent qualifications
_EDRQ514	DV: ASCED 514 Certificate III - Recent qualifications
_EDRQ521	DV: ASCED 521 Certificate II - Recent qualifications
_EDRQ524	DV: ASCED 524 Certificate I - Recent qualifications
_EDRQ611	DV: ASCED 611 Secondary School – Highest level - Recent qualifications
_EDRQ600	DV: ASCED 600 Secondary school - Lower level – Recent qualifications
_EDRQUNK	DV: ASCED Unknown - Not enough information - Recent qualifications

Current Marriage and Defacto Relationship Variables

The relationship section of the person questionnaires involve relatively complicated skips (especially from wave 2 onwards), so several partnering variables have been derived as set out in Table 17.

Table 17: Derived current marriage and defacto relationship variables

Variable	Description
_MRCURR	DV: Marital status from person questionnaire
_ORDFLT[1]	DV: NPQ: Years living together, first defacto excluding current
_ORDFRLT[2]	DV: NPQ: Years living together, most recent defacto excluding current

1. Waves 1 and from 4 onwards (NPQ)
2. Waves 2 and 3 only

Children Variables

Table 18 shows the various variables that have been created from the family formation section of the person questionnaires, including:

- the count of the number of the respondent's own resident and non-resident children (natural or adopted) of various ages, and the age of the respondent's own youngest child;
- the conversion into a common scale for the number of days or nights a child spends with their (other) parent; and
- the total child maintenance paid or received.

Table 18: Derived children variables

Variable	Description
All Children	
_TCYNG	DV: Age youngest own child (excl. foster/step). Weighted topcode.
Resident Children	
_TCR	DV: Number of own resident children
_TCR04	DV: Count of own resident children aged 0-4 yrs (G15)
_TCR514	DV: Count of own resident children aged 5-14 yrs (G15)
_TCR1524	DV: Count of own resident children aged 15-24 yrs (G15)
_TCR25	DV: Count of own resident children aged 25+ yrs (G15)
_RCYNG	DV: Age youngest resident own child (excl. foster/step).
_RCNGT	DV: Resident childs overnight stays with other parent (Days per annum)
_RCDAY	DV: Resident childs day visits with other parent (Days per annum)
_RCEFSPY[1]	DV: Resident child maintenance annual - paid - all children ($)
_RCEFSRY[1]	DV: Resident child maintenance received - annual - all children ($)
ARCEFSY[1]	DV: Child maintenance received - annual - all children ($)

Table 18: (c'td)

Variable	Description
Non-resident Children	
_TCNR	DV: Number of own non-resident children
_TCN04	DV: Count of own non-resident children aged 0-4 yrs (G3)
_TCN514	DV: Count of own non-resident children aged 5-14 yrs (G3)
_TCN1524	DV: Count of own non-resident children aged 15-24 yrs (G3)
_TCN25	DV: Count of own non-resident children aged 25+ yrs (G3)
_NCYNG	DV: Age youngest non-resident own child. Weighted topcode.
_NCNGT	DV: Overnight stays of non-resident child (Days per annum)
_NCDAY	DV: Day visits of non-resident child (Days per annum)
_NCEFSPY[1]	DV: Non-resident child maintenance annual - paid - all children ($)
_NCEFSRY[1]	DV: Non-resident child maintenance received - annual - all children ($)
ANCEFSY[1]	DV: Child maintenance paid - annual - all children ($)

1. In wave 1, the question only asked how much child maintenance they paid for non-resident children and how much they received for resident children. From wave 2 onwards, the questions were reworded to pay (_NCEFSPY, RCEFSPY) or receive (_NCEFSRY, RCEFSRY) for both non-resident and resident children.

Child Care Variables

The variables from the child care grids in the Household Questionnaire are used to make a number of summary variables (which are shown in Table 19). The children referred to in this section of the HQ are those living in the household aged under 15 and these are split into two groups:

- School-aged children – these children are of an age to attend school (that is, from aged 4 or 5, depending on the State, to 15).
- Children not yet at school – these children are aged 0 to 3 or 4, depending on the State, who are not yet of an age to attend school.[8]

The child care questions have changed a number of times across the waves in the following ways:

- The reason the child care was used – In wave 1, only information about child care used while the parents were working was collected. From wave 2, questions were included about the child care used so parent could undertake non-employment related activities (such as studying, excercising, shopping, etc).
- The level of detail collected for non-employment related child care – For waves 2 through 4, summary information was collected about the use of non-

[8] Up to wave 4, the questionnaire referred to the children not yet at school as 'pre-school' children. The shorter name was used in the questionnaire for space reasons but the interviewers were briefed on the intent of these questions to include all children who were not yet at school (not just those who aged 3 or 4 who are attending pre-school). The variable labels relating to the children not yet at school have been revised to use the 'not yet at school' terminology rather than the 'pre-school' terminology.

employment related child care. From wave 5, these grids contain a similar level of detail to the employment related child care grids.

- The level of detail collected for the cost of employment related child care – In Wave 1, the cost of each type of child care for each child was collected. From wave 2 onwards, the total cost for each type of child care for the two groups of children (school aged and those not yet at school) was collected.

- The level of detail for relatives looking after children – The types of child care that made reference to 'relatives' in waves 1 to 3 were split into 'grandparents' and 'other relatives' from wave 4.

The child care summary variables indicate whether a particular type of child care is used, along with the hours and cost (summed across the relevant children).

Table 19: Derived child care variables

Variable	Description
Child Care Used For School-Aged Children During Term Time While Parents Work	
_CSU_ME	DV: Childcare used - Me or my partner - Any school-age child - during term - While parents work
_CSU_BS	DV: Childcare used - The child's brother or sister - Any school-age child - during term - While parents work
_CSU_SF	DV: Childcare used - Child looks after self - Any school-age child - during term - While parents work
_CSU_WP	DV: Childcare used - Child comes to my workplace - Any school-age child - during term - While parents work
_CSU_OS	DV: Childcare used - Out of hours pw care at childs school - Any school-age child - during term - While parents work
_CSU_OE	DV: Childcare used - Out of hours pw care elsewhere - Any school-age child - during term time - While parents work
_CSU_RU[1]	DV: Childcare used - A relative who lives with us - Any school-age child - during term time - While parents work
_CSU_RE[1]	DV: Childcare used - A relative who lives elsewhere - Any school-age child - during term time - While parents work
_CSU_GU[2]	DV: Childcare used - Child's grandparent who lives with us - Any school-age child - during term time - While parents work
_CSU_GE[2]	DV: Childcare used - Child's grandparent who lives elsewhere - Any school-age child - during term time - While parents work
_CSU_AU[2]	DV: Childcare used - Other relative who lives with us - Any school-age child - during term time - While parents work
_CSU_AE[2]	DV: Childcare used - Other relative who lives elsewhere - Any school-age child - during term time - While parents work
_CSU_FO	DV: Childcare used - A friend or neighbour coming to our home - Any school-age child - during term time - While parents work
_CSU_FT	DV: Childcare used - A friend or neighbour in their home - Any school-age child - during term time - While parents work
_CSU_PS	DV: Childcare used - A paid sitter or nanny - Any school-age child - during term time - While parents work

Table 19: (c'td)

Variable	Description
_CSU_FD	DV: Childcare used - Family day care - Any school-age child - during term time - While parents work
_CSU_FC[2]	DV: Childcare used - Formal outside of school hours care - Any school-age child - during term time - While parents work
_CSU_OP	DV: Childcare used - Other parent not living in household/expartner - Any school-age child - during term time - While parents work
_CSU_BR	DV: Childcare used - Not applicable - Boarding school - Any school-age child - during term time - While parents work
_CSU_O1	DV: Childcare used - Other 1 (Specify) - Any school-age child - during term time - While parents work
_CSU_O2	DV: Childcare used - Other 2 (Specify) - Any school-age child - during term time - While parents work
_CSU_NA	DV: Childcare used - Not Answered - Any school-age child - during term time - While parents work
_CSH_BS	DV: Total hours - The childs brother or sister - All school-age children - during term time - While parents work
_CSH_SF	DV: Total hours - Child looks after self - All school-age children - during term time - While parents work
_CSH_WP	DV: Total hours - Child comes to my workplace - All school-age children - during term time - While parents work
_CSH_OS	DV: Total hours - Out of hours care at childs school - All school-age children - during term time - While parents work
_CSH_OE	DV: Total hours - Out of hours care elsewhere - All school-age children - during term time - While parents work
_CSH_RU[1]	DV: Total hours - A relative who lives with us - All school-age children - during term time - While parents work
_CSH_RE[1]	DV: Total hours - A relative who lives elsewhere - All school-age children - during term time - While parents work
_CSH_GU[2]	DV: Total hours pw - Childs grandparent who lives with us - All school-age children - during term time - While parents work
_CSH_GE[2]	DV: Total hours pw - Childs grandparent who lives elsewhere - All school-age children - during term time - While parents work
_CSH_AU[2]	DV: Total hours pw - Other relative who lives with us - All school-age children - during term time - While parents work
_CSH_AE[2]	DV: Total hours pw - Other relative who lives elsewhere - All school-age children - during term time - While parents work
_CSH_FO	DV: Total hours - A friend or neighbour coming to our home - All school-age children - during term time - While parents work
_CSH_FT	DV: Total hours - A friend or neighbour in their home - All school-age children - during term time - While parents work
_CSH_PS	DV: Total hours - A paid sitter or nanny - All school-age children - during term time - While parents work
_CSH_FD	DV: Total hours - Family day care - All school-age children - during term time - While parents work
_CSH_FC[2]	DV: Total hours pw - Formal outside of school hours care - All school-age children - during term time - While parents work

Table 19: (c'td)

Variable	Description
_CSH_O1	DV: Total hours - Other specify 1 - All school-age children - during term time - While parents work
_CSH_O2	DV: Total hours - Other specify 2 - All school-age children - during term time - While parents work
_CSCTC	DV: Childcare total cost ($) for all school age children during term, across all types of care. While parents work
Child Care Used For School-Aged Children During Holidays While Parents Work	
_CHU_ME	DV: Childcare used - Me or my partner - Any school-age child - during school holidays - While parents work
_CHU_BS	DV: Childcare used - The childs brother or sister - Any school-age child - during school holidays - While parents work
_CHU_SF	DV: Childcare used - Child looks after self - Any school-age child - during school holidays - While parents work
_CHU_WP	DV: Childcare used - Child comes to my workplace - Any school-age child - during school holidays - While parents work
_CHU_VS	DV: Childcare used - Vacation care at childs school - Any school-age child - during school holidays - While parents work
_CHU_VE	DV: Childcare used - Vacation care elsewhere - Any school-age child - during school holidays - While parents work
_CHU_RU[1]	DV: Childcare used - A relative who lives with us - Any school-age child - during school holidays - While parents work
_CHU_RE[1]	DV: Childcare used - A relative who lives elsewhere - Any school-age child - during school holidays - While parents work
_CHU_GU[2]	DV: Q9 Childcare used - Childs grandparent who lives with us - Any school-age child - during school holidays - While parents work
_CHU_GE[2]	DV: Childcare used - Childs grandparent who lives elsewhere - Any school-age child - during school holidays - While parents work
_CHU_AU[2]	DV: Childcare used - Other relative who lives with us - Any school-age child - during school holidays - While parents work
_CHU_AE[2]	DV: Childcare used - Other relative who lives elsewhere - Any school-age child - during school holidays - While parents work
_CHU_FO	DV: Childcare used - A friend or neighbour coming to our home - Any school-age child - during school holidays - While parents work
_CHU_FT	DV: Childcare used - A friend or neighbour in their home - Any school-age child - during school holidays - While parents work
_CHU_PS	DV: Childcare used - A paid sitter or nanny - Any school-age child - during school holidays - While parents work
_CHU_FD	DV: Childcare used - Family day care - Any school-age child - during school holidays - While parents work
_CHU_VC[2]	DV: Childcare used - Vacation care - Any school-age child - during school holidays - While parents work
_CHU_BR	DV: Childcare used - Not applicable - Boarding school - Any school-age child - during school holidays - While parents work
_CHU_O1	DV: Childcare used - Other 1 (Specify) - Any school-age child - during school holidays - While parents work

Table 19: (c'td)

Variable	Description
_CHU_O2	DV: Childcare used - Other 2 (Specify) - Any school-age child - during school holidays - While parents work
_CHU_NA	DV: Childcare used - Not Answered - Any school-age child - during school holidays - While parents work
_CHH_BS	DV: Total hours - The childs brother or sister - All school-age children - during school holidays - While parents work
_CHH_SF	DV: Total hours - Child looks after self - All school-age children - during school holidays - While parents work
_CHH_WP	DV: Total hours - Child comes to my workplace - All school-age children - during school holidays - While parents work
_CHH_VS	DV: Total hours - Vacation care at childs school - All school-age children - during school holidays - While parents work
_CHH_VE	DV: Total hours - Vacation care elsewhere - All school-age children - during school holidays - While parents work
_CHH_RU[1]	DV: Total hours - A relative who lives with us - All school-age children - during school holidays - While parents work
_CHH_RE[1]	DV: Total hours - A relative who lives elsewhere - All school-age children - during school holidays - While parents work
_CHH_GU[2]	DV: Total hours pw - Childs grandparent who lives with us - All school-age children - during school holidays - While parents work
_CHH_GE[2]	DV: Total hours pw - Childs grandparent who lives elsewhere - All school-age children - during school holidays - While parents work
_CHH_AU[2]	DV: Total hours pw - Other relative who lives with us - All school-age children - during school holidays - While parents work
_CHH_AE[2]	DV: Total hours pw - Other relative who lives elsewhere - All school-age children - during school holidays - While parents work
_CHH_FO	DV: Total hours - A friend or neighbour coming to our home - All school-age children - during school holidays - While parents work
_CHH_FT	DV: Total hours - A friend or neighbour in their home - All school-age children - during school holidays - While parents work
_CHH_PS	DV: Total hours - A paid sitter or nanny - All school-age children - during school holidays - While parents work
_CHH_FD	DV: Total hours - Family day care - All school-age children - during school holidays - While parents work
_CHH_VC[2]	DV: Total hours pw - Vacation care - All school-age children - during school holidays - While parents work
_CHH_O1	DV: Total hours - Other specify 1 - All school-age children - during school holidays - While parents work
_CHH_O2	DV: Total hours - Other specify 2 - All school-age children - during school holidays - While parents work
_CHCTC	DV: Childcare total cost ($) for all school age children during school holidays, across all types of care. While parents work
Child Care Used For Children Not Yet At School While Parents Work	
_CPU_ME	DV: Childcare used - Me or my partner - Any child not yet at school - While parents work
_CPU_BS	DV: Childcare used - The childs brother or sister - Any child not yet at school - While parents work

Table 19: (c'td)

Variable	Description
_CPU_RU[1]	DV: Childcare used - A relative who lives with us - Any child not yet at school - While parents work
_CPU_RE[1]	DV: Childcare used - A relative who lives elsewhere - Any child not yet at school - While parents work
_CPU_GU[2]	DV: Childcare used - Child's grandparent who lives with us - Any not yet at school child - While parents work
_CPU_GE[2]	DV: Childcare used - Child's grandparent who lives elsewhere - Any not yet at school child - While parents work
_CPU_AU[2]	DV: Childcare used - Other relative who lives with us - Any not yet at school child - While parents work
_CPU_AE[2]	DV: Childcare used - Other relative who lives elsewhere - Any not yet at school child - While parents work
_CPU_FO	DV: Childcare used - A friend or neighbour coming to our home - Any child not yet at school - While parents work
_CPU_FT	DV: Childcare used - A friend or neighbour in their home - Any child not yet at school - While parents work
_CPU_PS	DV: Childcare used - A paid sitter or nanny - Any child not yet at school - While parents work
_CPU_FD	DV: Childcare used - Family day care - Any child not yet at school - While parents work
_CPU_WD	DV: Childcare used - Long day care centre at workplace - Any child not yet at school - While parents work
_CPU_PD	DV: Childcare used - Private or community long day care centre - Any child not yet at school - While parents work
_CPU_KP	DV: Childcare used - Kindergarten/pre-school - Any child not yet at school - While parents work
_CPU_O1	DV: Childcare used - Other 1 (Specify) - Any child not yet at school - While parents work
_CPU_O2	DV: Childcare used - Other 2 (Specify) - Any child not yet at school - While parents work
_CPU_OP	DV: Childcare used - Other parent not living in household/expartner - Any child not yet at school - While parents work
_CPU_NA	DV: Childcare used - Not Answered - Any child not yet at school - While parents work
_CPH_BS	DV: Total hours pw - The childs brother or sister - All children not yet at school - While parents work
_CPH_RU[1]	DV: Total hours pw - A relative who lives with us - All children not yet at school - While parents work
_CPH_RE[1]	DV: Total hours pw - A relative who lives elsewhere - All children not yet at school - While parents work
_CPH_GU[2]	DV: Total hours pw - Child's grandparent who lives with us - All not yet at school children - While parents work
_CPH_GE[2]	DV: Total hours pw - Child's grandparent who lives elsewhere - All not yet at school children - While parents work
_CPH_AU[2]	DV: Total hours pw - Other relative who lives with us - All not yet at school children - While parents work
_CPH_AE[2]	DV: Total hours pw - Other relative who lives elsewhere - All not yet at school children - While parents work
_CPH_FO	DV: Total hours pw - A friend or neighbour coming to our home - All children not yet at school - While parents work

Table 19: (c'td)

Variable	Description
_CPH_FT	DV: Total hours pw - A friend or neighbour in their home - All children not yet at school - While parents work
_CPH_PS	DV: Total hours pw - A paid sitter or nanny - All children not yet at school - While parents work
_CPH_FD	DV: Total hours pw - Family day care - All children not yet at school - While parents work
_CPH_WD	DV: Total hours pw - Long day care centre at workplace - All children not yet at school - While parents work
_CPH_PD	DV: Total hours pw - Private or community long day care centre - All children not yet at school - While parents work
_CPH_KP	DV: Total hours pw - Kindergarten / preschool - All children not yet at school - While parents work
_CPH_O1	DV: Total hours pw - Other specify 1 - All children not yet at school - While parents work
_CPH_O2	DV: Total hours pw - Other specify 2 - All children not yet at school - While parents work
_CPCTC	DV: Childcare total cost ($) for all children not yet at school across all types of care - While parents work
Cost for Child Care Used While Parents Work (To Make Wave 1 Comparable)	
ACSC_WP	DV: Total cost pw - All school age children - Child comes to workplace - during term - while parents work
ACSC_OS	DV: Total cost pw - All school age children - Out of hours care at childs school - during term - while parents work
ACSC_OE	DV: Total cost pw - All school age children - Out of hours care elsewhere - during term - while parents work
ACSC_RU	DV: Total cost pw - All school age children - A relative who lives with us - during term - while parents work
ACSC_RE	DV: Total cost pw - All school age children - A relative who lives elsewhere - during term - while parents work
ACSC_FO	DV: Total cost pw - All school age children - A friend or neighbour coming to our home - during term - while parents work
ACSC_FT	DV: Total cost pw - All school age children - A friend or neighbour in their home - during term - while parents work
ACSC_PS	DV: Total cost pw - All school age children - A paid sitter or nanny - during term - while parents work
ACSC_FD	DV: Total cost pw - All school age children - Family day care - during term - while parents work
ACSC_O1	DV: Total cost pw - All school age children - Other specify 1 - during term - while parents work
ACSC_O2	DV: Total cost pw - All school age children - Other specify 2 - during term - while parents work
ACHC_WP	DV: Total cost pw - All school age children - Child comes to my workplace - during school holidays - while parents work
ACHC_VS	DV: Total cost pw - All school age children - Vacation care at childs school - during school holidays - while parents work
ACHC_VE	DV: Total cost pw - All school age children - Vacation care elsewhere - during school holidays - while parents work
ACHC_RU	DV: Total cost pw - All school age children - A relative who lives with us - during school holidays - while parents work
ACHC_RE	DV: Total cost pw - All school age children - A relative who lives elsewhere - during school holidays - while parents work

Table 19: (c'td)

Variable	Description
ACHC_FO	DV: Total cost pw - All school age children - A friend or neighbour coming to our home - during school holidays - while parents work
ACHC_FT	DV: Total cost pw - All school age children - A friend or neighbour in their home - during school holidays - while parents work
ACHC_PS	DV: Total cost pw - All school age children - A paid sitter or nanny - during school holidays - while parents work
ACHC_FD	DV: Total cost pw - All school age children - Family day care - during school holidays - while parents work
ACHC_O1	DV: Total cost pw - All school age children - Other specify 1 - during school holidays - while parents work
ACHC_O2	DV: Total cost pw - All school age children - Other specify 2 - during school holidays - while parents work
ACPC_RU	DV: Total cost pw - All children not yet at school - A relative who lives with us - while parents work
ACPC_RE	DV: Total cost pw - All children not yet at school - A relative who lives elsewhere - while parents work
ACPC_FO	DV: Total cost pw - All children not yet at school - A friend or neighbour coming to our home - while parents work
ACPC_FT	DV: Total cost pw - All children not yet at school - A friend or neighbour in their home - while parents work
ACPC_PS	DV: Total cost pw - All children not yet at school - A paid sitter or nanny - while parents work
ACPC_FD	DV: Total cost pw - All children not yet at school - Family day care - while parents work
ACPC_WD	DV: Total cost pw - All children not yet at school - Long day care centre at workplace - while parents work
ACPC_PD	DV: Total cost pw - All children not yet at school - Private or community long day care centre - while parents work
ACPC_KP	DV: Total cost pw - All children not yet at school - Kindergarten / preschool - while parents work
ACPC_O1	DV: Total cost pw - All children not yet at school - Other specify 1 - while parents work
ACPC_O2	DV: Total cost pw - All children not yet at school - Other specify 2 - while parents work
Child Care Used For School-Aged Children While Parents Do Not Work	
_CNSU_BS[3]	DV: Childcare used - The child's brother or sister - Any children school aged - not employment related
_CNSU_GU[3]	DV: Childcare used - Child's grandparent who lives with us - Any children school aged - not employment related
_CNSU_GE[3]	DV: Childcare used - Child's grandparent who elsewhere - Any children school aged - not employment related
_CNSU_AU[3]	DV: Childcare used - Other relative who lives with us - Any children school aged - not employment related
_CNSU_AE[3]	DV: Childcare used - Other relative who lives elsewhere - Any children school aged - not employment related
_CNSU_FO[3]	DV: Childcare used - A friend or neighbour coming to our home - Any children school aged - not employment related

Table 19: (c'td)

Variable	Description
_CNSU_FT[3]	DV: Childcare used - A friend or neighbour coming in their home - Any children school aged - not employment related
_CNSU_PS[3]	DV: Childcare used - A paid sitter/nanny - Any children school aged - not employment related
_CNSU_FD[3]	DV: Childcare used - Family day care - Any children school aged - not employment related
_CNSU_PD[3]	DV: Childcare used - Private/community long day care centre - Any children school aged - not employment related
_CNSU_FC[3]	DV: Childcare used - Formal outside of school hours care - Any children school aged - not employment related
_CNSU_01[3]	DV: Childcare used - Other specify 1 - All children school aged - not employment related
_CNSU_02[3]	DV: Childcare used - Other specify 2 - All children school aged - not employment related
_CNSU_NA[3]	DV: Childcare used - No answered - All children school aged - not employment related
_CNSU_NP[3]	DV: Childcare used - None - All children school aged - not employment related
_CNSH_BS[3]	DV: Total hours pw - The childs brother or sister - All children school aged - not employment related
_CNSH_GU[3]	DV: Total hours pw - Child's grandparent who lives with us - All children school aged - not employment related
_CNSH_GE[3]	DV: Total hours pw - Child's grandparent who elsewhere - All children school aged - not employment related
_CNSH_AU[3]	DV: Total hours pw - Other relative who lives with us - All children school aged - not employment related
_CNSH_AE[3]	DV: Total hours pw - Other relative who lives elsewhere - All children school aged - not employment related
_CNSH_FO[3]	DV: Total hours pw - A friend or neighbour coming to our home - All children school aged - not employment related
_CNSH_FT[3]	DV: Total hours pw - A friend or neighbour in their home - All children school aged - not employment related
_CNSH_PS[3]	DV: Total hours pw - A paid sitter/nanny - All children school aged - not employment related
_CNSH_FD[3]	DV: Total hours pw - Family day care - All children school aged - not employment related
_CNSH_PD[3]	DV: Total hours pw - Private/community long day care centre - All children school aged - not employment related
_CNSH_FC[3]	DV: Total hours pw - Formal outside of school hours care - All children school aged - not employment related
_CNSH_01[3]	DV: Total hours pw - Other 1 - All children school aged - not employment related
_CNSH_02[3]	DV: Total hours pw - Other 2 - All children school aged - not employment related
_NSCTC[3]	DV: Childcare total cost ($) for all school age children across all types of care. Not employment related
Child Care Used For Children Not Yet At School While Parents Do Not Work	
_CNPU_BS[3]	DV: Childcare used - The child's brother or sister - Any children not yet at school - not employment related
_CNPU_GU[3]	DV: Childcare used - Child's grandparent who lives with us - Any children not yet at school - not employment related
_CNPU_GE[3]	DV: Childcare used - Child's grandparent who lives elsewhere - Any children not yet at school - not employment related

Table 19: (c'td)

Variable	Description
_CNPU_AU[3]	DV: Childcare used - Other relative who lives with us - Any children not yet at school - not employment related
_CNPU_AE[3]	DV: Childcare used - Other relative who lives elsewhere - Any children not yet at school - not employment related
_CNPU_FO[3]	DV: Childcare used - A friend or neighbour coming to our home - Any children not yet at school - not employment related
_CNPU_FT[3]	DV: Childcare used - A friend or neighbour in their home - Any children not yet at school - not employment related
_CNPU_PS[3]	DV: Childcare used - A paid sitter/nanny - Any children not yet at school - not employment related
_CNPU_FD[3]	DV: Childcare used - Family day care - Any children not yet at school - not employment related
_CNPU_PD[3]	DV: Childcare used - Private/community long day care centre - Any children not yet at school - not employment related
_CNPU_KP[3]	DV: Childcare used - Kindergarten/pre-school - Any children not yet at school - not employment related
_CNPU_01[3]	DV: Childcare used - Other 1 (Specify) - Any children not yet at school - not employment related
_CNPU_02[3]	DV: Childcare used - Other 2 (Specify) - Any children not yet at school - not employment related
_CNPU_NA[3]	DV: Childcare used - No answer - Any children not yet at school - not employment related
_CNPU_NP[3]	DV: Childcare used - None - Any children not yet at school - not employment related
_CNPH_BS[3]	DV: Total hours pw - The child's brother or sister - All children not yet at school - not employment related
_CNPH_GU[3]	DV: Total hours pw - Child's grandparent who lives with us - All children not yet at school - not employment related
_CNPH_GE[3]	DV: Total hours pw - Child's grandparent who lives elsewhere - All children not yet at school - not employment related
_CNPH_AU[3]	DV: Total hours pw - Other relative who lives with us - All children not yet at school - not employment related
_CNPH_AE[3]	DV: Total hours pw - Other releative who lives elsewhere - All children not yet at school - not employment related
_CNPH_FO[3]	DV: Total hours pw - A friend or neighbour coming to our home - All children not yet at school - not employment related
_CNPH_FT[3]	DV: Total hours pw - A friend or neighbour in their home - All children not yet at school - not employment related
_CNPH_PS[3]	DV: Total hours pw - A paid sitter/nanny - All children not yet at school - not employment related
_CNPH_FD[3]	DV: Total hours pw - Family day care - All children not yet at school - not employment related
_CNPH_PD[3]	DV: Total hours pw - Private/community long day care centre - All children not yet at school - not employment related
_CNPH_KP[3]	DV: Total hours pw - Kindergarten/pre-school - All children not yet at school - not employment related
_CNPH_01[3]	DV: Total hours pw - Other specify 1 - All children not yet at school - not employment related
_CNPH_02[3]	DV: Total hours pw - Other specify 2 - All children not yet at school - not employment related

Table 19: (c'td)

Variable	Description
_NPCTC[3]	DV: Childcare total cost ($) for all children not yet at school across all types of care - not employment related

1. For waves 1 to 3 ('Relatives' were replaced by 'grandparents' and 'other relatives' from wave 4).

2. From wave 4.

3. From wave 5.

Employment Variables

The employment related derived variables are listed in Table 20. The occupation variables are coded to 4-digit Australian Standard Classification of Occupations (ASCO)[9] which is then used to produce:

- the 1-digit and 2-digit ASCO codes;
- ANU4 occupational status scale (which ranges from 0 to 100); and
- the 2-digit and 4-digit International Standard Classification of Occupation-88 (ISCO-88) codes (only the 4-digit ISCO is on the In-Confidence Release (unconfidentialised)).

The industry variables are coded to 4-digit Australian and New Zealand Standard Classification of Industry (ANZSIC)[10] which is then used to produce:

- the division level and 2-digit ANZSIC codes; and
- the 2-digit International Standard Industry Classification (ISIC) codes.

The 4-digit ASCO, ISCO and ANZSIC codes are available on the In-confidence Release (unconfidentialised) files only.

The history variables should first be consulted if you are attempting to piece together information about previous employment spells as some of the work may already be done.

In all waves but wave 2, the labour force status of individuals was asked on the Household Form, which provides useful information in the weighting and imputation processes for non-respondents. We have imputed the broad labour force status for all those people enumerated in wave 2 in three different ways. Firstly the HF labour force status was set to equal the PQ labour force status for individuals who completed an interview in wave 2. Any non-respondents in wave 2 who responded in wave 3 then had labour force status established from their response to the wave 3 calendar questions, which overlapped with the wave 2 interview date. The remaining non-respondents were imputed by a hotdeck imputation method which seeks to find a donor with a similar set of characteristics to the non-respondent (including the previous wave labour force status where available) and uses the donor's labour force status to impute for the non-respondents.

[9] ASCO, Second Edition 1997. From Release 7, we will use the Australian and New Zealand Standard Classification of Occupations (ANZSCO), First Edition 2006. We plan to backcode the occupation responses for waves 1 to 6 to this new codeframe.

[10] ANZSIC 1993 Edition. From Release 7, we will use the ANZSIC 2006 Edition. We plan to backcode the industry responses for waves 1 to 6 to this new codeframe.

Table 20: Derived employment variables

Variable	Description
_ESDTL	DV: Labour force status – detail
_ESBRD	DV: Labour force status – broad
_JBHRUC	DV: Hours per week usually worked in all jobs
_JBMHRUC	DV: Hours per week usually worked in main job
_JBHRQF	DV: Data Quality Flag: hours of work main job vs all jobs
_JBTPRHR	DV: Hours would like to work
_JBMOCC1	DV: Occupation 1-digit ASCO
_JBMOCC2	DV: Occupation 2-digit ASCO
_JBMOCCS	DV: ANU4 occupational status scale, current main job
_JBMI82	DV: ISCO-88 2-digit, Occupation current main job
_JBMI88[1]	DV: ISCO-88 4-digit, Occupation current main job
_JBMIND1	DV: C14 Current main job industry. ANZSIC division
_JBMIND2	DV: C14 Current main job industry. 2-digit ANZSIC
_JBMII2	DV: ISIC 3.1 2-digit Industry current main job
_WCPD[2]	DV: Days of paid workers compensation in last 12 months
_WCAPD[3]	DV: Days absent from work on paid workers compensation in last 12 months
_ALPD[2]	DV: Days of paid annual leave in last 12 months
_ALSK[2]	DV: days of paid sick leave in last 12 months
_ALOP[2]	DV: days of paid (maternity, paternity, bereavement, family, carers) leave in last 12 months
_ALUP[2]	DV: Days of unpaid leave in last 12 months
_ES	DV: Employment status in main job if currently employed
_JBMUABS	DV: ABS defined trade union membership (dont know=no)
_JBCASAB	DV: Casual worker (ABS definition: no paid holiday leave, no paid sick leave)
_JBOCCT	DV: Tenure in current occupation (years)
_JBEMPT	DV: Tenure with current employer (years)
_JST	DV: Weeks unemployed, missing if no exact duration
_UJLJOC1	DV: NPQ:D30 Occupation last job (not currently in paid work) 1-digit ASCO
_UJLJOC2	DV: NPQ:D30 Occupation last job (not currently in paid work) 2-digit ASCO
_UJLJOCS	DV: ANU4 occupational status scale, last job, not currently in paid work
_UJLJI82	DV: ISCO-88 2-digit, Occupation last job, not currently in paid work
_UJLJI88[1]	DV: ISCO-88 4-digit, Occupation last job, not currently in paid work
_UJLJIN1	DV: NPQ:D25 Industry, last job (not currently in paid work) ANZSIC division
_UJLJIN2	DV: NPQ:D25 Industry, last job (not currently in paid work) 2-digit ANZSIC
_UJLJII2	DV: ISIC 3.1 2-digit Industry previous job, if not currently in paid work
_UJLHRUC	DV: Hours per week worked in last job
_UJLJWS	DV: Pay in last job per annum ($)
_UJLJT	DV: Tenure with last employer (years)
_MOLT	DV: Months since did activity required by Centrelink/NP

Table 20: (c'td)

Variable	Description
_PJOOCC1	DV: CPQ: Occupation in previous job (currently employed) 1-digit ASCO
_PJOOCC2	DV: CPQ: Occupation in previous job (currently employed) 2-digit ASCO
_PJOOCCS	DV: ANU4 occupational status scale, CPQ: in previous job if job changed
_PJOI82	DV: ISCO-88 2-digit, CPQ: Occupation previous job if job changed
_PJOI88[1]	DV: ISCO-88 4-digit, CPQ: Occupation previous job if job changed
_PJOTOC1	DV: CPQ: Occupation in previous job (if currently unemployed) 1-digit ASCO
_PJOTOC2	DV: CPQ: Occupation in previous job (if currently unemployed) 2-digit ASCO
_PJOTOCS	DV: ANU4 occupational status scale, in previous job if unemployed
_PJOTI82	DV: ISCO-88 2-digit, CPQ: Occupation previous job if unemployed
_PJOTI88	DV: ISCO-88 4-digit, CPQ: Occupation previous job if unemployed
_PJOTIN1	DV: CPQ:D28 Previous job industry (if currently unemployed) ANZSIC division
_PJOTIN2	DV: CPQ:D28 Previous job industry (if currently unemployed) 2-digit ANZSIC
_PJOTII2	DV: ISIC 3.1 2-digit Industry previous job, if unemployed
_PJOIND1	DV: CPQ: Previous job industry (if currently employed). ANZSIC division
_PJOIND2	DV: CPQ: Previous job industry (if currently employed). 2-digit ANZSIC
_PJOII2	DV: ISIC 3.1 2-digit Industry previous job, if job changed
AJBPERM	DV: Permanently unable to work
BHGEBI	DV: Eligibility for interview - HF Labour force status - broad [imputed]
BHGEBF	DV: Imputation flag labour force status - broad
BHGEBI1 to BHGEBI14	DV: Eligibility for interview - HF Labour force status - broad [imputed]
BHGEBF1 to BHGEBF14	DV: Imputation flag labour force status - broad

1. Variables are only on the In-Confidence Release (unconfidentialised) files.

2. Wave 5 onwards.

3. Wave 6 onwards.

Calculating Hourly Wage Rates

The following is aimed at pointing you in the right direction if you want to calculate hourly wage rates. You would use the following derived variables:

- *_esbrd* DV: Labour force status – broad
- *_jbhruc* DV: Combined hrs per week usually worked in all jobs
- *_wscei* DV: Imputed current weekly gross wages & salary - all jobs ($) topcoded

The hourly wage rate can be calculated in SPSS as follows:

if (*aesbrd*=1 and *ajbhruc*>0 and *awscei*>0) hwr01 = rnd(*awscei/ajbhruc*).
if (*besbrd*=1 and *bjbhruc*>0 and *bwscei*>0) hwr02 = rnd(*bwscei/bjbhruc*).
…
if (f*esbrd*=1 and f*jbhruc*>0 and f*wscei*>0) hwr06 = rnd(f*wscei/fjbhruc*).

The above code calculates the hourly wage rate (across all their jobs) if the respondent:

(i) is employed;

(ii) has current wages and salaries; and

(iii) has usual hours worked in all jobs.

If you wish to look at those that are full and part time employed separately, use *_esdtl* (DV: Labour force status – detail) to define these groups. The cases that did not need to be imputed can be identified using the flag *_wscef* =0.

If you wish to look at the hourly wage in the respondent's main job, use *_wscmei* and *_jbmhruc*.

Please note that the questions about hours worked and income are asked in separate sections of the person questionnaire. As some respondents report low wages and salaries with high hours and vice versa, it is important that users are aware that there are some odd outliers when deriving hourly wage rates. This is, unfortunately, unavoidable.

Calendar Variables

The calendar contains over 1000 variables. Before you trawl through these variables and create your own summary variables, check if one of the derived calendar variables in Table 21 may help you. These derived variables typically relate to the financial year, while the calendar may stretch from 14 to 18 months, depending on the interview date.

Table 21: Derived calendar variables

Variable	Description
_CAPEFT	DV: Per cent time spent in ft education in last financial year
_CAPEPT	DV: Per cent time spent in pt education in last financial year
_CAPJ	DV: Per cent time spent in jobs in last financial year
_CAPUNE	DV: Per cent time spent unemployed in last financial year
_CAPNLF	DV: Per cent time spent not in the labour force in last financial year
_CAFNJ	DV: Number of jobs in last financial year
_CANTP	DV: Number of time periods answered in calendar

Family Variables

The relationship grid on the Household Form collects the relationship of everyone in the household to everyone else. This information is then used to assign people to family groups, identify what relationship they hold within the family, identify what type of family they belong to, and then identify what type of household they belong to.

The core relationships needed to make a family are a couple relationship or a parent-child relationship. There are several key points to note about how the families are defined:

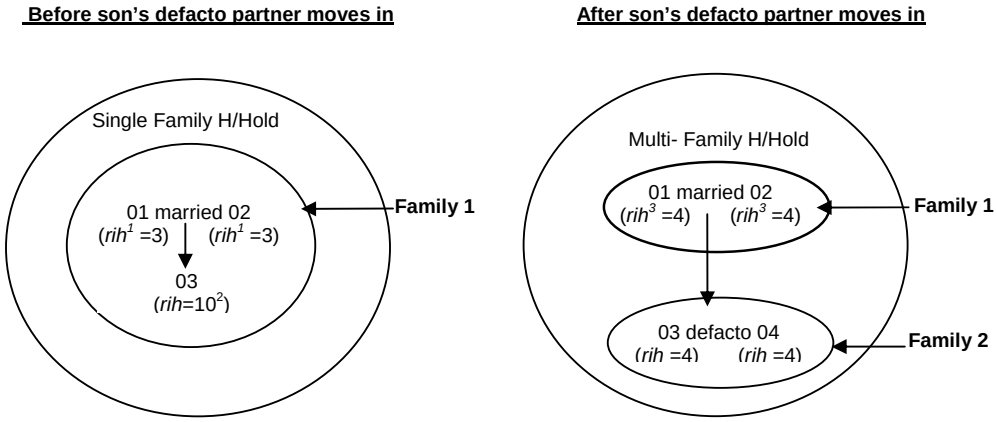
- A couple relationship takes precedence over a parent-child relationship (see Figure 15). In a household with mother, father, son and son's defacto, the son's couple relationship takes precedence over his child-parent relationship. This

household would be a multi-family household, with mother and father as a couple in one family and the son and his defacto in another family.

- The most recent generation has precedence over an older generation and the older generation is then considered another relative. Figure 16 illustrates this case. The core household is defined by the mother–daughter generation (Before Child). However, when the daughter has a daughter herself, that younger generation then takes precedence and forms the core household (After Child) and the first mother is considered to be a relative (a grandmother).

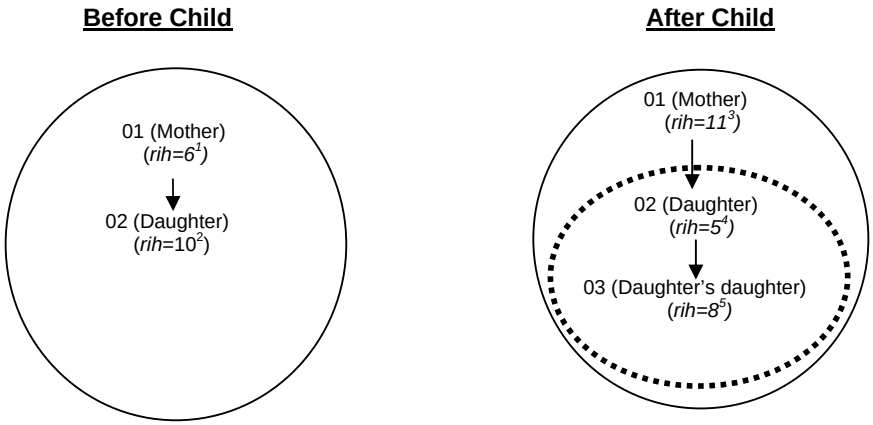
- Children aged under 15 living in a household without natural, adopted, step or foster parents are attached to their closest relative. If they are without relatives, then they are attached to the person thought most likely to form a parent-child relationship with that child.

Figure 15: Family where new defacto relationship is formed



1. *rih* = 3: Couple with non-dependent child
2. *rih* = 10: Non-dependent child
3. *rih* = 4: Couple without child

Figure 16: Family where new child is born

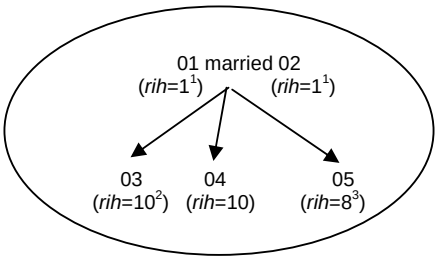


1. *rih* = 6: lone parent with non-dependent child
2. *rih* = 10: Non-dependent child
3. *rih* = 11: Other family member
4. *rih* = 5: Lone parent with dependent child
5. *rih* = 8: Dependent child < 15

For the family variables, children are defined as natural, adopted, step or foster children. Children are split into three classifications: children under 15, dependent students (who are aged 15-24, not employed full-time, are studying full-time, are living with at least one parent, and are without a partner or child of their own); and non-dependent children.[11]

The family and household labels used identify the most dependent type of children in the family without identifying the other types of children in the family. Figure 17 illustrates a family with a couple, one child under 15 and two non-dependent children. The family type would be classified as a "Couple with a child <15 without others" and the household type would be classified as a "Couple with a child <15 without others". "Others Related" or "Others unrelated" refer to those outside of the core family unit (such as aunts, uncles, or grandparents).

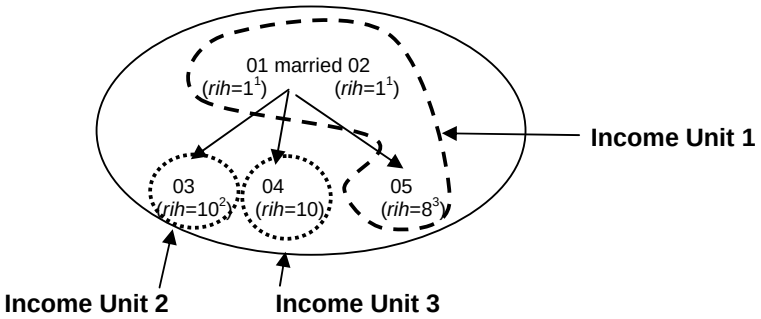
Figure 17: Family with child under 15 and non-dependent children



1. rih = 1: Couple with children < 15)
2. *rih* = 10: Non-dependent child
3. *rih* = 8: Dependent child < 15

The income units are derived from the family units and separate out the non-dependant children and other relatives from rest of the family. The previous scenario in Figure 17 is, therefore, broken into 3 income units. Figure 18 shows how this is done. The first income unit (1) includes mother, father and the child under 15. Each non-dependent child forms their own income unit (Income units 2 and 3).

Figure 18: Income units in a family with child under 15 and non-dependent children



1. rih = 1: Couple with children < 15)
2. *rih* = 10: Non-dependent child
3. *rih* = 8: Dependent child < 15

Along with the variables based on the relationship grid, a number of other variables are listed in Table 22, including identifiers for various people in the household and counts of

[11] Note that this definition of a dependent student is different to the full-time student identifier provided on the Responding Person File.

the number of people in certain age groups. The partner, father and mother identifiers were discussed in a preceding section on identifiers.

Table 22: Derived family variables

Variable	Description
_HHTYPE	DV: Household type
_HHRIH	DV: Relationship in household
_HHFAM[1]	DV: Family number
_HHFTY[1]	DV: Family type
_HHIU[1]	DV: Income unit
_HHPXID	DV: Partner's crosswave person number (7-digit)
_HHFXID	DV: Father's crosswave person number (7-digit)
_HHMXID	DV: Mother's crosswave person number (7-digit)
_HHPRTID	DV: Partner's 2-digit person number within household
_HHFID	DV: Father's 2-digit person number within household
_HHMID	DV: Mother's 2-digit person number within household
_HHYNG	DV: Age of youngest person in household. Weighted topcode.
_HHOLD	DV: Age of oldest person in household. Weighted topcode.
_HH0_4	DV: Number of persons aged 0-4 years at June 30
_HH5_9	DV: Number of persons aged 5-9 years at June 30
_HH10_14	DV: Number of persons aged 10-14 years at June 30
_HHADULT	DV: Number of persons aged 15+ years at June 30
_HHIVWS	DV: Number of completed interviews in household
_HGINT	DV: Interviewed flag
_HHPQ	DV: Type of person interview
_HGWSLI	DV: Weeks since last interviewed
_HGENUM	DV: Enumerated flag

1. On the Household File, these variables are listed for each person, that is _hhrih01 to _hhrih16, _hhfam01 to _hhfam16, _hhfty01 to _hhfty16, and _hhiu01 to _hhiu16. (Note that variables for persons 13 and 14 are only included from wave 2 and person 15 and 16 are only included from wave 6.)

Health Variables

Each wave the SF-36 Health Survey instrument is included within the Self-Completion Questionnaire. The SF-36 Health Survey is an internationally recognised diagnostic tool for assessing functional health status and well-being. It comprises 36 items which provide multi-item scales measuring each of eight distinct health concepts. Following the scoring rules outlined in Ware et al. (2000), each of these eight scales has been transformed into a standardised 0-100 index. The individual scores for each of these indices have been provided as derived variables in the data set. In addition, the SF-6D health state classification has also been derived from the SF-36 (as outlined in Brazier, Roberts and Deverill, 2002).

From wave 6, respondents are asked to record their height and weight in the Self-Completion Questionnaire. This is used to derive their body mass index. Further

information on the quality of the height and weight data is provided in Wooden and Watson (2008).

The health variables are listed in Table 23.

Table 23: Derived health variables

Variable	Description
_GHPF	DV: SF-36 physical functioning – transformed
_GHRP	DV: SF-36 role-physical – transformed
_GHBP	DV: SF-36 bodily pain – transformed
_GHGH	DV: SF-36 general health – transformed
_GHVT	DV: SF-36 vitality – transformed
_GHSF	DV: SF-36 social functioning – transformed
_GHRE	DV: SF-36 role-emotional – transformed
_GHMH	DV: SF-36 mental health – transformed
_GHRHT	DV: SF-36 reported health transitions - raw
_GHSF6D	DV: SF-6D Health state Classification
_BMHT[1]	DV: SCQ: Body Mass Index. Height in centimetres
_BMWT[1]	DV: SCQ: Body Mass Index. Weight in kilograms
_BMI[1]	DV: SCQ: Body Mass Index. BMI
_BMIGP[1]	DV: SCQ: Body Mass Index. BMI group

1. From wave 6.

Housing Variables

Table 24 lists a couple of derived housing variables provided on the datasets. The associated questions in the Household Questionnaire allow numerous ways in which the respondent could have answered (for example, per month, per fortnight, per week, etc).

Also, as of Release 6, home value has been imputed using the method described later for the imputation of household wealth.

Table 24: Derived housing variables

Variable	Description
_HSRNT	DV: Rent usual repayments $ per month
_HSMG	DV: Mortgage usual repayments $ per month
_HSSL	DV: Second mortgage usual repayments $ per month
_HSVALUI	DV: Home value ($) [imputed]
_HSVALUF	DV: Imputation flag home value

Time Use Variables

Table 25 lists derived time use variables which combine the hours and minutes spent in a week on various activities.

Table 25: Derived time use variables

Variable	Description
_LSEMP	DV: Combined hrs/mins per week - Paid employment
_LSCOM	DV: Combined hrs/mins per week - Travelling to/from paid employment
_LSERR	DV: Combined hrs/mins per week - Household errands
_LSHW	DV: Combined hrs/mins per week - Housework
_LSOD	DV: Combined hrs/mins per week - Outdoor tasks
_LSCHD	DV: Combined hrs/mins per week - Playing with your children
_LSOCD	DV: Combined hrs/mins per week - Playing with other peoples children
_LSVOL	DV: Combined hrs/mins per week - Volunteer/Charity work
_LSCAR	DV: Combined hrs/mins per week - Caring for disabled/elderly relative

Household Expenditure Variables

Table 26 lists the derived household-level expenditure variables, as collected in the Self-Completion Questionnaire from wave 5.

While the person in the household responsible for the household bills was asked to complete these questions, sometimes more than one person in a household provided answers. The variables with prefix *_xp* are the derived annualised response for each person who provided a response to these questions. The variables with the prefix *_hx* averages the responses across all individuals who provided a response to these expenditure questions. Most users will use the *_hx* variables.

Table 26: Derived household expenditure variables

Variable	Description
Household Expenditure (Averaged Across Individuals Providing Response)	
_HXYGROC	DV: SCQ Household annual expenditure - Groceries ($)
_HXYALC	DV: SCQ Household annual expenditure - Alcohol ($)
_HXYCIG	DV: SCQ Household annual expenditure - Cigarettes and tobacco ($)
_HXYPUBT	DV: SCQ Household annual expenditure - Public transport and taxis ($)
_HXYMEAL	DV: SCQ Household annual expenditure - Meals eaten out ($)
_HXYHSGE[1]	DV: SCQ Household annual expenditure - Hobbies, sports, gambling, entertainment ($)
_HXYMVF	DV: SCQ Household annual expenditure - Motor vehicle fuel ($)
_HXYCLTH[1]	DV: SCQ Household annual expenditure - Clothing and footwear ($)
_HXYMCF[2]	DV: SCQ Household annual expenditure - Mens clothing and footwear ($)
_HXYWG[2]	DV: SCQ Household annual expenditure - Womens clothing and footwear ($)
_HXYCCF[2]	DV: SCQ Household annual expenditure - Childrens clothing and footwear ($)

Table 26: (c'td)

Variable	Description
_HXYTEL[1]	DV: SCQ Household annual expenditure - Telephone rent and calls (excl. internet charges) ($)
_HXYTELI[2]	DV: SCQ Household annual expenditure - Telephone rent, calls and internet charges ($)
_HXYHOL	DV: SCQ Household annual expenditure - Holidays and holiday travel ($)
_HXYPHI	DV: SCQ Household annual expenditure - Private health insurance ($)
_HXYOI[2]	DV: SCQ Household annual expenditure - Other insurance (home/contents/motor vehicle) ($)
_HXYHLTH[1]	DV: SCQ Household annual expenditure - Health care ($)
_HXYHLTP[2]	DV: SCQ Household annual expenditure - Fees paid to health practitioners ($)
_HXYPHRM[2]	DV: SCQ Household annual expenditure - Medicines, prescriptions, pharmaceuticals, alternative medicines ($)
_HXYELEC[1]	DV: SCQ Household annual expenditure - Electricity bills ($)
_HXYGAS[1]	DV: SCQ Household annual expenditure - Gas bills ($)
_HXYOHF[1]	DV: SCQ Household annual expenditure - Other heating fuels ($)
_HXYUTIL	DV: SCQ Household annual expenditure - Electricity, gas and other heating fuel bills ($)
_HXYHMRN	DV: SCQ Household annual expenditure - Home repairs/renovations/maintenance ($)
_HXYMVR	DV: SCQ Household annual expenditure - Motor vehicle repairs/maintenance ($)
_HXYEDUC	DV: SCQ Household annual expenditure - Education fees ($)
_HXYNCAR[2]	DV: SCQ Household annual expenditure - New motor vehicles, motorbikes or other vehicles ($)
_HXYUCAR[2]	DV: SCQ Household annual expenditure - Used motor vehicles, motorbikes or other vehicles ($)
_HXYCOMP[2]	DV: SCQ Household annual expenditure - Computers and related devices ($)
_HXYTVAV[2]	DV: SCQ Household annual expenditure - Televisions, home entertainment systems and other AV equipment ($)
_HXYWG[2]	DV: SCQ Household annual expenditure - Whitegoods such as ovens, fridges etc ($)
_HXYFURN[2]	DV: SCQ Household annual expenditure - Furniture ($)
Annualised Household Expenditure (Individuals Providing a Response)	
_XPYGROC	DV: SCQ Annual household expenditure - person level - Groceries ($)
_XPYALC	DV: SCQ Annual household expenditure - person level - Alcohol ($)
_XPYCIG	DV: SCQ Annual household expenditure - person level - Cigarettes and tobacco ($)
_XPYPUBT	DV: SCQ Annual household expenditure - person level - Public transport and taxis ($)
_XPYMEAL	DV: SCQ Annual household expenditure - person level - Meals eaten out ($)
_XPYHSGE[1]	DV: SCQ Annual household expenditure - person level - Hobbies, sports, gambling, entertainment ($)
_XPYMVF	DV: SCQ Annual household expenditure - person level - Motor vehicle fuel ($)
_XPYCLTH[1]	DV: SCQ Annual household expenditure - person level - Clothing and footwear ($)
_XPYMCF[2]	DV: SCQ Annual household expenditure - person level - Mens clothing and footwear ($)
_XPYWG[2]	DV: SCQ Annual household expenditure - person level - Womens clothing and footwear ($)

Table 26: (c'td)

Variable	Description
_XPYCCF[2]	DV: SCQ Annual household expenditure - person level - Childrens clothing and footwear ($)
_XPYTEL[1]	DV: SCQ Annual household expenditure - person level - Telephone rent and calls (excl. internet charges) ($)
_XPYTELI[2]	DV: SCQ Annual household expenditure - person level - Telephone rent, calls and internet charges ($)
_XPYHOL	DV: SCQ Annual household expenditure - person level - Holidays and holiday travel ($)
_XPYPHI	DV: SCQ Annual household expenditure - person level - Private health insurance ($)
_XPYOI[2]	DV: SCQ Annual household expenditure - person level - Other insurance (home/contents/motor vehicle) ($)
_XPYHLTH[1]	DV: SCQ Annual household expenditure - person level - Health care ($)
_XPYHLTP[2]	DV: SCQ Annual household expenditure - person level - Fees paid to health practitioners ($)
_XPYPHRM[2]	DV: SCQ Annual household expenditure - person level - Medicines, prescriptions, pharmaceuticals, alternative medicines ($)
_XPYELEC[1]	DV: SCQ Annual household expenditure - person level - Electricity bills ($)
_XPYGAS[1]	DV: SCQ Annual household expenditure - person level - Gas bills ($)
_XPYOHF[1]	DV: SCQ Annual household expenditure - person level - Other heating fuels ($)
_XPYUTIL	DV: SCQ Annual household expenditure - person level - Electricity, gas and other heating fuel bills ($)
_XPYHMRN	DV: SCQ Annual household expenditure - person level - Home repairs/renovations/maintenance ($)
_XPYMVR	DV: SCQ Annual household expenditure - person level - Motor vehicle repairs/maintenance ($)
_XPYEDUC	DV: SCQ Annual household expenditure - person level - Education fees ($)
_XPYNCAR[2]	DV: SCQ Annual household expenditure - person level - New motor vehicles, motorbikes or other vehicles ($)
_XPYUCAR[2]	DV: SCQ Annual household expenditure - person level - Used motor vehicles, motorbikes or other vehicles ($)
_XPYCOMP[2]	DV: SCQ Annual household expenditure - person level - Computers and related devices ($)
_XPYTVAV[2]	DV: SCQ Annual household expenditure - person level - Televisions, home entertainment systems and other AV equipment ($)
_XPYWG[2]	DV: SCQ Annual household expenditure - person level - Whitegoods such as ovens, fridges etc ($)
_XPYFURN[2]	DV: SCQ Annual household expenditure - person level - Furniture ($)

1. Only in wave 5.

2. From wave 6.

Personality Variables

In wave 5, respondents were asked questions about their personality character traits based on the Big 5 Personality Inventory (taken from Saucier, 1994). Five scales have been created and are listed in Table 27 below. Reliability and factor analyses were undertaken, resulting in some variation from Saucier's scales. The five scales are composed by taking the average of the following items:

- Extroversion – talkative, bashful (reversed), quiet (reversed), shy (reversed), lively, and extroverted.
- Agreeableness - sympathetic, kind, cooperative, and warm.
- Conscientiousness - orderly, systematic, inefficient (reversed), sloppy (reversed), disorganised (reversed), and efficient.
- Emotional stability - envious (reversed), moody (reversed), touchy (reversed), jealous (reversed), temperamental (reversed), and fretful (reversed).
- Openness to experience - deep, philosophical, creative, intellectual, complex, imaginative.

The higher the score, the better that personality character trait describes the respondent.

Table 27: Derived personality variables

Variable	Description
_PNEXTRV	Personality scale - Extroversion
_PNAGREE	Personality scale - Agreeableness
_PNCONSC	Personality scale - Conscientiousness
_PNEMOTE	Personality scale - Emotional stability
_PNOPENE	Personality scale - Openness to experience

Religion

In wave 4, respondents were asked about their religion. *_religb* describes their broad religion classification (using the Australian Standard Classification of Religious Groups 1996).

Income Variables and Income Imputation

Income, Tax and Family Benefits Model

Figure 19, Figure 20 and Figure 21 show how the numerous income questions in the Person Questionnaire are combined together to form several financial year income components and a windfall income component on the responding person file, enumerated person file and household file respectively. The Family Tax Benefit and Maternity Allowance are calculated on the interim income to produce a total financial year income.[12] The Child Care Benefit is also calculated but not included in total financial year income (as it is considered a social transfer in kind rather than a cash benefit).[13]

Current wages and salaries and current benefits are asked about separately from the financial year questions.

Since Release 4, the income components have been imputed for both respondents and non-respondents within responding households. The enumerated file, as a result, contains component level data (rather than just total financial year income and windfall income as occurred in earlier releases). This has also permitted the calculation of these components at the household level as detailed in Figure 19. Market income, private income and Australian public transfers have also been calculated.

The HILDA income tax model calculates the financial year tax typically payable for an Australian taxpayer in the circumstances akin to those of the respondent. It does not attempt to calculate every individual variation in tax available under the Australian taxation system. Only the major components (income tax, business income tax, Medicare Levy, private pensions tax, deductions and offsets) contributing to income tax are estimated for the individual. When aggregated, these variables compare favourably with the national aggregates. The following key points should be noted about the income tax model:

- The input data are the imputed income variables and the data collected in the personal questionnaire. The components which the Australian Tax Office (ATO) treats as taxable income are summed: wages and salaries, business income, investment income and Australian pensions and benefits.

- Deductions are calculated as a percentage of income for 20 income ranges, the average deductions for each income range ranging from 6% for low incomes to 4% for the highest incomes (Taxation Statistics 1999-2000, ATO, 2002, CD Table s3.8). Gross income is reduced by deductions.

- Business income is separated from general income and then business tax is calculated. Business incomes up to $50,000 are taxed at the same rate as labour incomes. For business income exceeding $50,000 the rates applied are 15 percent up to $100,000, 10 percent up to $500,000 and 6 percent beyond $500,000. These rates reflect what is actually paid on business incomes (Taxation Statistics 1999-2000, ATO, 2002, CD Table s3.10).

[12] The Maternity Allowance is allocated to all families with newborn children and included in the Australian pensions and benefits.

[13] This change has been in place since Release 4.0. In earlier Releases, the Child Care Benefit was included in the total financial year income and the Maternity Allowance was only recorded if the respondent reported it.

- The four standard marginal tax rates are applied for non-retired people who earn just labour incomes (Table 28). A low income offset is incorporated into the rates for those earning up to $20,000.

- Low tax rates are applied to retired people. The rates we impute reflect what is actually paid by retired people on different levels of income (Taxation Statistics 1999-2000, ATO, 2002, CD Table s3.11). Non-respondents are presumed to be retired if aged over 65.

- The Medicare Levy is estimated as a flat 1.5%. HILDA does not collect private health insurance status, so the Medicare surcharge cannot be applied. An adjustment is made for seniors.

- As an approximation, low income pension and benefit recipients (taxable income less than $20,000) are deemed to pay no income tax.

- The largest offsets are dividend imputation and eligible termination payments, but these are not collected in HILDA. As an approximation, an average national offset of 2% of taxable income is applied as a flat rate to all taxpayers.

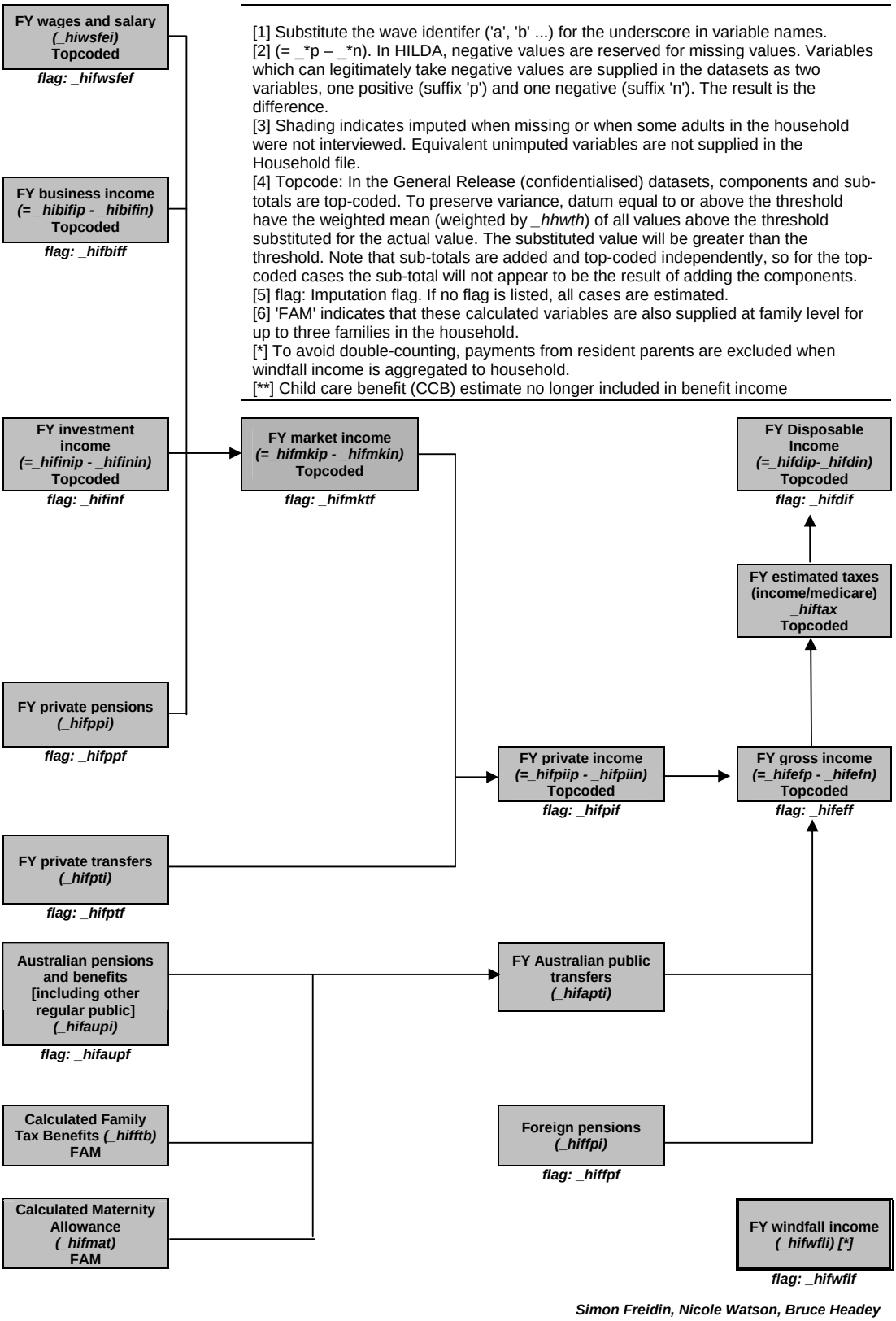
- As an approximation, private pensions are taxed at a flat rate of 5%. The same rate is applied to Workers' Compensation.

- Total income tax is calculated as the sum of income tax, business tax, Medicare Levy and private pensions tax less offsets.

Table 28: Australian Resident Income Tax Rates, Waves 1- 6

Wave	*Income*	*Tax Rate*
1, 2, 3 (Financial Years 2000-01, 2001-02, 2002-03)	$0 - $6000	Nil
	$6001 - $20000	Nil plus 17c for each $ over $6000
	$20001 - $50000	$2380 plus 30c for each $ over $20000
	$50001 - $60000	$11380 plus 42c for each $ over $50000
	$60001 and over	$15580 plus 47c for each $ over $60000
4 (Financial Year 2003-04)	$0 - $6000	Nil
	$6001 - $21600	Nil plus 17c for each $ over $6000
	$21601 - $52000	$2652 plus 30c for each $ over $21600
	$52001 - $62500	$11772 plus 42c for each $ over $52000
	$62501 and over	$16182 plus 47c for each $ over $62500
5 (Financial Year 2004-05)	$0 - $6000	Nil
	$6001 - $21600	Nil plus 17c for each $ over $6000
	$21601 - $58000	$2652 plus 30c for each $ over $21600
	$58001 - $70000	$13572 plus 42c for each $ over $58000
	$70001 and over	$18612 plus 47c for each $ over $70000
6 (Financial Year 2005-06)	$0 - $6000	Nil
	$6001 - $21600	Nil plus 15c for each $ over $6000
	$21601 - $63000	$2340 plus 30c for each $ over $21600
	$63001 - $95000	$14760 plus 42c for each $ over $63000
	$95001 and over	$28200 plus 47c for each $ over $95000

Figure 19: Financial Year Income Model: Household



[1] Substitute the wave identifer ('a', 'b' ...) for the underscore in variable names.
[2] (= _*p – _*n). In HILDA, negative values are reserved for missing values. Variables which can legitimately take negative values are supplied in the datasets as two variables, one positive (suffix 'p') and one negative (suffix 'n'). The result is the difference.
[3] Shading indicates imputed when missing or when some adults in the household were not interviewed. Equivalent unimputed variables are not supplied in the Household file.
[4] Topcode: In the General Release (confidentialised) datasets, components and sub-totals are top-coded. To preserve variance, datum equal to or above the threshold have the weighted mean (weighted by *_hhwth*) of all values above the threshold substituted for the actual value. The substituted value will be greater than the threshold. Note that sub-totals are added and top-coded independently, so for the top-coded cases the sub-total will not appear to be the result of adding the components.
[5] flag: Imputation flag. If no flag is listed, all cases are estimated.
[6] 'FAM' indicates that these calculated variables are also supplied at family level for up to three families in the household.
[*] To avoid double-counting, payments from resident parents are excluded when windfall income is aggregated to household.
[**] Child care benefit (CCB) estimate no longer included in benefit income

Simon Freidin, Nicole Watson, Bruce Headey

Figure 20: Financial Year Income Model: Enumerated Person

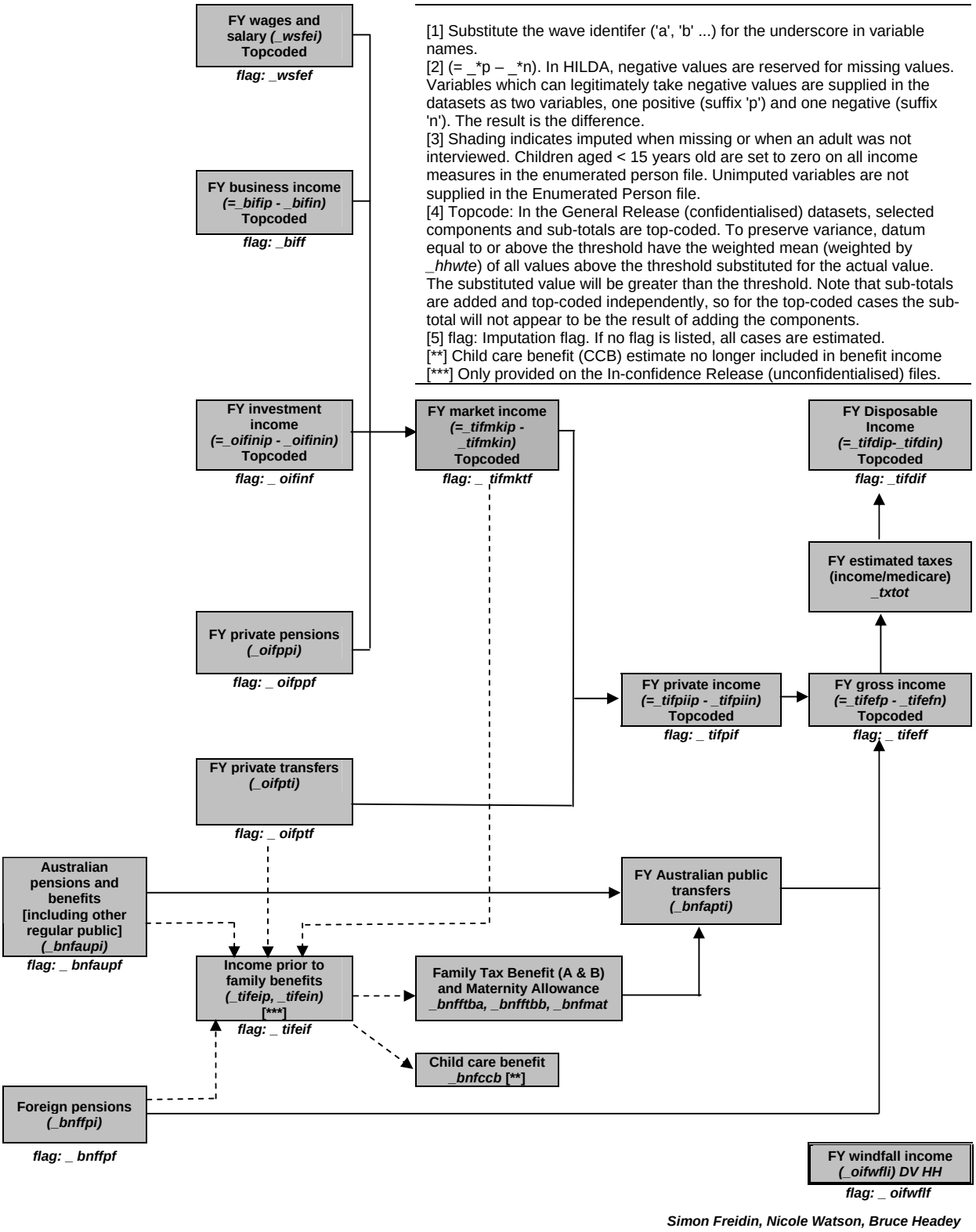
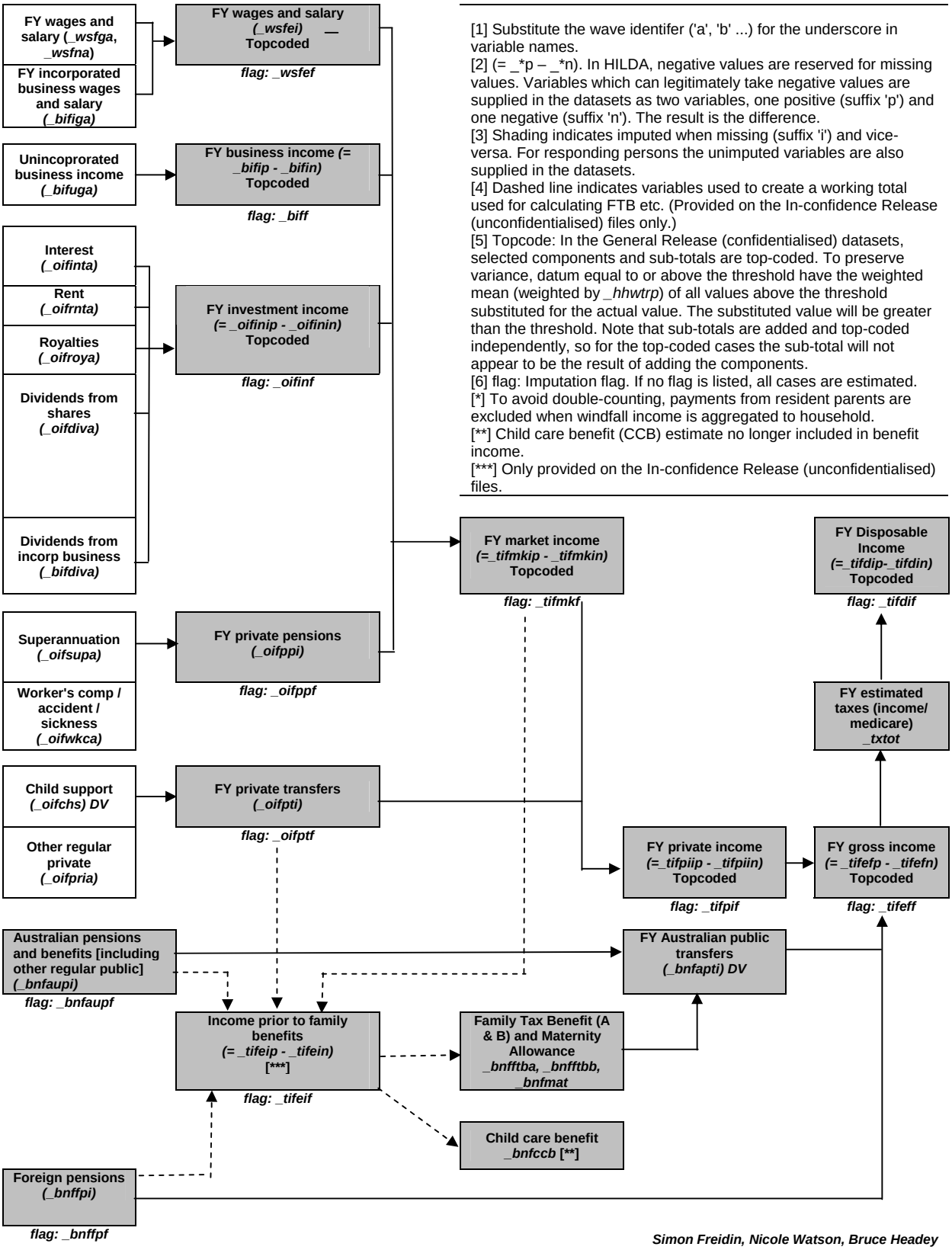


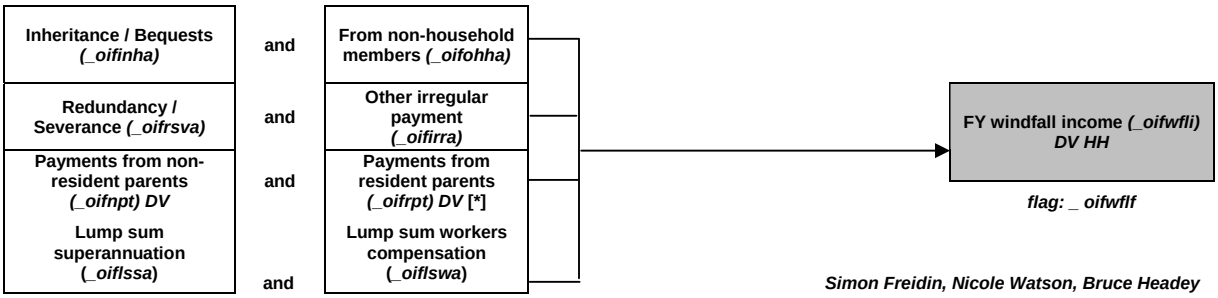
Figure 21: Financial Year Income Model: Responding Person



[1] Substitute the wave identifer ('a', 'b' ...) for the underscore in variable names.
[2] (= _*p – _*n). In HILDA, negative values are reserved for missing values. Variables which can legitimately take negative values are supplied in the datasets as two variables, one positive (suffix 'p') and one negative (suffix 'n'). The result is the difference.
[3] Shading indicates imputed when missing (suffix 'i') and vice-versa. For responding persons the unimputed variables are also supplied in the datasets.
[4] Dashed line indicates variables used to create a working total used for calculating FTB etc. (Provided on the In-confidence Release (unconfidentialised) files only.)
[5] Topcode: In the General Release (confidentialised) datasets, selected components and sub-totals are top-coded. To preserve variance, datum equal to or above the threshold have the weighted mean (weighted by *_hhwtrp*) of all values above the threshold substituted for the actual value. The substituted value will be greater than the threshold. Note that sub-totals are added and top-coded independently, so for the top-coded cases the sub-total will not appear to be the result of adding the components.
[6] flag: Imputation flag. If no flag is listed, all cases are estimated.
[*] To avoid double-counting, payments from resident parents are excluded when windfall income is aggregated to household.
[**] Child care benefit (CCB) estimate no longer included in benefit income.
[***] Only provided on the In-confidence Release (unconfidentialised) files.

Simon Freidin, Nicole Watson, Bruce Headey

Figure 21: Financial Year Income Model: Responding Person (c'td)



A list of additional derived income variables are provided in Table 29 (those that are directly related to the income imputation are provided later in Table 31). There are several issues to take note of in this table:

- Wages and salaries were asked of respondents for their main job, then for all their other jobs combined. The suffix 'g' and 'e' refer to gross and estimated gross incomes – where the respondent didn't know their gross income, their after tax income was asked for and this was translated back into an estimated gross income. The 'e' variables will have fewer cases with missing wages and salaries than the 'g' variables, as the 'e' variables include all the known 'g' values.

- The variable labels indicate when top-coding has occurred. The actual value replacing the top-coded value will be the weighted mean of the top-coded units (see section on Confidentialisation).

- Child support is calculated from the questions asked about the children in the family formation grid, rather than from the single category listed in the 'other income' question in the income section. This is because it is more likely the respondent would provide a more accurate response to the detailed questions rather than the broad 'catch all' question.

- The components feeding into the 'windfall' income are those thought irregular (such as inheritances, redundancies, payments from parents, lump sum superannuation payouts, lump sum workers compensation payouts).

- In wave 1, respondents were asked how different their current wage and salary income was from one year ago. This has been provided in dollar terms in *awsly*.

The imputation method and derived variables are discussed in the following sections.

Table 29: Other derived income variables

Variable	Description
Current wages and salaries and current benefits	
_WSCG	DV: All jobs, current gross wages per week ($). Weighted topcode.*
_WSCMG	DV: Main job, current weekly gross wages & salary ($). Weighted topcode.*
_WSCOG	DV: Other jobs, current weekly gross wages & salary ($). Weighted topcode.*
Financial year income – Unimputed variables	
AWSLY	DV: Gross weekly current wages & salary (from all jobs) one year ago ($)
_WSFG	DV: Financial year gross wages & salary ($).Weighted topcode.*
_WSFG	DV: Financial year gross wages & salary ($).Weighted topcode.*
_OIINT	DV: Financial year interest including nil ($)
_OIRNTN	DV: Financial year rental income including nil ($) Negative value
_OIRNTP	DV: Financial year rental income including nil ($) Positive value
_OIDIV	DV: Financial year dividends including nil ($)
_OIROY	DV: Financial year royalties including nil ($)
_OIDVRY	DV: Financial year dividends plus royalties including nil ($)
_TIFMKTP	DV: Financial year market (factor) income ($) Positive values. Weighted topcode.*
_TIFMKTN	DV: Financial year market (factor) income ($) Negative values
_TIFPRIP	DV: Financial year private income ($). Positive values. Weighted topcode.*
_TIFPRIN	DV: Financial year private income ($). Negative values
Financial year income – Estimated CCB, FTB A, FTB B, income tax and medicare levy	
_HIFCCB	DV: Household Child Care Benefit ($) financial year
_BNCCBF1	DV: Family number 1 Child Care Benefit ($) financial year
_BNCCBF2	DV: Family number 2 Child Care Benefit ($) financial year
_BNCCBF3	DV: Family number 3 Child Care Benefit ($) financial year
_BNFTAF1	DV: Family number 1 Family Tax Benefit A ($) financial year
_BNFTAF2	DV: Family number 2 Family Tax Benefit A ($) financial year
_BNFTAF3	DV: Family number 3 Family Tax Benefit A ($) financial year
_BNFTBF1	DV: Family number 1 Family Tax Benefit B ($) financial year
_BNFTBF2	DV: Family number 2 Family Tax Benefit B ($) financial year
_BNFTBF3	DV: Family number 3 Family Tax Benefit B ($) financial year
_BNMATF1	DV: Family number 1 Maternity Allowance ($) financial year
_BNMATF2	DV: Family number 2 Maternity Allowance ($) financial year
_BNMATF3	DV: Family number 3 Maternity Allowance ($) financial year
_TXINC	DV: Financial year taxes - income tax - estimate ($). Weighted topcode.*
_TXMED	DV: Financial year taxes - medicare - estimate ($). Weighted topcode.*

* See section on Confidentialisation for explanation of top-coding.

Imputation Method

Since Release 3, the primary method for imputing income is based on a method developed by Little and Su (1989). This longitudinal imputation method incorporates trend and individual level information into the imputed amounts by using a multiplicative model based on row (person) and column (wave) effects. The model is of the form:

imputation = (row effect) x (column effect) x (residual).

Ideally, the record with missing information (called the recipient) should be imputed using information from a record with complete information (called the donor) that has similar characteristics for the variable of interest. The Little and Su methodology was improved by extending it to take into account additional characteristics of the donors and recipients. Donors and recipients are matched within imputation classes which have similar characteristics. The imputation classes used were age groups defined by the following ranges: 15-19, 20-24, 25-34, 35-44, 45-54, 55-64, 65+.[14] The formulae for the Little and Su method are provided in Appendix 3, together with a worked example.

For some cases, such as new entrants interviewed in the latest wave who did not respond to some income questions, the imputation method used was the nearest neighbour regression method adopted in Release 2 (Watson, 2004a).

For respondents with item non-response (that is, where some questions during their interview were not answered), the income components have been imputed and the totals are the sum of the relevant components. These components and totals are available on the responding person file.

The income components for non-respondents within responding households have also been imputed. Prior to using the Little and Su method for non-respondents, income components were determined to be zero or non-zero using a population carryover method.[15] The Little and Su method was used to determine the non-zero amounts. However, for some cases, the Little and Su method could not be used (such as a non-responding new entrant in the latest wave). For these cases, the income totals were imputed first using the nearest neighbour regression method and then the income components were taken from the same donor. The components and totals for non-respondents are available on the enumerated person file (along with the components and totals for responding persons).

Imputed components and totals are also available at the household level on the household file.

Table 30 shows the proportion of missing cases that were imputed by each imputation method.[16] The proportions are summarized across all income variables that have been imputed. Ideally all records would be imputed by the Little and Su method and although this is mostly the case for responding persons sufficient information is not always available. The nearest neighbor regression method is a fall-back method that is used when

[14] Age groups were used to create the imputation classes because it is a simple characteristic and it is known for almost all donors and recipients. For a few cases, age was missing and was therefore imputed from a person with a similar relationship structure to the missing case. Not all income variables were imputed using imputation classes. The variables where donors and recipients were matched with imputation classes were current wages and salaries, current benefits, financial year wages and salaries, Australian Government pensions and rent income.

[15] Zeros are carried forward or backward from the surrounding waves with the same probability as that observed in the complete cases.

[16] For the proportion of cases which are missing, see Table 39.

the Little and Su method cannot be applied and is used more among enumerated persons as this group includes non-respondents within responding households. Non-responding cases are the only group that undergoes the additional population carryover imputation step.

Less information is available for non-responding persons within responding households when compared to responding persons and as a result the quality of the imputation is slightly poorer. However, the income components are still provided to enable these components to be available at the household level.

Improvements to the income imputation methodology are ongoing.[17] Further revisions to the income imputation methodology are expected.

Table 30: Proportion of missing cases imputed by imputation method

	Wave					
Imputation Method	1	2	3	4	5	6
Responding Persons						
Nearest Neighbour	13.7	4.3	5.4	4.6	4.3	7.2
Little & Su	86.3	95.7	94.6	95.4	95.7	92.8
Enumerated Persons						
Nearest Neighbour	54.7	37.4	40.3	39.6	41.2	49.1
Little & Su	32.9	38.8	39.5	36.2	40.5	39.2
Carryover	12.4	23.8	20.2	24.2	18.4	11.7

Imputed Income Variables

All income imputation was undertaken at the derived variable level, leaving the original data unchanged. In the main, both the pre-imputed and post-imputed variables are available in the datasets, along with an imputation flag, so that it is easy to choose between using the pre-imputed data or the post-imputed data.

An overview of the pre- and post-imputed income variables is provided in Table 31. We have deviated from the general style of presenting the derived variables in this manual in the hope that it is clearer from the following table how the post-imputed variables and flags relate to the pre-imputed variables.

[17] A detailed review of the performance of the imputation methods is reported in Starick and Watson (2007).

Table 31: Imputed income variables

	Pre-imputed	*Post-imputed*	*Flag*
Responding person file			
Current income			
Wages and salaries – all jobs	_wsce	_wscei	_wscef
Wages and salaries – main job	_wscme	_wscmei	_wscmef
Wages and salaries – other jobs	_wscoe	_wscoei	_wscoef
Benefits	_bncaup	_bncaupi	_bncaupf
Financial year income			
Wages and salaries	_wsfe	_wsfei	_wsfef
Australian govt pensions	_bnfaup	_bnfaupi	_bnfaupf
Foreign govt pensions	_bnffp	_bnffpi	_bnffpf
Business income	_bifn, _bifp	_bifin, _bifip	_biff
Investments	_oifinvn, _oifinvp	_oifinin, _oifinip	_oifinf
Private pensions	_oifpp	_oifppi	_oifppf
Private transfers	_oifpt	_oifpti	_oifptf
Total FY income [a]	Not provided	_tifefn, _tifefp	_tifeff
Windfall income	_oifwfl	_oifwfli	_oifwflf
Enumerated person file			
Current income			
Wages and salaries – all jobs	Not provided	_wscei	_wscef
Wages and salaries – main job	Not provided	_wscmei	_wscmef
Wages and salaries – other jobs	Not provided	_wscoei	_wscoef
Benefits	Not provided	_bncaupi	_bncaupf
Financial year income			
Wages and salaries	Not provided	_wsfei	_wsfef
Australian govt pensions	Not provided	_bnfaupi	_bnfaupf
Foreign govt pensions	Not provided	_bnffpi	_bnffpf
Business income	Not provided	_bifin, _bifip	_biff
Investments	Not provided	_oifinin, _oifinip	_oifinf
Private pensions	Not provided	_oifppi	_oifppf
Private transfers	Not provided	_oifpti	_oifptf
Total FY income [a]	Not provided	_tifefn, _tifefp	_tifeff
Windfall income	Not provided	_oifwfli	_oifwflf
Household file			
Current income			
Wages and salaries – all jobs	Not provided	_hiwscei	_hifwscef
Wages and salaries – main job	Not provided	_hiwscmi	_hifwscmf
Wages and salaries – other jobs	Not provided	_hiwscoi	_hifwscof
Benefits	Not provided	_hicaupi	_hicaupf
Financial year income			
Wages and salaries	Not provided	_hiwsfei	_hifwsfef
Australian govt pensions	Not provided	_hifaupi	_hifaupf
Foreign govt pensions	Not provided	_hiffpi	_hiffpf
Business income	Not provided	_hibifin, _hibifip	_hifbiff
Investments	Not provided	_hifinin, _hifinip	_hifinf
Private pensions	Not provided	_hifppi	_hifppf
Private transfers	Not provided	_hifpti	_hifptf
Total FY income	Not provided	_hifefn, _hifefp	_hifeff
Windfall income	Not provided	_hifwfli	_hifwflf

a. The following variables use total person financial year income (_tifefn, _tifefp) in their calculations: income tax (_txinc), and medicare (_txmed). Use _tifeff as imputation flag for these variables.

Wealth Variables and Wealth Imputation (Wave 2 and 6 Special Topic)

Wealth Model

In waves 2 and 6 a special wealth module was incorporated into the questionnaires. The Household Questionnaire contained the majority of the wealth questions and we endeavoured to ask these of the person knowing the most about the household finances. These questions covered the following topics:

- Cash and equity investments, trust funds, life insurance;
- Home and other property assets and debts;
- Business assets and debts;
- Children's bank accounts;[18]
- Collectables and vehicles; and
- Overdue household bills (in wave 6 only[19]).

Also, each respondent was asked some questions about their personal wealth in the Person Questionnaire, including:

- Bank accounts and credit card debt;
- Superannuation;
- HECS debt; and
- Other personal debts[20].

Figure 22 shows how the wealth components are combined together to form the total household wealth. The boxes with the broken lines highlight the variables that come from the Person Questionnaires. From Release 6, the imputation for non-respondents has been conducted at the wealth component level, so the household-level components are the sum of all persons in the household.[21]

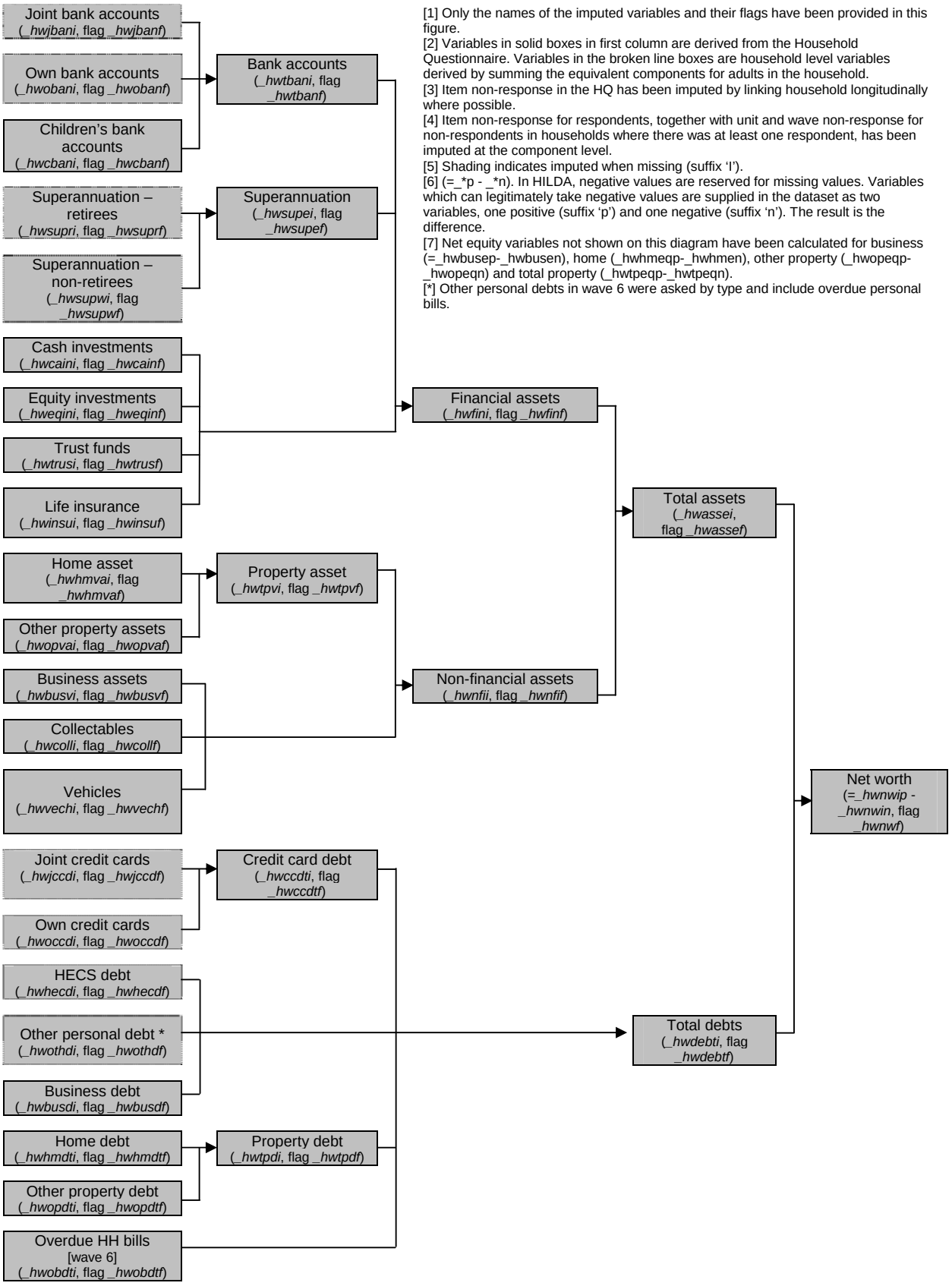
[18] That is, bank accounts of people in the household aged under 15.

[19] Overdue household bills were explicitly asked for in wave 6. It was assumed that this was captured in the 'any other debt' question asked in wave 2 (though perhaps not well).

[20] In wave 6, these other personal debts were asked for at a more disaggregated level and overdue personal bills were also explicitly asked for.

[21] For Release 2 to 5, please note that the imputation for non-respondents was only conducted at the total assets and debts level. As a result the household-level components that summed these person-level components was just the sum of ***responding*** persons only. This will explain some of the differences observed for these variables between releases.

Figure 22: Wealth Model Diagram, Household-level



[1] Only the names of the imputed variables and their flags have been provided in this figure.
[2] Variables in solid boxes in first column are derived from the Household Questionnaire. Variables in the broken line boxes are household level variables derived by summing the equivalent components for adults in the household.
[3] Item non-response in the HQ has been imputed by linking household longitudinally where possible.
[4] Item non-response for respondents, together with unit and wave non-response for non-respondents in households where there was at least one respondent, has been imputed at the component level.
[5] Shading indicates imputed when missing (suffix 'I').
[6] (=_*p - _*n). In HILDA, negative values are reserved for missing values. Variables which can legitimately take negative values are supplied in the dataset as two variables, one positive (suffix 'p') and one negative (suffix 'n'). The result is the difference.
[7] Net equity variables not shown on this diagram have been calculated for business (=_hwbusep-_hwbusen), home (_hwhmeqp-_hwhmen), other property (_hwopeqp-_hwopeqn) and total property (_hwtpeqp-_hwtpeqn).
[*] Other personal debts in wave 6 were asked by type and include overdue personal bills.

Several equity variables (assets less debts) not described in the previous figure are provided on the household file. These are business equity, home equity, other property equity, and total property equity.

These variables, together with the unimputed versions of the sub-totals described in Figure 22 are provided in Table 32 (variables relating directly to the wealth imputation are provided later in Table 35).

Table 32: Other derived wealth variables at household-level

Variable	Description
_HWBEIP	DV: Household wealth: Business: equity [positive values] [imputed] ($). Weighted topcode
_HWBEIN	DV: Household wealth: Business: equity [negative values] [imputed] ($)
_HWBEF	DV: Imputation flag Household wealth: Business: equity ($)
_HWBUSEP	DV: Household wealth: Business: equity [positive values] ($). Weighted topcode
_HWBUSEN	DV: Household wealth: Business: equity [negative values] ($)
_HWHMEIP	DV: Household wealth: Home: Equity [positive values] [imputed] ($). Weighted topcode
_HWHMEIN	DV: Household wealth: Home: Equity [negative values] [imputed] ($)
_HWHMEF	DV: Imputation flag Household wealth: Home: Equity ($)
_HWHMEQP	DV: Household wealth: Home: Equity [positive values] ($). Weighted topcode
_HWHMEQN	DV: Household wealth: Home: Equity [negative values] ($)
_HWOPEIP	DV: Household wealth: Other property: Equity [positive values] [imputed] ($). Weighted topcode
_HWOPEIN	DV: Household wealth: Other property: Equity [negative values] [imputed] ($)
_HWOPEF	DV: Imputation flag Household wealth: Other property: Equity ($)
_HWOPEQP	DV: Household wealth: Other property: Equity [positive values] ($). Weighted topcode
_HWOPEQN	DV: Household wealth: Other property: Equity [negative values] ($)
_HWTPEIP	DV: Household wealth: Total property: Equity [positive values] [imputed] ($). Weighted topcode
_HWTPEIN	DV: Household wealth: Total property: Equity [negative values] [imputed] ($)
_HWTPEF	DV: Imputation flag Household wealth: Total property: Equity ($)
_HWTPEQP	DV: Household wealth: Total property: Equity [positive values] ($). Weighted topcode
_HWTPEQN	DV: Household wealth: Total property: Equity [negative values] ($)
_HWTPVAL	DV: Household wealth: Total property: Value ($). Weighted topcode
_HWHMHL	DV: Household wealth: Home: Home loans ($)
_HWHMOL	DV: Household wealth: Home: Other loans ($)
_HWHMEQL	DV: Household wealth: Home: Equity loan ($)
_HWTPDT	DV: Household wealth: Total property: Debt ($). Weighted topcode
_HWSUPER	DV: Household wealth: Total superannuation ($). Weighted topcode
_HWTBANK	DV: Household wealth: Total bank accounts ($). Weighted topcode
_HWFIN	DV: Household Financial Assets ($). Weighted topcode
_HWNFIN	DV: Household Non-Financial Assets ($). Weighted topcode

Imputation Method

With the repeat of the wealth questions in wave 6, we can now impute the wealth variables via a longitudinal method.[22]

Person-level wealth variables are imputed using the Little and Su method. The nearest neighbour regression method[23] is used as a fall-back method when the Little and Su method does not work. Wealth variables for non-respondents have been imputed at the wealth component level (rather than for total assets and total debts as was done in Release 2 through to 5) so that the household wealth can be provided at the component level by summing the person-level wealth of the adults in the household.

Household-level wealth variables are also imputed using the Little and Su method, where the household can be longitudinally linked[24] between waves 2 and 6 (*bhwlink* is an indicator variable for whether the wave 2 household could be linked to a wave 6 household). Otherwise the nearest neighbour regression method is used.

The household-level wealth variable for home value was collected in all waves of the survey. This variable was imputed through waves 1 to 6 via the same method used in the two wave wealth module approach. Households were longitudinally linked from one wave to another (*_hhlink* is an indicator variable showing whether a household was linked to another household in the next wave) with the Little and Su method applied wherever possible. The nearest neighbor regression method was used as a fall-back method again.

Where we knew the wealth component was non-zero (for example, a screener question identified them as having the particular wealth component), only a non-zero donor was permitted. For situations where it was unknown if the individual or household had the asset, a filter process[25] was used to determine if they received a non-zero value or not.

Table 33 and Table 34 show the proportion of missing cases that were imputed by each imputation method.[26] In the first table the proportions are summarized across all wealth variables that have been imputed. As with income it is preferable to have all records imputed by the Little and Su method but with only two waves of wealth data sufficient information was not always available. Non respondents in the enumerated person group were less likely to be imputed by the Little and Su method (for similar reasons as in income imputation) and any households not linked between the two waves, regardless of their situation, were imputed via the nearest neighbor regression method. Table 34 shows a much higher proportion of imputed records via the Little and Su method for home value due to better household linking between consecutive waves (rather than the four-wave difference experienced with the wealth module imputation).

[22] For Release 2 through to 5, the wealth imputation for wave 2 used a nearest neighbour regression method (see Watson, 2004a for more details).
[23] The reported value of a donor with the nearest predicted value from a regression model to the recipient's is used to replace the missing value.
[24] Households have been linked over time for the purposes of imputing wealth if they have the same adult (aged 18+) membership. Deaths are not counted as a membership change.
[25] The nearest neighbor regression method was run prior to the normal imputation step. All unknown records who were assigned a donor with a zero value were considered not to have the wealth item. All other records were imputed with a non-zero amount and the Little and Su method was used where possible to determine the non-zero amount.
[26] For the proportion of cases which are missing, see Table 41 and Table 42.

Table 33: Proportion of missing cases imputed by imputation method (wealth), waves 2 and 6

	Wave	
Imputation Method	2	6
Person level wealth items (responding persons)		
Nearest Neighbour	38.1	40.8
Little & Su	61.9	59.2
Person level wealth items (enumerated persons)		
Nearest Neighbour	73.3	67.7
Little & Su	26.7	32.3
Household level wealth items		
Nearest Neighbour	56.4	62.6
Little & Su	43.6	37.4

Table 34: Proportion of missing cases imputed by imputation method (home value), waves 1 - 6

	Wave					
Imputation Method	1	2	3	4	5	6
Home value (households)						
Nearest Neighbor	26.0	5.3	15.2	14.4	14.6	32.7
Little & Su	74.0	94.7	84.8	85.6	85.4	67.3

Imputed Wealth Variables

Table 35 outlines the imputed wealth variables included on the wave 2 and 6 files.

Table 35: Imputed wealth variables

	Pre-imputed	*Post-imputed*	*Flag*
Responding person file			
Assets			
Joint bank accounts	_pwjbank	_pwjbani	_pwjbanf
Own bank accounts	_pwobank	_pwobani	_pwobanf
Superannuation – retirees	_pwsupr	_pwsupri	_pwsuprf
Superannuation – non-retirees	_pwsupwk	_pwsupwi	_pwsupwf
Debts			
HECS debt	_pwhecdt	_pwhecdi	_pwhecdf
Joint credit cards	_pwjccdt	_pwjccdi	_pwjccdf
Own credit cards	_pwoccdt	_pwoccdi	_pwoccdf
Other personal debt	_pwothdt	_pwothdi	_pwothdf

Table 35: (c'td)

	Pre-imputed	*Post-imputed*	*Flag*
Enumerated person file			
Assets			
Joint bank accounts	Not provided	_pwjbani	_pwjbanf
Own bank accounts	Not provided	_pwobani	_pwobanf
Superannuation – retirees	Not provided	_pwsupri	_pwsuprf
Superannuation – non-retirees	Not provided	_pwsupwi	_pwsupwf
Debts			
HECS debt	Not provided	_pwhecdi	_pwhecdf
Joint credit cards	Not provided	_pwjccdi	_pwjccdf
Own credit cards	Not provided	_pwoccdi	_pwoccdf
Other personal debt	Not provided	_pwothdi	_pwothdf
Household file			
Assets			
Joint bank accounts	_hwjbank	_hwjbani	_hwjbanf
Own bank accounts	_hwobank	_hwobani	_hwobanf
Children's bank accounts	_hwcbank	_hwcbani	_hwcbanf
Superannuation – retirees	_hwsupr	_hwsupri	_hwsuprf
Superannuation – non-retirees	_hwsupwk	_hwsupwi	_hwsupwf
Business assets	_hwbusva	_hwbusvi	_hwbusvf
Cash investment	_hwcain	_hwcaini	_hwcainf
Equity investment	_hweqinv	_hweqini	_hweqinf
Collectables	_hwcoll	_hwcolli	_hwcollf
Home asset	_hwhmval	_hwhmvai	_hwhmvaf
Other property assets	_hwopval	_hwopvai	_hwopvaf
Life insurance	_hwinsur	_hwinsui	_hwinsuf
Trust funds	_hwtrust	_hwtrusi	_hwtrusf
Vehicles value	_hwvech	_hwvechi	_hwvechf
Total household assets	_hwasset	_hwassei	_hwassef
Debts			
HECS debt	_hwhecdt	_hwhecdi	_hwhecdf
Joint credit cards	_hwjccdt	_hwjccdi	_hwjccdf
Own credit cards	_hwoccdt	_hwoccdi	_hwoccdf
Other personal debt	_hwothdt	_hwothdi	_hwothdf
Business debt	_hwbusdt	_hwbusdi	_hwbusdf
Home debt	_hwhmdt	_hwhmdti	_hwhmdtf
Other property debt	_hwopdt	_hwopdti	_hwopdtf
Overdue household bills[1]	_hwobdt	_hwobdti	_hwobdtf
Total household debts	_hwdebt	_hwdebti	_hwdebtf
Net worth	_hwnetwp, _hwnetwn	_hwnwip, _hwnwin	_hwnwf

1. Variable only in wave 6.

Weights

Cross-Sectional Weights

Wave 1

In wave 1, we essentially had a complex cross-sectional survey. The initial (or design) weights are derived from the probability of selecting the households into the sample. These household weights are initially adjusted according to information collected about all selected households (both responding and non-responding) and further adjusted so that weighted household estimates from the HILDA Survey match several known household-level benchmarks.

The person-level weights are based on the household-level weights, with adjustments made based on information collected about all the people listed in the responding households. These weights are also adjusted to ensure that the weighted person estimates match several known person-level benchmarks.

The benchmarks were reviewed for Release 4 and these changes have been carried over to later releases. The changes made to the weighting process include:

- The household and enumerated person weights are determined at the same time (rather than sequentially as was done in earlier releases). This is known as integrated weighting. The weights are adjusted to the household benchmarks at the same time as they are adjusted to the enumerated person benchmarks. The household weight will be the same as the enumerated weight for each person in the household, resulting in identical estimates where the same concept can be determined from the two files.[27]

- Due to the demands placed on the weights through the integrated weighting process, some of the benchmarks used have been simplified. Also, following some concerns about the representativeness of the sample, some additional benchmarks on marital status and occupation have been included (based on the ABS Labour Force Survey).

In summary, the household benchmarks have been revised to:

- Number of adults by number of children; and

- State by part of State.[28]

The enumerated person benchmarks have been revised to:

- Sex by broad age;

[27] For example, the number of people living in a household with two people can be derived by two methods. Firstly, this can be calculated from the household file by estimating the number of two person households and multiplying by two. Secondly, it can be estimated from the enumerated file by summing the weights of people living in two person households.

[28] Prior to Release 4, the household benchmarks were number of adults by number of children by **broad geography** and State by part of State (the bolded text indicates what has been dropped).

- State by part of State;
- Labour force status; and
- Marital status.[29]

The responding person benchmarks were simplified in Release 6. They have been revised to:

- State by part of State;
- Sex by broad age;
- State by labour force status;
- Marital status; and
- Occupation.[30]

The person benchmarks for State, part of State, sex and age are from the Estimated Residential Population figures produced by the ABS based on the 2001 Census, updated for births, deaths, immigration, emigration and interstate migration. The household benchmarks are now also based on the 2001 Census and are similarly updated from that time point.[31] The remaining benchmarks come from the ABS Labour Force Survey.

From Release 5 onwards, the very remote parts of New South Wales, Queensland, South Australia, Western Australia and the Northern Territory have been excluded from the benchmarks, which is in line with the practice adopted in similar large-scale surveys run by the ABS.[32] Information about the other aspects of the weighting procedure can be found in Watson and Fry (2002).

Wave 2 onwards

From wave 2 onwards, the 'selection' of the sample is dependent on the wave 1 responding sample and the household and individual attrition after wave 1. The cross-sectional weights for wave 2 onwards 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 from wave 1 into the relevant wave household. Following this, non-response adjustments are made which require within-sample modelling of non-response probabilities and benchmarking to known population estimates at both the household and person level.

[29] Prior to Release 4, the enumerated person benchmarks were **State by part of State** by sex by broad age, and **State by part of State** by labour force status (the bolded text indicates what has been dropped, but note that State by part of State is now included as a separate benchmark). The marital status benchmark has been added from Release 4.
[30] Prior to Release 6, the responding person benchmarks were **State by part of State** by sex by broad age; State by **part of State** by labour force status; marital status by **broad age**; and occupation by **broad geography** (the bolded text indicates what has been dropped, but note that State by part of State is now included as a separate benchmark). From Release 4 the marital status and occupation benchmarks have been included.
[31] Prior to Release 5, only household estimates based on the 1996 Census were available.
[32] Prior to Release 5, only the sparsely settled parts of the Northern Territory were excluded.

The benchmarks used in Release 4 and later have been amended as described above. Other aspects of the weighting process for wave 2 onwards are detailed in Watson (2004b).[33]

Longitudinal Weights

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.

In Release 4 and later, the benchmarks for the longitudinal weights have been modified to mirror those used in the cross-section weights.

The longitudinal enumerated person benchmarks have been revised to:

- Sex by broad age;
- State by part of State;
- Labour force status; and
- Marital status.[34]

The longitudinal responding person benchmarks have been revised to mirror the Release 6 changes made to the cross-sectional responding person weights. The benchmarks are:

- Sex by broad age;
- State by part of State;
- State by labour force status;
- Marital status; and
- Occupation.[35]

From Release 6, we have provided longitudinal weights for the balanced panel of responding persons or enumerated persons from every wave to every other wave and for the balanced panel of any combination of a pair of waves.[36] These weights adjust for attrition from the initial wave and are benchmarked back to the key characteristics of the initial wave. For instance if you were interested in a panel of respondents from waves 2 through 6, the weight provided for this panel would adjust for attrition from the balanced

[33] While this paper is written in relation to the wave 2 weighting, the process in later waves follows the same methodology.
[34] Prior to Release 4, the longitudinal enumerated person benchmarks were **State by part of State** by sex by age broad, and **State by part of State** by labour force status (the bolded text indicates what has been dropped, but note that State by part of State is now included as a separate benchmark). The marital status benchmark has been added from Release 4 onwards.
[35] Prior to Release 6, the longitudinal responding person benchmarks were **State by part of State** by sex by broad age; State by **part of State** by labour force status; marital status by **broad age**; and occupation by **broad geography** (the bolded text indicates what has been dropped, but note that State by part of State is now included as a separate benchmark). From Release 4 the marital status and occupation benchmarks have been included.
[36] Prior to Release 6, weights were only provided for the balanced panel of respondents or enumerated persons from wave 1 to every other wave.

panel from wave 2 to 6 and would ensure key characteristics of the wave 2 population are matched.

Other aspects of the longitudinal weights are described in Watson (2004).

Replicate Weights

Replicate weights have been provided for users to calculate standard errors that take into account the complex sample design of the HILDA Survey. These weights can be used by the SAS GREGWT macro, the STATA 'svy jackknife' commands (more detail is provided below on *Calculating Standard Errors*), or you can write your own routine to use these weights. As of Release 6, weights for 45 replicate groups are provided.

Weights Provided on the Data Files

Table 36 provides a list of the weights provided on the data files together with a description of those weights. The longitudinal weights provided on the enumerated and responding person files are the ones you are most likely to use, though other longitudinal weights are provided on the Longitudinal Weights File.

Irrespective of the changes made to the construction of the weights, some changes are expected to the weights with each new 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 (very occasionally).

Table 36: Weights

File	*Weights*	*Description*
Household File	_hhwth	The household weight is the cross-section population weight for all households responding in the relevant wave. Note the sum of these household weights for wave 1 is approximately 7.4 million.
	_hhwths	This is the cross-section household population weight rescaled to the sum of the sample size for the relevant wave (i.e. 7682 responding households in wave 1). Use this weight when the statistical package requires the weights to sum to the sample size.
	_hhwte01 to _hhwte16	The enumerated person weights are provided on both the household file and the enumerated person file. See description below.
	_rwh1 to _rwh45	Cross-section household population replicate weights.
Enumerated Person File	_hhwte	The enumerated person weight is the cross-section population weight for all people who are usual residents of the responding households in the relevant wave (this includes children, non-respondents and respondents). The sum of these enumerated person weights for wave 1 is 19.0 million.
	_hhwtes	This is the cross-section enumerated person population weight rescaled to the sum of the sample size for the relevant wave (i.e. for wave 1, 19914 enumerated persons). Use this weight when the statistical package requires the weights to sum to the sample size.

Table 36: (c'td)

File	*Weights*	*Description*
	_lnwte	This longitudinal enumerated person weight is the longitudinal population weight for all people who were enumerated (i.e. in responding households) each wave from wave 1 to the wave where this variable resides. This weight applies to children, non-respondents, intermittent respondents, and full respondents in responding households. blnwte is for the balanced panel of enumerated persons from wave 1 to 2; clnwte is for the balanced panel from wave 1 to 3; dlnwte is for the balanced panel from wave 1 to 4, etc. These variables are also on the *Longitudinal Weights File*, but are named differently: *wlea_b*; *wlea_c*; *wlea_d*, etc. We expect to drop _lnwte in future.
	_rwe1 to _rwe45	Cross-section enumerated person population replicate weights.
	_rwlne1 to _rwlne45	Longitudinal enumerated person population replicate weights
Responding Person File	_hhwtrp	The responding person weight is the cross-section population weight for all people who responded in the relevant wave (i.e. they provided a personal interview). The sum of these responding person weights for wave 1 is 15.0 million.
	_hhwtrps	This is the cross-section responding person population weight rescaled to sum to the number of responding persons in the relevant wave (i.e. 13,969 in wave 1). Use this weight when the statistical package requires the sum of the weights to be the sample size.
	_lnwtrp	This longitudinal responding person weight is the longitudinal population weight for all people responding (i.e. provided an interview) each wave from wave 1 to the wave where this variable resides. blnwtrp is for the balanced panel of respondents from wave 1 to 2; clnwtrp is for the balanced panel from wave 1 to 3; dlnwtrp is for the balanced panel from wave 1 to 4, etc. These variables are also on the *Longitudinal Weights File*, but are named differently: *wlra_b*; *wlra_c*; *wlra_d*, etc. We expect to drop _lnwtrp in future.
	_rwrp1 to _rwrp45	Cross-sectional responding person population replicate weights
	_rwlnr1 to _rwlnr45	Longitudinal responding person population replicate weights.
Longitudinal Weights File*	wle*t1_tn*	Longitudinal enumerated person weight for the balanced panel of all people who were enumerated (i.e. part of a responding household) each wave from wave *t1* to *tn*. Wave letters are used in place to *t1* and *tn*. For example, wlec_f is the longitunal enumerated person weight for the balanced panel from wave 3 to 6.
	wle*t1tn*	Longitudinal enumerated person weight for the balanced panel of all people who were enumerated (i.e. part of a responding household) in wave *t1* and *tn*. Wave letters are used in place to *t1* and *tn*. The paired longitudinal weights do not restrict individuals in any way based on their response status in waves between *t1* and *tn*. For example, wlecf is the longitunal enumerated person weight for the balanced panel of enumerated people in wave 3 and 6 (they may or may not have been enumerated in other waves).

Table 36: (c'td)

File	*Weights*	*Description*
	wlr*t1_tn*	Longitudinal responding person weight for the balanced panel of all people who were interviewed each wave from wave *t1* to *tn*. Wave letters are used in place to *t1* and *tn*. For example, wlrc_f is the longitunal responding person weight for the balanced panel of respondents from wave 3 to 6.
	wlr*t1tn*	Longitudinal responding person weight for the balanced panel of all people who were interviewed in wave *t1* and *tn*. Wave letters are used in place to *t1* and *tn*. The paired longitudinal weights do not restrict individuals in any way based on their response status in waves between *t1* and *tn*. For example, wlrcf is the longitunal responding person weight for the balanced panel of respondents in wave 3 and 6 (they may or may not have been responding in other waves).

* Replicate weights for the weights provided on the Longitudinal Weights File are available on request. Email hilda-inquiries@unimelb.edu.au.

Advice on Using Weights

Which Weight to Use

For some users, the array of weights on the dataset may 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 4 interviews (either at the household level or person level) then you would use the wave 4 cross-section weights. Similarly, if you are only using wave 3 information, then you would use the wave 3 cross-section weights, and so on. If you want to infer how people have changed across the five years between waves 1 and 6, then you would use the longitudinal weights for waves 1 through 6.

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

- What proportion of households rent in 2006? We would use the cross-section household weight for wave 6 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 2003? This is a question that can only be answered from the responding person file using the cross-section responding person weight for wave 3. We would identify those reportedly

working in professional occupations and take the weighted average of their wages and salaries.

- For how many years have people been poor between 2001 and 2006? 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 years people were poor between wave 1 and wave 6, and apply the longitudinal enumerated person weight (*flnwte* or equivalently *wlea_f*) for those people enumerated every wave between wave 1 and 6.

- What proportion of people have changed their employment status between 2002 and 2006? This question can only be answered by considering the responding persons in both waves. We would use the longitudinal responding person weight for the pair of waves extracted from the Longitudinal Weight File (*wlrbf*) and construct a weighted cross-tabulation of the employment status of respondents in wave 2 against the employment status of respondents in wave 6.

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 HILDA survey has a complex survey design that needs to be taken into account when calculating standard errors. 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 later waves.

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 later waves have not been produced.

- Use of the SPSS Release 12 add-on module "SPSS Complex Samples". The add-on module produces standard errors via the Taylor Series approximation. SPSS does not have a built in feature to handle replicates weights.

- Use of SAS procedures SURVEYMEANS, SURVEYREG, SURVEYFREQ and SURVEYLOGISTIC (the last two only in version 9 onwards). The SAS procedures produce standard errors via the Taylor Series approximation. SAS does not have a built in feature to handle replicates weights.

- Use of GREGWT macro in SAS – Some users within FaHCSIA, ABS 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 using the replicate weights.

- Use of 'svy' commands in STATA – Stata has a set of survey commands that deal with complex survey designs. Using the 'svyset' commands, the clustering, stratification and weights can be assigned. You can request the standard errors be calculated using the Jackknife method using 'svy jackknife' and the replicate weights. 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.

A User Guide for calculating the standard errors in HILDA is provided as part of our technical paper series, see Hayes (2008). Example code is provided in SAS, SPSS and STATA.

To assist you in the calculation of appropriate standard errors, the wave 1 area (cluster), and proxy stratification variables have been included on the master file. These are listed in Table 37 and need to be specified for the SPSS, SAS and Stata Taylor Series approximation standard error calculations suggested above. Any new entrants to the household are assigned to the same sample design information as the permanent sample member. As of Release 6 the proxy stratification variable (*ahhstrat*) has replaced major statistical region (*ahhmsr*) on the master file as the variable to be used in the Taylor Series approximation method. The new stratification variable is essentially a collapsed area unit variable that approximates the effect of both the systematic selection and stratification of the survey selection better than only using the variable for the major statistical region.

Table 37: Sample design variables

Variable	Description	Design element
AHHRAID	DV: randomised area id	Cluster
AHHSTRAT	DV: Wave 1 Strata	Proxy stratification

Also, a few users may be interested in the sample design weight in wave 1 before any benchmark or non-response adjustments have been made. This is available on the household file as *ahhwtdsn*.

DOCUMENTATION

Documentation Choices

Before you get lost in the array of documentation, it is worth pausing to consider how you work and what documentation is available to you. You will not need to look at all pieces of documentation that have been prepared in order to use the datasets efficiently.

There are four main pathways through the documentation:

- Marked-up questionnaires and the derived variable coding framework for each wave – you would use these if you were familiar with the questionnaires and wanted to know what extra variables have been included;
- File-based coding framework for each wave – you would use this if you were roughly aware of what variables were in what files and were interested in a range of different topics;
- Subject-level coding framework for each wave – you would use this if you were interested in a couple of different topics;
- Cross-wave variable listing – you would use this if you were frequently using variables across the various waves, and were happy to find out the codes used when you started using the variables.

The codeing frameworks have been provided on the DVD (as .pdf documents) as well as via an on-line data dictionary.

You should also consider which files you want to print out and which you are happy to look at electronically. We have found that the marked-up questionnaires are best printed. The rest are best looked at on screen where there are search functions available.[37]

While frequencies of the variables have been provided, it is expected that you might only refer to these files for some simple queries with the variable name in mind (for example, how many employed people do we have in the sample, or what are the codes used for question R3).

Also, as you may have already seen, the previous chapter of this manual provides an overview of the topics covered in the questionnaires and the derived variables created.

These tools are described in more detail below.

Marked-Up Questionnaires

Beside each question in the questionnaires, the associated variable name has been added. Derived variables are not included, only the variables that relate directly to the question asked. See Figure 23 for an example.

---

[37] In Adobe Acrobat, you would begin a search by clicking on the button that looks like this:

.

Figure 23: Example of the marked-up questionnaires

H9 **I am now going to ask you about the amount of <u>contact</u> you have with your (youngest) child who lives elsewhere.**

About how many <u>nights</u> each week, fortnight or month does this child usually stay overnight with you?

If respondent refers to weeks rather than nights, record number of full weeks instead of nights.

If overnight contact is sparse, interviewer to get estimate for 3, 6 or 12 month period.

Zero overnight stays in a year.....997

Else:

Record number of nights		Record number of full weeks
	OR	
Per week..... 1		per…
Fortnight..... 2		Fortnight..... 2
4 weeks..... 3		4 weeks..... 3
3 months 4		3 months 4
6 months 5		6 months 5
Year.....6		Year 6

```
ANCNGTH   H9 Youngest non-resident
              child overnight stays -
              answered nights or weeks
ANCNGTN   H9 Youngest non-resident
              child overnight stays -
              number of nights
ANCNGTNP  H9 Youngest non-resident
              child overnight stays -
              nights - period
ANCNGTW   H9 Youngest non-resident
              child overnight stays -
              number of weeks
ANCNGTWP  H9 Youngest non-resident
              child overnight stays -
              weeks - period
```

Variable Listings

Derived Variable Listing

The derived variable listing contains all the extra variables created from those collected in the questionnaires. This listing shows the following:

- on which file the variable can be found;
- the variable name;
- the label describing the variable;
- what values the variable can take; and
- to which population the variable relates.

Figure 24 shows the derived variable associated with the variables listed on the questionnaire in Figure 23.

Figure 24: Example of the derived variable listing

File	Variable	Data Item	Categories	Population
Children - Non Resident: Derived Variables				
RP	ANCNGT	DV: Overnight stays of non-resident child (Days per annum)	[Days]	Respondents with non-resident children
Conversion from days per week, fortnight etc to days per annum. If answered in weeks, converted to days per annum				

File-Based Listing

For each file provided (except for the Combined File), there is a file-based variable listing. This listing contains:

- the questionnaire and question number;
- the variable name and label describing the variable;
- the values that each variable can take;
- the population to which the variable relates; and
- for derived variables, a brief explanation of how the variable was derived.

In this listing, the derived variables are interspersed with the variables directly from the questionnaires. See Figure 25 below:

Figure 25: Example of the file-based listing

Qn	*Variable*	*Data Item*	*Categories*	*Population*
PQ H9	ANCNGTH	Youngest non-resident child overnight stays - answered nights or weeks	0 Zero day visits in a year 1 Nights 2 Weeks	Parents with resident children aged 24 or less who have a non resident parent
PQ H9	ANCNGTN	Youngest non-resident child overnight stays - number of nights	[Nights] 0 Zero day visits in a year	Parents with resident children aged 24 or less who have a non resident parent

Subject Listing

The subject listing is similar to the file-based listing, but includes the variables of all files together in one listing. There is an index at the beginning and the broad subject name is at the top of each page to help you navigate through the 170 to 300 page document. See Figure 26 below:

Figure 26: Example of the subject listing

File		Variable	Data Item	Categories	Population
RP	PQ H9	ANCNGTH	Youngest non-resident child overnight stays - answered nights or weeks	0 Zero day visits in a year 1 Nights 2 Weeks	Parents with resident children aged 24 or less who have a non -resident parent
RP	PQ H9	ANCNGTN	Youngest non-resident child overnight stays - number of nights	0 Zero day visits in a year [Nights]	Parents with resident children aged 24 or less who have a non -resident parent

Cross-Wave Variable Listing

The cross-wave variable listing is probably the most useful tool of all the documentation options. It provides information on the file where the variable can be found, the label and in which wave the variable has been asked. For the particular example provided in Figure 27, we can see that these questions have changed from section H in wave 1 to section G in later waves, and that the question numbering has changed slightly between waves 2 and 3.

Figure 27: Example of the cross-wave variable listing

File	Variable	Data Item	Wave 1	2	3	4	5	6
RP	_NCNGTH	Youngest non-resident child overnight stays - answered nights or weeks	PQ H9	PQ G9	PQ G9b	PQ G9b	PQ G11b	PQ G9b
RP	_NCNGTN	Youngest non-resident child overnight stays - number of nights	PQ H9	PQ G9	PQ G9b	PQ G9b	PQ G11b	PQ G9b
RP	_NCNGTNP	Youngest non-resident child overnight stays - nights - period	PQ H9	PQ G9	PQ G9b	PQ G9b	PQ G11b	PQ G9b
RP	_NCNGTW	Youngest non-resident child overnight stays - number of weeks	PQ H9	PQ G9	PQ G9b	PQ G9b	PQ G11b	PQ G9b
RP	_NCNGTWP	Youngest non-resident child overnight stays - weeks - period	PQ H9	PQ G9	PQ G9b	PQ G9b	PQ G11b	PQ G9b

Frequencies

The frequencies are a simple listing of the categories for each question and the number of cases falling into each category. Figure 28 provides an example of the listing.

Figure 28: Example of the frequencies

ancngtnp H9 Youngest non-resident child overnight stays - nights - period

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 Week	62	.4	20.5	20.5
	2 Fortnight	104	.7	34.4	55.0
	3 4 weeks	55	.4	18.2	73.2
	4 3 months	10	.1	3.3	76.5
	5 6 months	2	.0	.7	77.2
	6 Year	69	.5	22.8	100.0
	Total	302	2.2	100.0	
Missing	-1 Not asked	13667	97.8		
Total		13969	100.0		

On-line Data Dictionary

For Release 6, we have included a new addition to our HILDA documentation: the On-line Data Dictionary. As this is the first time the On-line Data Dictionary has been made available, we would welcome your feedback and suggestions.

The On-line Data Dictionary can be accessed via the HILDA website:

http://www.melbourneinstitute.com/hilda/onlinedd/Default.aspx

This on-line system is is designed to provide easy access to HILDA metadata. The database essentially provides the user with the information available in HILDA coding frameworks (.pdf).

The On-lin Data Dictionary allows users to search HILDA metadata four different ways:

- by keyword;
- by subject area;
- by variable name; and
- by derived variable name.

A help page (accessed by clicking on the help icon at the bottom right of the page) provides instructions on how to use the system along with example screen shots.

This system is still a work in progress, so you can expect it will be added to during 2008. Note that the questionnaire text is only currently available for wave 6 (the text for other waves will be added in due course).

DATA QUALITY ISSUES

Summary of Data Quality Issues

There are several technical and discussion papers that discuss the data quality issues that we are aware of in the datasets. These papers can be found on the HILDA website. A summary of these data quality issues is provided in Table 38. As further research is carried out on a variety of data quality issues, this table will be added to.

Table 38: Summary of the data quality issues in the HILDA data

Topic / variable	Problem	Where to get more information
Sample Representativeness		
Wave 1 non-response	The wave 1 response rate was 66% and non-respondents were more likely to be living in Sydney, male or unmarried, aged 20 to 24 or 65+, or born in non-English speaking country.	Watson and Wooden (2002a, pp. 3-8)
Attrition	The attrition rates from wave 2 are provided in Table 59. Attritors are more likely to be living in Sydney and Melbourne; aged 15 to 24 years; single or living in a de facto marriage; born in a non-English-speaking country; Aboriginal or Torres Strait Islander; living in a flat, unit or apartment; relatively low levels of education; unemployed; or working in blue-collar or low-skilled occupations.	Watson and Wooden (2006); Watson and Wooden (2004a, pp.2-14)
Missing data		
Item non-response		
General level of item non-response	Overall, the level of non-response in the HF, HQ and PQs is generally relatively low – less than 2 per cent. The item non-response rates in the SCQ are higher – averaging around 2.5 to 2.8 per cent	Watson and Wooden (2002a, p9); Watson and Wooden (2004a, p15)
Missing income data	10-16 per cent of respondents did not provide details for all financial year income components, resulting in 22 to 29 per cent of households with missing financial year income. Analysis of Wave 1 data shows that individuals with missing financial year information were more likely to be female; living in Sydney and rural WA; or attach a high importance to their financial situation. The income data is imputed.	Section above 'Income Variables and Income Imputation'; Section below 'Missing income data and the extent of income imputation'; Watson and Wooden (2002a, pp. 9-12)

Table 38: (c'td)

Topic / variable	Problem	Where to get more information
Missing wealth data	14 per cent of respondents did not provide allperson-level wealth details and 20 per cent of households did not provide all household-level wealth details, resulting in 39 per cent of households with missing wealth data (in wave 2) and 29 per cent in wave 6. The wealth data is imputed.	Section above 'Wealth Variables and Wealth Imputation'; Section below 'Missing wealth data and the extent of wealth imputation'; Watson and Wooden (2004a, pp. 21-24)
Family background	People living with both parents in wave 1 were not asked the family background questions on the assumption that this could be derived from the parent's interview. However, not all parents responded or it was impossible to determine what the parent was doing when the respondent was aged 14.	Watson and Wooden (2002a, pp. 12-13)
Permanently unable to work	452 respondents were incorrectly coded as 'permanently unable to work' at D21 in the PQ (interviewers were meant to check back to D6, but many used the response at D20 to code D21). As a result, the questions for those not in paid employment were not asked (such as whether looking for work, main activity, whether they would like work, and whether they have retired). Note that the retirement questions were asked in later waves.	Watson and Wooden (2002a, pp. 13-14)
Incomplete households		
Part-responding households	8 to 10 per cent of households are partially responding (that is, some but not all adults in the household provide an interview). When using derived variables that sum information across individuals in the household (for example, income or wealth variables), there will be more missing data.	Watson and Wooden (2002a, p14); Table 51, Table 53 to Table 57 below.
Accuracy of the data		
Questionnaire design issues		
Childcare costs	The child care grids in the HQ are very complex and require the parent to split the costs by the type of children (those of school aged and those not yet at school). There is some (small amount of) evidence that some respondents struggled to do this, with the same amount being reported for the two groups of children when the number of children in each group is not the same.	Watson and Wooden (2002a, p15)

Table 38: (c'td)

Topic / variable	Problem	Where to get more information
Current wages and salaries	There are some respondents who reported having current wages and salaries but who: • did not report having a job (13 respondents in wave 1). • were recorded as an employer (414 respondents in wave 1). There were also some respondents who did not report having current wages and salaries but who: • were recorded as an employee of their own business (126 respondents in wave 1). • were recorded as an employee (16 respondents in wave 1). There may be some circumstances that can explain these apparent discrepancies (for example, a spouses who have income from the family business but who do not actually work in the business).	Watson and Wooden (2002a, pp.15-16)
Calendar	In wave 1, we tried to separate jobs out based on whether they were full or part-time and asked the interviewers to record job numbers so we can identify jobs changing between part-time and full-time. However, this was not completed by the interviewers very often or was (mistakenly) not entered by the processing team. Interviewers also did not have sufficient instruction on how to treat breaks in employment (such as long-term leave or infrequent hours). The design of the calendar was modified between wave 1 and 2 due to some problems identified with the calendar in wave 1.	Watson and Wooden (2002a, p16)
Marital status	The HF and PQ in wave 1 asked whether respondents were 'legally' married with the intent of asking about a a 'registered' marriage. We suspect some defacto couples reported they were 'legally' married because they have certail legal rights under the Australian legal system. From wave 1, we have revised the questions to talk about 'registered' marriages. As a result, there may be inconsistencies between wave 1 and later waves.	Watson and Wooden (2002a, p16)

Table 38: (c'td)

Topic / variable	Problem	Where to get more information
Time use	While we undertake a large amount of checking and editing on the time use questions in the SCQ, it is likely that problems remain. The problem areas are: • Excessive hours reported suggest respondents find it difficult to think in terms of hours in a week. • The same hours may be recorded against multiple tasks if respondents are doing more than one thing at a time (eg looking after children while doing the housework). • Some confusion was caused by the layout of the boxes as some respondents tried to record both hours and minutes. The design of the time use question has undergone some revision since wave 1 to try to address these problems, but it is expected that errors still occur.	Watson and Wooden (2002a, p17)
Leave entitlements	In the wave 1 SCQ, respondents were asked about their access to paid and unpaid maternity leave in their current job. To avoid additional skips for men, a 'not applicable' option was provided. However, 1535 men provided answers to these questions, presumably answering whether other employees at their workplace had access to maternity leave. Also, older females selected 'not applicable' because they were not planning on using such leave. The questions were revised in wave 2.	Watson and Wooden (2002a, p17)
Moving house	In wave 2, we asked movers when they moved to their current address, but did not ask when they left their previous address. For people who move twice in a year, we do not know the exact length of tenure at the former address. The questionnaire was amended in wave 3.	Watson and Wooden (2004a, p30)
Duration of defacto relationship	In waves 2 and 3, we asked those completing the NPQ how long their most recent defacto relationship started and how long it lasted. This is inconsistent with wave 1, where we asked about the first such relationship and from wave 4 these questions have been reverted to the original ones.	Watson and Wooden (2004a, p30)
Data collection issues		
Sex	A small number of individuals had their sex corrected in the next wave (in wave 2, 37 people's sex was corrected). Note that the latest sex and date of birth is applied back through the earlier waves. This may lead to some subsequentially introduced inconsitencies in the question skips that rely on age or sex.	Watson and Wooden (2004a, pp.25-26)

Table 38: (c'td)

Topic / variable	Problem	Where to get more information
Date of birth	A relatively small number of corrections are applied to a person's date of birth in the next wave. (In wave 2, there were 50 people with a major change to their date of birth and 451 with a minor change. In later years, the number of changes were less and usually to replace dates of birth that were missing for new entrants to the household.) Note that the latest sex and date of birth is applied back through earlier waves.	Watson and Wooden (2004a, pp.25-26)
Working hours	In Wave 1, respondents were asked to compare their current hours with those a year ago. 26 cases reported hours a year ago that were inconsistent with their answer of whether they were more or less. The answer to the later was changed to reflect the former. Similarly, a small number of cases (in wave 1, there were 7) were inconsitent with their answer to whether they wanted to work more or less and the number of hours they wanted to work. Generally the answer to whether they wanted more or less hours was altered. For those with two jobs, some recorded more hours in all jobs that was less than their main job (in wave 1, there were 13). The hours in all jobs were usually set to -6 (unbelievable value). For those who work at home, some recorded more hours worked at home than in their main job (in wave 1, there were 33). Where this could not be resolved by looking at the hours worked in all jobs for multiple job holders, the hours worked at home were usually set to -6 (unbelievable value).	Watson and Wooden (2002a, p19)
Interviewer observations	Interviewes were required to complete observations of the dwelling and of the PQ interview. Unfortunately, not all interviewers completed this. For example, in wave 1, about about 0.1-0.4 per cent of cases had missing values.	Watson and Wooden (2002a, p20)
Mode effects and social desirability / acquiescence bias	Differences observed are quite small in absolute terms. Items tested: • difference between reported health in PQ and SCQ in wave 1; • whether responses tempered by presence of other adults during the interview.	Watson and Wooden (2002a, pp.21-22)

Table 38: (c'td)

Topic / variable	Problem	Where to get more information
Cross-form comparisons		
HF and PQ	Few questions are asked more than once. Proportion of cases where answers differed in wave 1 between HF and PQ: • 10% for long-term health condition; • 6.1% for labour force status; • 0.4% for marital status. Note HF and PQ may be done on different days and answered by different people. Also the questions were not identically worded.	Watson and Wooden (2002a, p22)
Cross-wave inconsistencies		
Marital status changes	Respondents are asked whether they changed their marital status since the last wave interviewed. Some report a different status but say there has been no change (for example, there were 258 respondents reporting no changing their marital staus since wave 1 but who had a different status). Most of these errors are recall errors but a small number may also be transcription errors by the interviewer.	Watson and Wooden (2004a, p27)
Address changes	Address changes can be identified through either a comparison of actual addresses recorded on the HF undertaken by Nielsen or via a question in the PQ. In wave 2, for example, 119 people indicated in their PQ that they had not changed address, but the address recorded was different and 141 people said they had moved, but the HF address was the same.	Watson and Wooden (2004a, p27)
Employment status changes	Respondents are asked to recall whether they were employed or not at the previous interview. In wave 2, for example, 4.6 of those employed in wave 1 did not recall being employed then and 6.8 per cent of those not employed in wave 1 recalled that they were. A very detailed analysis is given in Goode (2007). The majority of mistakes are made by those who change employment states between interviews. Variables significantly associated with making a mistake are being in full time education, the number of children, the time elapsed between interviews (possibly) and the number of jobs reported in the employment calendar.	Goode (2007); Watson and Wooden (2004a, p27-28)

Table 38: (c'td)

Topic / variable	Problem	Where to get more information
Calendar matching	There is a two to six month overlap (or seam) in the activity calendar collected each wave. Of those who had at least one job in the calendar seam, 19 per cent provided job spell information that was inconsistent between wave 1 and wave 2. 1.8 per cent of the jobs matched within 1 month, 0.7 matched within 3 months, 2.1 matched but with an error of more than 3 months and 14.8 per cent had at least one job that could not be matched.	
Comparison with external data		
General	Generally, the estimates are quite close for labour market, housing, demographic and health variables.	Watson and Wooden (2002a, pp.24-26)
Income	Compared to the ABS Survey of Income and Housing Costs, HILDA reports higher wages and salaries, and investment income.	Watson and Wooden (2004a, pp. 17-21) Note income estimates in Watson and Wooden (2002a, pp.24-26) are not imputed so not a fair comparison.
Wealth	Comparison with ABS and RBA suggest HILDA slightly understates the volume of financial assets, is much closer to the RBA than the ABS for non-financial assets, and is much lower (20 per cent) on debts than the ABS and RBA estimates.	Watson and Wooden (2004a, pp.22-24)
Height and weight	HILDA compares reasonably well with the ABS National Nutrition Survey but HILDA has a greater proportion of obese people but also lower item non-response.	Wooden et al. (2008)

Some more detailed information on the amount of missing income data and the extent of the income imputation is provided below.

Missing Income Data and the Extent of Income Imputation

The proportion of cases with missing income data are provided in Table 39. For most income variables, the proportion of missing income falls each wave. Part of the reason for the decline in the proportion of missing income may be because respondents are becoming more comfortable with the survey. For respondents with item non-response, the variables with the highest proportion of missing cases are still business income, investments and private transfers.

Table 40 shows how much of the mean income was imputed for each wave. For responding people with item non-response, 5.0 percent of total financial year income was imputed in wave 6, compared to 7.9 percent in wave 1. Including the imputed income totals for non-respondents within responding households (but excluding children), the percentage of total financial year income imputed for enumerated persons is 13.4 percent in wave 6.

This shows that while approximately one in ten responding persons are missing some component of financial year income in wave 6, only one twentieth of the mean income comes from imputed values and the remainder is from reported values. At the household level, one in five households is missing some component of financial year income and one seventh of the mean income is from imputed values.

Table 39: Proportion of cases with missing income data, waves 1 – 6

	Wave					
Variable	1	2	3	4	5	6
Responding Persons (non-zero cases only)						
Current income (per week)						
Wages and salaries (main job)	4.6	3.1	2.8	2.7	2.4	2.2
Wages and salaries (other jobs)	16.7	13.9	13.2	13.0	12.9	11.1
Benefits	3.2	2.0	2.0	1.8	1.6	1.0
Financial year income						
Wages and salaries	7.9	6.9	5.5	3.8	4.5	4.6
Aust govt pensions	2.1	2.1	1.3	2.0	1.4	1.0
Foreign govt pensions	0.5	2.7	0.0	0.5	2.4	0.5
Business income	29.1	28.6	27.4	19.4	21.7	18.6
Investments						
Interest income	19.5	18.6	13.9	11.0	11.3	12.8
Dividends and royalties	14.6	14.5	11.9	9.2	10.2	11.3
Rent income	20.3	14.7	14.9	11.3	10.5	10.3
Private pensions	6.3	4.7	3.3	4.1	4.9	3.9
Private transfers	8.0	23.1	15.8	14.4	21.2	13.4
Total FY income	15.7	14.9	12.1	9.6	10.7	10.3
Windfall income						
Windfall income	4.0	2.8	3.2	2.7	2.1	4.6
Enumerated Persons (zero and non-zero cases, excluding children)						
Current income (per week)						
Wages and salaries (main job)	10.0	8.6	7.9	8.3	7.3	7.0
Wages and salaries (other jobs)	8.4	7.6	7.0	7.5	6.6	6.3
Benefits	8.6	7.5	7.0	7.3	6.4	6.1
Financial year income						
Wages and salaries	12.1	10.9	9.6	9.0	8.7	8.6
Aust govt pensions	8.3	7.7	6.8	7.5	6.4	6.1
Foreign govt pensions	7.7	7.0	6.4	6.9	6.0	5.8
Business income	10.3	9.6	9.0	8.7	8.0	7.4
Investments						
Interest income	12.0	11.2	9.5	9.3	8.6	8.9
Dividends and Royalties	11.5	10.7	9.4	9.0	8.4	8.4
Rent income	9.2	8.3	7.8	7.8	6.9	6.8
Private pensions	8.0	7.3	6.6	7.1	6.3	6.1
Private transfers	7.9	7.6	7.0	7.3	6.8	6.2
Total FY income	21.3	20.1	17.2	15.3	15.5	15.1
Windfall income						
Windfall income	7.9	7.2	6.7	7.1	6.2	6.2

Table 39 (c'td)

	Wave					
Variable	1	2	3	4	5	6
Households (zero and non-zero cases)						
Current income (per week)						
Wages and salaries (main job)	14.2	12.3	11.2	12.2	10.9	10.4
Wages and salaries (other jobs)	11.9	10.8	10.0	10.8	10.0	9.3
Benefits	12.1	10.6	9.9	10.6	9.6	8.9
Financial year income						
Wages and salaries	17.0	15.7	13.8	13.0	12.8	12.8
Aust govt pensions	11.6	10.8	9.6	10.7	9.5	8.9
Foreign govt pensions	10.6	9.8	8.9	9.8	8.9	8.4
Business income	14.4	13.3	12.7	12.3	11.7	10.7
Investments	21.2	19.8	16.9	16.2	15.6	16.0
Private pensions	11.3	10.2	9.3	10.2	9.4	8.8
Private transfers	10.9	10.8	9.8	10.6	10.0	9.1
Total FY income	29.4	28.0	24.0	21.8	22.3	21.5
Windfall income						
Windfall income	10.9	10.0	9.3	10.2	9.1	9.1

Table 40: Mean financial year income ($) (including imputed values) and proportion of mean income imputed, waves 1 – 6 (weighted)

	Wave					
Variable	1	2	3	4	5	6
Responding persons						
Wages and salaries						
Mean	20,580	21,058	21,791	22,707	24,333	26,297
Proportion Imputed	5.9	4.6	3.7	2.7	3.2	3.5
Total income						
Mean	27,317	28,335	29,204	30,680	33,025	35,672
Proportion Imputed	7.9	6.9	5.9	4.5	5.0	5.0
Enumerated persons						
Wages and salaries						
Mean	21,079	21,689	22,307	23,055	24,805	26,850
Proportion Imputed	15.1	14.9	13.8	12.8	12.3	12.1
Total income						
Mean	27,836	28,958	29,803	31,072	33,497	36,276
Proportion Imputed	16.2	16.4	15.6	14.0	14.1	13.4
Households						
Wages and salaries						
Mean	42,368	43,471	44,776	46,404	49,874	53,934
Proportion Imputed	15.1	14.9	13.8	12.8	12.3	12.1
Total household income						
Mean	55,949	58,042	59,821	62,541	67,350	72,868
Proportion Imputed	16.2	16.4	15.6	14.0	14.1	13.4

Missing Wealth Data and the Extent of Wealth Imputation

The proportion of cases with missing wealth data are provided in Table 41. This table has two columns for each wave to highlight the proportion of respondents who answered the wealth question with a wealth band.[38] Wealth bands are strictly adhered to in the imputation of any wealth value (that is the imputed value must fall within the reported band) and greatly improve the quality of imputation. Treating cases where a wealth band is available as missing unfairly over represents the missingness problem so both situations have been provided. Missing cases for responding person and household level wealth items are reported as a proportion of non-zero cases only to more clearly show how much of a problem missing data is. However, not all missing cases required a non-zero impute (most cases do but for some it is unknown if they have the asset or debt and they can receive a zero impute) so the proportions give are a slight overestimation.

For most wealth variables, the proportion of missing income falls between wave 2 and wave 6. Part of the reason for the decline in the proportion of missing wealth may be because respondents are becoming more comfortable with the survey. In some situations where a wealth band option has been introduced, or an existing wealth band has been continued, there has been an increased proportion of missing values (when counting the wealth band as missing data). For respondents with item non-response, the variables with the highest proportion of missing cases are superannuation for retirees and those not retired. At the household level the largest amount of missingness is for trust funds, life insurance, business debt and business value. Each of the household level items are for situations where only a small amount of households actually have the asset or debt so the actual number of cases to be imputed is quite small.

Treating wealth band information as a response, nearly 39 percent of wave 2 households have some component of net worth missing. In wave 6 this proportion has dropped to 29 percent.

Table 42 shows the proportion of cases with missing home value which has generally declined over time.

Table 43 and Table 44 give the weighted mean wealth value (including imputed values) along with what proportion of the mean is attributed to imputed values. For all of the household wealth totals presented in Table 43, there has been a drop in the proportion imputed between wave 2 and wave 6. Home value (in Table 44) showed a small decrease in how much the mean was imputed between waves 2 to 5.

Comparing the table of missing values against the weighted means show that despite nearly 45 percent of households in wave 6 missing some component of net worth only 9.1 percent of the mean net worth wealth value was contributed by imputation.

[38] A wealth band is two values which the respondent believes their actual value to be within. The bands differ between some variables.

Table 41: Proportion of cases with missing wealth data including and excluding wealth band responses, waves 2 and 6

	Wave 2		*Wave 6*	
Variable	*inc. bands*	*excl. bands*	*inc. bands*	*excl. bands*
Responding persons (non-zero cases only)				
Joint bank accounts	9.8	-	6.0	-
Own bank accounts	4.6	-	3.3	-
Superannuation, retirees	20.1	-	19.7	12.2
Superannuation, not retired	17.3	10.7	27.5	13.6
HECS debt	10.6	-	7.6	-
Joint credit card debt	10.1	-	7.6	-
Own credit card debt	3.6	-	3.1	-
Other Debt	2.4	-	1.8	-
Enumerated persons (zero and non-zero cases)				
Joint bank accounts	11.3	-	8.4	-
Own bank accounts	9.8	-	7.9	-
Superannuation, retirees	8.0	-	7.0	6.5
Superannuation, not retired	17.1	13.3	23.0	14.4
HECS debt	7.8	-	6.4	-
Joint credit card debt	7.7	-	6.2	-
Own credit card debt	7.6	-	6.2	-
Other Debt	7.5	-	6.2	-
Household wealth items (non-zero cases only)				
Children's bank accounts	6.2	-	4.6	-
Business value	20.1	-	17.5	7.8
Cash investments	11.6	-	12.3	7.1
Equity investments	15.3	-	13.3	4.4
Collectibles	14.0	-	15.1	8.1
Other property value	4.6	-	0.5	-
Life insurance	24.9	-	28.5	16.9
Trust funds	35.7	-	35.8	26.7
Vehicles: Value	2.3	-	1.5	-
Business debt	22.9	-	11.6	8.1
Home Value	7.6	-	4.2	-
Home: All debt	5.4	-	4.2	-
Other property: Debt	7.1	-	5.9	-
Overdue bills: Debt	-	-	2.2	-
Household totals (zero and non-zero cases)				
Financial Assets	36.3	31.6	40.6	24.7
Non-Financial Assets	10.9	-	7.5	5.3
Total Assets	41.0	36.6	43.8	27.5
Financial Liabilities	15.1	-	12.3	12.2
Net Worth	43.0	38.9	44.9	29.4

Table 42: Proportion of households with missing home value data, waves 1 – 6

	Wave					
Variable	1	2	3	4	5	6
Home value (households) (non-zero cases only)						
Home value	5.9	7.6	5.6	4.0	3.3	4.2

Table 43: Mean wealth value ($) (including imputed values) and proportion of mean value imputed, waves 2 and 6 (weighted)

	Wave	
Variable	2	6
Households		
Financial assets		
Mean	152,081	219,158
Proportion Imputed	18.8	18.5
Non-financial assets		
Mean	315,398	507,143
Proportion Imputed	7.8	4.3
Total Assets		
Mean	467,473	726,301
Proportion Imputed	11.4	8.6
Total Liabilities		
Mean	65,485	113,661
Proportion Imputed	6.1	6.2
Net Worth		
Mean	401,980	612,640
Proportion Imputed	12.3	9.0

Table 44: Mean home value ($) (including imputed values) and proportion of mean value imputed, waves 1 – 6 (weighted)

	Wave					
Variable	1	2	3	4	5	6
Households						
Home Value						
Mean	200,168	232,206	271,135	298,830	315,108	345,298
Proportion Imputed	4.9	7.3	5.2	3.3	3.2	4.0

THE HILDA SAMPLE

Sample Design

Overview

In line with leading panel studies conducted in other countries, the sampling unit is the household, and members of those households will be traced over an indefinite life. The wave 1 sample is then automatically extended over time by following rules that add to the sample:

- 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.[39]

These following rules, in combination with the initial sample that is intended to be representative of all Australian households, provide a mechanism for ensuring that the panel retains its cross-sectional representativeness over time.

While all members of the selected households are defined as members of the sample, individual interviews are only conducted with those aged 15 years and over. Some limited information about people under 15, however, is collected from an appropriate adult member of the household.[40]

Reference Population

The reference population for wave 1 was all members of private dwellings in Australia, with the following exceptions:

- certain diplomatic personnel of overseas governments, customarily excluded from censuses and surveys;
- overseas residents in Australia (that is, persons who had stayed or intended to stay in Australia less than one year);
- members of non-Australian defence forces (and their dependents) stationed in Australia;
- residents of institutions (such as hospitals and other health care institutions, military and police installations, correctional and penal institutions, convents and monasteries) and other non-private dwellings (such as hotels and motels); and
- people living in remote and sparsely populated areas.

---

[39] See the section on Following Rules for more information about who is temporarily and permanently added to the sample.

[40] This approach is consistent with the British Household Panel Study (BHPS), with the difference that in the BHPS only people aged 16 years and over are interviewed. The lower age chosen for the HILDA Survey simply reflects our desire to conform to Australian Bureau of Statistics (ABS) standards adopted in its Labour Force Survey.

Further, to ensure that all members of the in-scope population have the same probability of selection, dwellings that were not primary places of residence (for example, holiday homes) were also excluded.

These coverage rules are broadly in line with those adopted by the Australian Bureau of Statistics (ABS) in the monthly Labour Force Survey supplements.[41] There are, however, two major differences. First, unlike the ABS, individuals at boarding schools, halls of residence and university colleges were included in the reference population for wave 1. Second, and again in contrast to ABS practice, military personnel who reside in private dwellings are part of the reference population for wave 1.

Note that while all members of the selected households are defined as members of the sample, individual interviews were only conducted with those aged 15 years and over.

Sampling Units

The sampling unit is the household defined, following the ABS, as 'a group of people who usually reside and eat together'[42]. The ABS clarifies how this definition is operationalised. Specifically, a household is either:

- a one-person household, that is, a person who makes provision for his or her own food or other essentials for living without combining with any other person to form part of a multi-person household; or

- a multi-person household, that is, a group of two or more persons, living within the same dwelling, who make common provision for food or other essentials for living. The persons in the group may pool their incomes and have a common budget to a greater or lesser extent; they may be related or unrelated persons, or a combination of both.

In general, persons who live in more than one household were only treated as members of the household where they spent most of their time. People who lived in another private dwelling for more than 50 per cent of the time were not treated as part of the household. Visitors to the household were also not treated as part of the household. Finally, people who usually lived in the household but were temporarily absent for work, school or other purposes were treated as part of the household, and this meant that a small proportion of interviews were conducted in locations other than at the household address.

Note again that we varied from the ABS practice in how we treat children attending boarding schools and halls of residence while studying. Specifically, while these dwellings are out of scope in wave 1, such individuals were treated as members of sampled households provided they spent at least part of the year in the sampled dwelling.

Sample Selection

The households were selected using a multi-staged approach. First, a sample of 488 Census Collection Districts (CDs) were selected from across Australia (each of which consists of approximately 200 to 250 households). Second, within each of these CDs, a sample of 22 to 34 dwellings was selected, depending on the expected response and occupancy rates of the area. The selections were made after all dwellings within each of

[41] ABS, Labour Statistics: Concepts, Sources and Methods (ABS Cat. No. 6102.0), ABS, Canberra, 2001.
[42] ABS, Statistical Concepts Library (ABS Cat. No. 1361.30.001), ABS, Canberra, 2000.

the CDs were fully listed. Finally, within each dwelling, up to three households were selected to be part of the sample.

Watson and Wooden (2002b) provides further details of the sampling methodology.

Following Rules

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 (CSMs) include all members of wave 1 households (including children). Any children born to or adopted by CSMs are also classified as CSMs. Further, all new entrants to a household who have a child with a CSM are converted to CSM status. CSMs remain in the sample indefinitely. All other people who share a household with a CSM in wave 2 or later are considered Temporary Sample Members (TSMs).

Where the household has moved, split or moved and split, the interviewers and office staff track the CSMs. The CSMs (along with their new household) are then interviewed, where applicable, at their new address or by phone.[43] TSMs that split from a household and are no longer part of a household with a CSM are not followed. However, if the TSM is converted to a CSM, then they are followed for interview as any CSM would be.

[43] Note that if a child CSM moves without any other adult CSMs, they are followed to their new household and the eligible members of that household are then interviewed.

DATA COLLECTION

The data collection task was subcontracted to Nielsen, a private market research company with a strong background in undertaking survey research in the area of social policy. The majority of the data were collected through face-to-face interviews.

Pilot Testing

The questionnaires are developed over the 9-month period prior to the main fieldwork for each wave. This pilot testing involves:

- Skirmish with a small number of participants (10-15 for waves 1-4 and 30 from wave 5) conducted in an office environment.
- Dress Rehearsal with a sample of urban and rural households (approximately 180 for waves 1-4 in NSW and 580 from wave 5 in NSW and Victoria).

In waves 1 and 2, a Pre-Test was also conducted in between the Skirmish and Dress Rehearsal with 30 Sydney households. From wave 3 onwards, the Pre-Test sample has been rolled into the Dress Rehearsal sample.

Questionnaire Length

Table 45 provides the average time taken to complete each of the questionnaires each wave.

Table 45: Average time (minutes) taken to complete questionnaires, waves 1 to 6

	Wave					
Questionnaire	*1*	*2*	*3*	*4*	*5*	*6*
Household Form (responding households)	4[a]	5[a]	5[a]	5[a]	6[a]	6[a]
Household Questionnaire	6.2	10.0	6.6	6.5	6.2	10.7
Person Questionnaire	34.4	-	-	-		
Continuing Person Questionnaire	-	30.5	30.1	28.1	31.7	31.3
New Person Questionnaire	-	36.2	34.2	37.7	37.5	37.1
Self-Completion Questionnaire	20[a]	20[a]	20[a]	20[a]	30[a]	30[a]

a. Approximate minutes as not timed

Interviewers

The number of interviewers used for each wave of the fieldwork is given in Table 46, together with the percentage of interviewers that were new to the HILDA Survey.

All interviewers and supervisors attended a two-day briefing session prior to each wave. From wave 2 onwards, the new interviewers received an extra day of training. All interviewers are provided with a manual covering the details of the questionnaires and fieldwork procedures.

Table 46: Number of interviewers and percentage of new interviewers each wave

	Wave					
	1	*2*	3	*4*	*5*	*6*
Number of interviewers used	139	142	128	135	133	141
Percentage of new interviewers	-	37%	25%	20%	20%	29%

Fieldwork Process

Data Collection Mode

The vast majority of the data were collected though face-to-face interviews. While telephone interviews and assisted interviews were conducted to ensure a high response rate, they are only used as a last resort. Table 47 provides the proportion of people interviewed by telephone in each wave. Due to the fact that some households moved outside of the 488 areas originally selected across Australia in wave 1 and the desire to interview as many people as possible, more telephone interviews are necessary in later waves. The overall incidence of telephone interviews has increased from 0.3 per cent in wave 1 to 6.6 per cent in wave 6.

Table 47: Proportion of respondents interviewed by telephone (%)

Sample Member Type	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Previous respondents	-	2.7	4.4	5.0	5.3	5.8
Previous non-respondents	-	7.7	10.6	17.1	21.0	27.0
Previous child, now turned 15	-	3.2	4.9	4.9	5.4	5.6
New entrants (TSM)	-	7.5	8.6	8.3	13.7	9.7
All wave respondents	0.3	3.0	4.8	5.6	6.5	6.6

Timeline

The interviews are conducted annually with the interviewer briefing occurring in mid-August each year. In wave 1, all but a few interviews were completed by December 2001. From wave 2 onwards, the fieldwork has been extended by several months into the following year to focus on tracking and interviewing hard-to-find cases.

Table 48 provides details of the fieldwork dates and Table 49 shows how the individual interviews are spread across each fieldwork period. For those interviewed in the next wave, most are interviewed within one month of the anniversary of the previous interview (as shown inTable 50). Up to 4 per cent of the interviews are, however, conducted more than three months before or after the anniversary of the interview in the previous wave.

Table 48: Fieldwork dates and proportion of interviews post December

	Fieldwork period		
Wave	Beginning of fieldwork	End of fieldwork	Percentage of fieldwork post December
1	24 August 2001	23 January 2002	0.4%
2	21 August 2002	19 March 2003	2.3%
3	21 August 2003	4 March 2004	1.8%
4	19 August 2004	28 February 2005	2.3%
5	24 August 2005	10 March 2006	4.0%
6	23 August 2006	9 March 2007	2.2%

Table 49: Distribution of individual interviews by month, waves 2 to 6

	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Total
Wave 1	147	5621	5153	1959	1036	53	-	-	13969
%	1.1	40.2	36.9	14.0	7.4	0.4	-	-	100.0
Wave 2	741	7272	3233	1379	116	1	276	23	13041
%	5.7	55.8	24.8	10.6	0.9	0.0	2.1	0.2	100.0
Wave 3	986	7373	2920	1058	157	27	206	1	12728
%	7.7	57.9	22.9	8.3	1.2	0.2	1.6	0.0	100.0
Wave 4	1540	7453	2229	761	139	24	262	-	12408
%	12.4	60.1	18.0	6.1	1.1	0.2	2.1	-	100.0
Wave 5	410	6803	3896	943	198	197	310	2	12759
%	3.2	53.3	30.5	7.4	1.6	1.5	2.4	0.0	100.0
Wave 6	527	7392	3611	907	181	141	142	4	12905
%	4.1	57.3	28.0	7.0	1.4	1.1	1.1	0.0	100.0

Table 50: Time between interview dates and anniversary of previous wave interview, waves 2 to 6

	Less than 1 month	1 to less than 2 months	2 to less than 3 months	3 to less than 4 months	4 to less than 5 months	5 to less than 6 months	Total ivwd in both waves
Wave 2	7985	2412	1242	217	106	31	11993
%	66.6	20.1	10.4	1.8	0.9	0.3	100.0
Wave 3	9046	1478	608	77	172	101	11482
%	78.8	12.9	5.3	0.7	1.5	0.9	100.0
Wave 4	9030	1363	512	103	153	115	11276
%	80.1	12.1	4.5	0.9	1.4	1.0	100.0
Wave 5	8741	1514	621	129	265	67	11337
%	77.1	13.4	5.5	1.1	2.3	0.6	100.0
Wave 6	9174	1503	577	165	243	46	11708
%	78.4	12.8	4.9	1.4	2.1	0.4	100.0

Survey Notification Material

In wave 1, the selected households were sent a primary approach letter and a brochure approximately one week prior to when the interviewer was scheduled to make contact with the household. This pre-interview material marketed the survey to respondents as a study about 'Living in Australia' and, among other things, emphasised that participation was voluntary and provided a means for sample members to opt out of the survey prior to an interviewer calling.

From wave 2 onwards, a primary approach letter and newsletter were sent to the last known address of the households approximately one month prior to when the interviewer was scheduled to make contact with the household. The newsletter provides respondents with some results from the previous wave. In addition to the posted pre-interview material, households with people who had not been part of the household in the previous wave were given a New Entrants Brochure. This brochure provided more information about the purpose of the study, why they had been asked to participate, and a method to opt out of the study if they chose to. A follow-up newsletter has also been introduced from wave 3 onwards.

A copy of the primary approach letters, brochures and newsletters are available from the HILDA website: http://www.melbourneinstitute.com/hilda/doc/doc_respinfo.htm.

Respondent Incentives

From wave 5, individuals responding with an interview received $25 and a $25 bonus was received by each fully responding household (i.e., each eligible member if the household provided an interview). This cash incentive structure is different to the one used in earlier waves.

Previously, in waves 1 to 4, a $50 cash incentive was offered to households where all eligible household members completed the Person Questionnaire. If this did not occur, a $20 payment was offered to households if at least one interview was obtained.

The availability of the incentive was made clear in both primary approach letter and the brochure/newsletter.

Call routine, Follow-Up and Refusal Aversion

In wave 1, the fieldwork was conducted in two stages. The first stage involved the interviewer working in an area over a three-week period. They visited each selected household according to the specified call-back pattern.[44] This achieved approximately 65 per cent of the interviews from each area. The remainder of each workload was then consolidated into intensive follow-up workloads and reassigned to the most experienced interviewers. They again visited each of these households according to the specified call-back pattern. These interviewers obtained the remaining 35 per cent of the interviews from each area.

From wave 2 onwards, a tracking component is incorporated into the fieldwork, splitting it into three distinct periods.[45] All households were issued into the field for the first period, and where all the interviews had not been completed, they were reissued into the field in the next period.[46] If a household could not be found at either one of these stages, they were put into tracking and once found were issued back into the current period if found quickly or more generally into a later period. The third period was used to finalise households that had to be traced and could not be immediately issued back into the field and also to contact some households where it was deemed beneficial to contact them in the third time (for example, a household member may have been away from the household at earlier contacts or they may have been temporarily unwell or busy).

Foreign Language Interviews

Language difficulties between the interviewer and the potential respondent were most often resolved by another member of the household acting as an interpreter. However, a small number of interviews each wave are conducted with a professional interpreter present during the interview.

Interviewer Monitoring

Several methods were used to ensure the fieldwork quality was consistent and maintained throughout the fieldwork collection period. These methods focused on the training, experience, in-field checking and monitoring of the interviewers.[47]

[44] Six or more calls were made to all selected households until a final household outcome was achieved. These calls were made over a minimum of a five-day period, with typically three calls on weekdays and at least three calls on weekends.

[45] For details on the tracking procedures adopted, see Watson and Wooden (2004b).

[46] When initially making contact with a household, the interviewer had up to six calls to make contact and a further six calls to undertake all of the interviews once contact had been made. If a household had to be put into tracking and was found, the initial call allocation to make contact with the household was carried over to the next period of the fieldwork. When following up a household, the interviewer had a total of five calls to finalise the household.

[47] See Watson and Wooden (2002b) for details of these monitoring methods.

Response Rates

A summary of the outcomes of the wave 1 fieldwork is provided in Table 51 and Table 52. Table 51 reveals that from the 11,693 households identified as in-scope, interviews were completed with all eligible members of 6872 households and with at least one eligible member of a further 810 households. The household response rate was, therefore, 66 per cent. Wooden, Freidin and Watson (2002) provide some comparisons of this response rate to other similar studies and conclude there are good reasons to be extremely satisfied with the rate of response obtained.

Table 51: Wave 1 Household Outcomes

Sample outcome	*Number*	*Percent*
Addresses issued	12,252	
Less out-of-scope (vacant, non-residential, foreign)	804	
Plus multi-households additional to sample	245	
Total households	*11,693*	*100.0*
Refusals to interviewer	2,670	22.8
Refusals to fieldwork company (via 1800 number or email)	431	3.7
Non-response with contact	469	4.0
Non-contact	441	3.8
Fully responding households	6,872	58.8
Partially responding households	810	6.9

The wave 1 person-level outcomes are provided in Table 52. Within the 7682 households interviewed, there were 19,914 people, resulting in an average of 2.6 persons per household. Of these people 4787 were under 15 years of age on the preceding 30 June and hence were ineligible for an interview in wave 1. This provided a sample of 15,127 eligible persons, 13,969 of whom completed the Person Questionnaire.

Table 52: Wave 1 Person Outcomes

Sample outcome	*Number*	*Percent*
Enumerated persons	19,914	
Ineligible children (under 15)	4,787	
Eligible adults	*15,127*	*100.0*
Refusals to interviewer	597	3.9
Refusals to fieldwork company (via 1800 number or email)	31	0.2
Non-response with contact	218	1.4
Non-contact	312	2.1
Responding individuals	13,969	92.3

Table 53 through to Table 57 show the household outcomes for waves 2 through 6. The household response rate (including fully and partially responding households) ranges from 87.0 per cent in wave 2 to 76.3 per cent in wave 6.

It is also constructive to consider the household outcomes for two groups – those that responded in the previous wave and those that didn't.[48] The household response rates for those households responding in the previous wave are:

- wave 2 – 87.0 per cent;
- wave 3 – 90.6 per cent;
- wave 4 – 92.3 per cent;
- wave 5 – 94.6 per cent; and
- wave 6 – 94.7 per cent.

In Table 58 and Table 59 we report a summary of the person-level response in waves 2 to 6. Of the 13,969 people interviewed in wave 1:

- 11,993 were re-interviewed in wave 2;
- 11,190 were re-interviewed in wave 3;
- 10,565 were re-interviewed in wave 4;
- 10,391 were re-interviewed in wave 5; and
- 10,085 were re-interviewed in wave 6.

The number interviewed in all six waves is 8864.

A common measure of the re-interviewing success is the attrition rate, calculated as the proportion of respondents in the previous wave that did not provide an interview in the current wave, excluding those that are out of scope (that is, those that have died or moved overseas). The wave-on-wave attrition rates, as shown in Table 59, are:

- wave 2 – 13.2 per cent;
- wave 3 – 9.6 per cent;
- wave 4 – 8.4 per cent;
- wave 5 – 5.6 per cent; and
- wave 6 – 5.2 per cent.

The attrition rates recorded in the early waves of the HILDA Survey are slightly higher than surveys such as the British Household Panel Study (BHPS), which achieved attrition rates in waves 2 and 3 of 12.4 percent and 7.8 percent respectively (after excluding proxy interviews). We believe they compare favorably given the comparative waves of the BHPS were conducted 10 years earlier and it has been generally accepted that response rates to surveys have been falling. Indeed, the wave 2 and 3 attrition rates for the recent BHPS Welsh sub-sample were 15.0 percent and 9.6 percent respectively and those for the recent BHPS Scottish sub-sample were 12.2 percent and 8.1 percent respectively (and these figures include proxy interviews which are not permitted in the HILDA Survey). The attrition

[48] Only responding households in wave 1 were issued in wave 2, so the closest comparison in the household response rate to be made in later waves is for households responding in the previous wave.

rate in the HILDA Survey is noticeably higher than the BHPS in both Waves 4 and 6, but did drop below the BHPS attrition rate in Wave 5 by 0.8 percentage points.

Table 53: Wave 2 Household Outcomes

Sample Outcome	*Number*	*Percent*
Households issued	7,682	
Plus split households	712	
Less out of scope households (due to death or move overseas)	68	
Total households	*8,326*	*100.0*
Refusals to interviewer	490	5.9
Refusals to fieldwork company (via 1800 number or email)	132	1.6
Non-response with contact	134	1.6
Non-contact, not lost to tracking	75	0.9
Lost to tracking	250	3.0
Fully responding households	6,541	78.6
Partially responding households	704	8.5

Table 54: Wave 3 Household Outcomes

Sample Outcome	*All households*		*Wave 2 responding HH*		*Wave 2 non-responding HH*	
	Number	*%*	*Number*	*%*	*Number*	*%*
Households from wave 2	8371		7245		1126	
Plus split households	463		394		69	
Less out of scope households (due to death or move overseas)	156		78		78	
Total households	*8678*	*100.0*	*7561*	*100.0*	*1117*	*100.0*
Refusals to interviewer	722	8.3	395	5.2	327	29.3
Refusals to fieldwork company (via 1800 number or email)	271	3.1	80	1.1	191	17.1
Non-response with contact	163	1.9	103	1.4	60	5.4
Non-contact, not lost to tracking	59	0.7	43	0.6	16	1.4
Lost to tracking	367	4.2	90	1.2	277	24.8
Fully responding households	6463	74.5	6289	83.2	174	15.6
Partially responding households	633	7.3	561	7.4	72	6.4

Table 55: Wave 4 Household Outcomes

Sample Outcome	*All households*		*Wave 3 responding HH*		*Wave 3 non-responding HH*	
	Number	*%*	*Number*	*%*	*Number*	*%*
Households from wave 3[a]	8738		7096		1642	
Plus split households	397		344		53	
Less out of scope households (due to death or move overseas)	247		136		111	
Total households	*8888*	*100.0*	*7304*	*100.0*	*1584*	*100.0*
Refusals to interviewer	871	9.8	312	4.3	559	35.3
Refusals to fieldwork company (via 1800 number or email)	339	3.8	52	0.7	287	18.1
Non-response with contact	199	2.2	117	1.6	82	5.2
Non-contact, not lost to tracking	93	1.0	43	0.6	50	3.2
Lost to tracking[b]	399	4.5	37	0.5	362	22.9
Fully responding households	6304	70.9	6122	83.8	182	11.5
Partially responding households	683	7.7	621	8.5	62	3.9

Notes: a Excludes 55 dead households and 41 empty households identified in Wave 2 and 3.
b Includes 321 households that were identified as untraceable in Wave 2 and 3.

Table 56: Wave 5 Household Outcomes

Sample Outcome	*All households*		*Wave 4 responding HH*		*Wave 4 non-responding HH*	
	Number	*%*	*Number*	*%*	*Number*	*%*
Households from wave 4[a]	9037		6987		2050	
Plus split households	388		329		59	
Less out of scope households (due to death or move overseas)	287		134		153	
Total households	*9138*	*100.0*	*7182*	*100.0*	*1956*	*100.0*
Refusals to interviewer	944	10.3	224	3.1	720	36.8
Refusals to fieldwork company (via 1800 number or email)	369	4.0	12	0.2	357	18.3
Non-response with contact	177	1.9	85	1.2	92	4.7
Non-contact, not lost to tracking	81	0.9	41	0.6	40	2.0
Lost to tracking[b]	442	4.8	31	0.4	411	21.0
Fully responding households	6495	71.1	6251	87.0	244	12.5
Partially responding households	630	6.9	538	7.5	92	4.7

Notes: a Excludes 95 dead households and 99 empty households identified in Wave 2 to 4.
b Includes 370 households that were identified as untraceable in an earlier wave.

Table 57: Wave 6 Household Outcomes

Sample Outcome	*All households*		*Wave 5 responding HH*		*Wave 5 non-responding HH*	
	Number	*%*	*Number*	*%*	*Number*	*%*
Households from wave 5[a]	9297		7125		2172	
Plus split households	397		373		24	
Less out of scope households (due to death or move overseas)	343		155		188	
Total households	*9351*	*100.0*	*7343*	*100.0*	*2008*	*100.0*
Refusals to interviewer	1097	11.7	240	3.3	857	42.7
Refusals to fieldwork company (via 1800 number or email)	399	4.3	11	0.1	388	19.3
Non-response with contact	190	2.0	82	1.1	108	5.4
Non-contact, not lost to tracking	51	0.5	32	0.4	19	0.9
Lost to tracking[b]	475	5.1	25	0.3	450	22.4
Fully responding households	6538	69.9	6396	87.1	142	7.1
Partially responding households	601	6.4	557	7.6	44	2.2

Notes: a Excludes 146 dead households and 173 empty households identified in Wave 2 to 5.
b Includes 407 households that were identified as untraceable in an earlier wave.

Table 58: Wave 2 to 6 Person Outcomes Against Wave 1 Person Outcomes

	Wave 1			*New Entrants each wave*					
	Resp.	*Non-resp*	*Child*	2	3	4	5	*6*	*Total*
Wave 2									
Respondent	11993	222	250	576	-	-	-	-	13041
Non-respondent	1824	904	61	210	-	-	-	-	2999
Out-of-scope	152	32	19	-	-	-	-	-	203
Child	-	-	4457	345	-	-	-	-	4802
Wave 3									
Respondent	11190	223	462	356	497	-	-	-	12728
Non-respondent	2464	886	165	154	156	-	-	-	3825
Out-of-scope	315	49	37	334	-	-	-	-	735
Child	-	-	4123	287	364	-	-	-	4774
Wave 4									
Respondent	10565	209	666	284	287	397	-	-	12408
Non-respondent	2898	879	289	138	113	167	-	-	4484
Out-of-scope	506	70	63	445	322	-	-	-	1406
Child	-	-	3769	264	295	332	-	-	4660

Table 58: (c'td)

	Wave 1			*New Entrants each wave*					
	Resp.	*Non-resp*	*Child*	2	3	4	5	*6*	*Total*
Wave 5									
Respondent	10392	238	909	261	230	247	482	-	12759
Non-respondent	2958	844	354	97	77	81	108	-	4519
Out-of-scope	619	76	67	514	424	279	-	-	1979
Child	-	-	3457	259	286	289	356	-	4647
Wave 6									
Respondent	10085	245	1146	233	191	200	311	494	12905
Non-respondent	3144	833	477	96	76	61	81	115	4883
Out-of-scope	740	80	67	553	483	364	256	-	2543
Child	-	-	3097	249	267	271	298	343	4525
Total	13969	1158	4787	1131	1017	896	946	952	24856

Table 59: Individual response rates (%) for the HILDA Survey, waves 2 to 6 compared

	All people					*People in responding h'hold in previous wave*				
	W2	*W3*	*W4*	*W5*	*W6*	*W2*	*W3*	*W4*	*W5*	*W6*
Previous wave respondent	86.8	90.4	91.6	94.4	94.8	86.8	90.4	91.6	94.4	94.8
Previous wave non-respondent	19.7	17.6	12.7	14.7	8.4	19.7	19.8	18.1	25.3	18.2
Previous wave child	80.4	71.3	70.7	74.6	75.4	80.4	81.8	81.2	87.3	89.5
New entrant this wave	73.3	76.1	70.4	81.7	81.1	73.3	78.5	71.8	85.4	81.0

Table 60 shows the response rates for the Self Completion questionnaire, calculated as the percentage of respondents to which an SCQ could be matched.

Table 60: Self Completion Questionnaire response rate by wave

	Person Questionnaire	*Matched Self Completion Questionnaire*	*% Matched*
Wave 1	13969	13055	93.5
Wave 2	13041	11638	89.2
Wave 3	12728	11747	92.3
Wave 4	12408	11393	91.8
Wave 5	12759	11460	90.0
Wave 6	12905	11709	90.7

Attrition Bias

Attrition is generally only a serious concern when it is non-random (that is, when the persons that attrit from the panel have characteristics that are systematically different from those who remain).

Table 61 provides figures on the proportion of wave 1 respondents who were re-interviewed in wave 6 disaggregated by various sample characteristics. For those persons interested in the balanced panel, the proportion of wave 1 respondents who have been interviewed in every wave is also provided. People who have died or moved overseas are excluded from these figures. These results indicate that the re-interview rate is lowest among people with following wave 1 characteristics:

- living in Sydney and Melbourne;
- aged 15 to 24 years;
- single or living in a de facto marriage;
- born in a non-English-speaking country;
- Aboriginal Torres Strait Islander;
- living in a flat, unit or apartment;
- relatively low levels of education;
- unemployed; or
- working in blue-collar or low-skilled occupations.

The variance in attrition over the six waves is particularly marked with respect to age, country of birth, labour force and occupation.

The disparity in the re-interview rates for wave 1 respondents re-interviewed in wave 6 across the different characteristics is not as great as for those interviewed every wave. The most striking example of this is indigenous status – while only 51.8 per cent of indigenous respondents in wave 1 have been re-interviewed every wave, 70.8 per cent were re-interviewed in wave 6. This indicates that the groups with low re-interview rates in the balanced panel are still engaged with the study.

Overall, attrition is not random. While we can make adjustments for the attrition through the sample weights, these adjustments are only as good as our ability to measure differential attrition.

The attrition rates are discussed at length in Watson and Wooden (2004c and 2006).

Table 61: Percentage of Wave 1 Respondents Re-interviewed by Selected Sample Characteristics

Wave 1 characteristics	*In all waves (%)*	*In wave 6 (%)*	*Wave 1 characteristics*	*In all waves (%)*	*In wave 6 (%)*
Area			Indigenous status		
Sydney	62.9	73.4	Indigenous	51.8	70.8
Rest of NSW	70.7	78.8	Non-indigenous	67.9	76.3
Melbourne	65.7	75.8	Education attainment		
Rest of Victoria	64.9	73.2	Year 11 or below	64.1	73.1
Brisbane	72.1	78.8	Year 12	63.7	74.4
Rest of Qld	69.4	77.2	Certificate	67.3	75.4
Adelaide	71.3	77.8	Diploma	72.8	80.7
Rest of SA	66.2	76.8	Degree or higher	77.3	83.9
Perth	68.2	74.4	Dwelling type		
Rest of WA	65.7	75.7	House	68.0	76.6
Tasmania	71.3	79.4	Semi-detached	69.0	77.0
Northern Territory	81.5	89.4	Flat, unit, apartment	63.7	72.7
ACT	69.3	76.2	Other	61.7	70.0
Sex			Labour force status		
Male	66.0	74.7	Employed full-time	66.8	75.7
Female	69.1	77.6	Employed part-time	69.8	78.9
Age group (years)			Unemployed	55.8	68.4
15–19	54.2	68.6	Not in labour force	69.0	76.3
20–24	54.9	68.8	Employment status in main job*		
25–34	65.2	75.6	Employee	67.8	76.7
35–44	69.8	77.4	Employer	66.8	75.4
45–54	69.9	77.0	Own account worker	69.9	77.7
55–64	74.3	80.6	Contributing family worker	68.2	83.6
65–74	78.8	84.0	Occupation*		
75+	65.4	71.5	Managers / administrators	70.4	79.8
Marital status			Professionals	76.2	84.1
Married	70.6	78.0	Associate professionals	67.0	76.0
De facto	63.9	75.0	Tradespersons	62.6	72.1
Separated	67.5	75.4	Advanced clerical / service	66.3	76.1
Divorced	74.8	83.0	Intermediate clerical / sales / service	68.1	76.5
Widowed	75.8	79.4			
Single	58.8	70.6	Intermediate production / transport	61.5	69.0
Country of birth			Elementary clerical / sales / service	65.0	76.2
Australia	69.6	78.0	Labourers	60.8	71.5
Overseas					
Main English-speaking	69.3	76.5	Total	67.6	76.2
Other	56.0	66.5	Number responding	8864	10085

Note: * Employed sub-sample only

Data Processing

Data Entry

The data from the Household Form, Household Questionnaire and Person Questionnaires were manually entered into a database. The keyed numerical data was double-entered to minimize data entry errors. The keyed verbatim responses were only entered once as these were only used for coding purposes and any mis-entered data could be easily identified and corrected. During data entry, the data was checked using range, logical and consistency edits. Where necessary the data entry was suspended until the identified problem was resolved.

The data from the Self-Completion Questionnaire were scanned into a database using a mark-sense recognition system. A sample of forms, together with forms that had a high proportion of missing responses, were visually inspected to ensure the scanning process was working correctly. All inappropriately marked multiple responses were also visually inspected, and where no single response could be clearly differentiated from the scanned image, the following rules (based on the SF-36 rules for coding problems in Ware et al, 2000) were applied:

- If a respondent marked two responses that were adjacent to each other that were part of a scale, one was randomly chosen.
- If a respondent marked two responses for an item and they were not adjacent to each other, the item was coded as -5 “invalid multiple response”.
- If a respondent marked three or more responses for an item, the item was coded as -5.

Once the data from all forms were entered, consistency in the data recorded on the various forms was checked. Any discrepancies were then investigated and resolved.

Coding Responses

The coding of the occupation and industry questions was done in the office prior to data entry. Occupation questions are coded to four-digit Australian Standard Classification of Occupation 1997 (ASCO) and industry questions are coded to four-digit Australian and New Zealand Standard Industry Classification 1993 (ANZSIC).[49]

The questionnaires also contained items for which a partial list was provided with an ‘other, please specify’ category and some of these lists needed to be further extended. These codeframe extensions occurred once a sufficient number of forms had been entered so that common responses could be identified. Some backcoding into the original codeframe in the questionnaire was also done where the interviewer had written down an ‘other’ response that actually fell into the codeframe already provided in the questionnaire.

The accuracy of the coding was monitored on a continual basis. Senior coders verified 10 per cent of the occupation and industry codes. Any discrepancies identified were discussed between the coders and corrected, thus identifying problematic areas for further

[49] For Release 7, we plan to recode the occupation and industry data for waves 1 to 6 to the 2006 version of the Australian and New Zealand Standard Classification of Occupations and the 2006 version of ANZSIC. From Wave 7, only these new codeframes will be used.

investigation and providing feedback to the coders. The coding of the ‘other, please specify’ responses were checked through a blind re-coding of all cases. Wherever problems were identified, these were investigated and resolved.

HILDA USER TRAINING

We aim to conduct a HILDA User Training sessions at least every two years. (The last HILDA User Training was held on 18 July 2007 prior to the HILDA Survey Research Conference which occurred on 19-20 July 2007.)

At this stage, we do not have any plans to run a HILDA User Training session in 2008. If your organisation has a sufficiently large number of users (at least 15) who would like training, you can request a special training session for those in your organisation. A registration fee will apply. Please contact Nicole Watson to discuss your request (n.watson@unimelb.edu.au).

Should we undertake a general User Training in 2008, details will be circulated to the HILDA email list and will be posted to the HILDA website. (To subscribe to the HILDA email list, go to www.ecom.unimelb.edu.au/iaesrwww/hilda/mail/hilda-l.html.)

GETTING MORE INFORMATION

No doubt there will be questions this manual does not answer. There are a number of other ways to get more information about the HILDA Survey data:

- Go to the HILDA website – copies of all survey instruments and various discussion and technical papers can be viewed and downloaded. You will also find the order forms for the datasets along with a growing bibliography of research papers that use the HILDA Survey data.

- Contact the HILDA team at the Melbourne Institute – if you have lost your password, or you have questions about the data files or variables email hilda-inquiries@unimelb.edu.au.

- Contact the HILDA team at FaHCSIA – if you have a query about getting access to the data, ensure you have read the details on the HILDA website about accessing the data (http://www.melbourneinstitute.com/hilda/data.html), and if your questions are not answered there, then email HILDA@fahcsia.gov.au.

- Circulate a message to the HILDA email list – all users of the data are automatically subscribed to the HILDA email list when you apply for the data. You will receive an email confirmation that you have been subscribed. If your question could be answered by the broader HILDA user community, please feel free to send your question to this group (email: hilda-l@unimelb.edu.au).

REFERENCES

ABS (1994), *Australian and New Zealand standard industrial classification: alphabetic coding index.* Australian Bureau of Statistics, Canberra; Department of Statistics, Wellington. Cat No 1293.0.

ABS (1997), *Australian Standard Classification of Occupations,* Australian Bureau of Statistics, Canberra. Cat No 1220.0

Brazier, J.E., Roberts, J.R. and Deverill, M, (2002), 'The estimation of a preference-based measure of health from the SF-36', *Journal of Health Economics,* vol. 21, pp. 271-292.

Cummins, R.A. (1996), 'The domains of life satisfaction: An attempt to order chaos', *Social Indicators Research*, vol. 38, no. 3, pp. 303-332.

Galinsky, E. (1999), *Ask the Children*: *What America's Children Really think about Working Parents*, William Morrow & Co, New York.

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

Hayes, C (2008), 'HILDA Standard Errors: User Guide', HILDA Project Technical Paper Series No. 2/08, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Henderson, S., Duncan-Jones, P., McAuley, H. and Ritchie, K. (1978), 'The patient's primary group'*, British Journal of Psychiatry*, vol. 132, pp. 74-86.

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.

Little, R.J.A., and Su, H.L. (1989) 'Item Non-Response in Panel Surveys' in Panel Surveys, edited by Kasprzyk, D., Duncan, G.J., Kalton, G., Singh, M.P., John Wiley and Sons, New York.

Manski, C. and Straub, J. (2000), 'Worker Perceptions of Job Insecurity in the Mid-1990s', *Journal of Human Resources*, vol. 35, no. 3, pp. 447-479.

Marshall, M.L., and Barnett, R.C. (1993), 'Work family strains and gains among two-earner couples', *Journal of Community Psychology*, vol. 21, pp. 64-78.

Saucier, G. (1994), 'Mini-Markers: A Brief Version of Goldberg's Unipolar Big-Five Markers', *Journal of Personality Assessment*, vol. 63, no. 3, pp. 506-516.

Starick, R. and Watson, N. (2007), 'Evaluation of Alternative Income Imputation Methods for the HILDA Survey', HILDA Project Technical Paper Series No. 1/07, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

VandenHeuvel, A. and Wooden, M. (1995), 'Self-employed contractors in Australia: How many and who are they?', *The Journal of Industrial Relations* , vol. 37, no. 2, pp. 263–280.

Ware, J.E., Snow, K.K., Kosinski, M. and Gandek, B. (2000), *SF-36 Health Survey: Manual and Interpretation Guide*, QualityMetric Inc., Lincoln, RI.

Watson, N (2004a), 'Income and Wealth Imputation for Waves 1 and 2', HILDA Project Technical Paper Series No. 3/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N (2004b), 'Wave 2 Weighting', HILDA Project Technical Paper Series No. 4/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N, and Fry, T (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 (2002a), 'Assessing the Quality of the HILDA Survey Wave 1 Data', HILDA Project Technical Paper Series No. 4/02, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N, and Wooden, M (2002b), '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, and Wooden, M (2004a), 'Assessing the Qulity 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 Survey Methodology', HILDA Project Technical Paper Series No. 1/04, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N, and Wooden, M (2004c), 'Sample Attrition in the HILDA Survey', *Australian Journal of Labour Economics*, vol. 7, June, pp. 293-308.

Watson, N, and Wooden, M (2006) 'Modelling Longitudinal Survey Response: The Experience of the HILDA Survey', HILDA Project Discussion Paper Series No. 2/06, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Wooden, M, Freidin, S, and Watson, N (2002), 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1', *The Australian Economic Review*, vol. 35, no. 3, pp. 339-48.

Wooden, M, Watson, N, Aguis, P, and Freidin, S (2008), 'Assessing the Quality of the Height and Weight Data in the HILDA Survey', HILDA Project Technical Paper Series No. 1/08, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

APPENDIX 1a: SUMMARY OF HILDA SURVEY CONTENT, WAVES 1 – 6

The following tables provide a guide to topics covered in the HILDA Survey across the first six waves. If you are interested in which specific variables are available each waves, you should refer to the cross-wave index provided with the documentation on Release 6 of the HILDA DVD.

HOUSEHOLD FORM	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Sex [a]	X	X	X	X	X	X
Date of birth [a]	X	X	X	X	X	X
Fraction of time spent living at address	X					
English language ability of household members	X					
Disabilities of household members	X	X	X	X	X	X
Marital status of household members	X					
Employment status of household members	X		X	X	X	X
Household relationships	X	X	X	X	X	X
Entrants – reasons for, and date of, joining household		X	X	X	X	X
Movers – reasons for, and date of, leaving household		X	X	X	X	X

a Pre-printed from Wave 2 onwards.

HOUSEHOLD QUESTIONNAIRE	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Child Care						
Difficulties with child care (12 items [b])	X	X	X	X	X	X
Care during school term time – hours and cost by type	X [c]	X	X	X	X	X
Care during school holidays – hours and cost by type	X [c]	X	X	X	X	X
Care for children not yet at school while working – hours and cost by type	X [c]	X	X	X	X	X
Care while not working – hours and cost by type [d]		X	X	X	X	X
Receipt of Child Care Benefit	X	X	X	X	X	X
Receipt of Family Tax Benefit			X	X	X	X
Housing						
No. of bedrooms	X	X	X	X	X	X
Ownership status	X	X	X	X	X	X
Landlord type	X	X	X	X	X	X
Rent payments	X	X	X	X	X	X
Boarders	X	X	X	X	X	X
How housing provided if live rent free	X	X	X	X	X	X
Notional rent (if live rent free)	X	X	X	X	X	X

HOUSEHOLD QUESTIONNAIRE (c'td)	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Dwelling type [e]	X	X	X	X	X	X
Condition of dwelling (interviewer assessed) [e]	X	X	X	X	X	
Housing Wealth						
Owner IDs and share owned		X				X
First home buyer	X	X				
Year home purchased		X				X
Purchase price of home		X				X
Current value of home	X	X	X	X	X	X
Value of housing debt	X	X	X	X	X	X
Value of housing loans repayments	X	X	X	X	X	X
Year expect housing loan to be paid off	X	X	X	X	X	X
Value of initial housing loans		X				X
Value of housing loan when last refinanced						X
Other Household Assets *[special module]*						
Value of other properties		X				X
Value of equity investments		X				X
Value of trust funds		X				X
Value of children's bank accounts		X				X
Value of other cash-type investments		X				X
Value of business assets		X				X
Value of vehicles		X				X
Value of life insurance		X				X
Value of collectibles		X				X
Other Household Debts *[special module]*						
Value of business debt		X				X
Overdue household bills						X
Other						
Number of motor vehicles	X					
Weekly expenditure on groceries / food [f]	X		X	X	X	
Weekly expenditure on meals out [f]	X		X	X	X	
Adequacy of household income	X					
Total household income (bands)				X	X	X

b One item added from Wave 3 onwards.
c In Wave 1 all child care items related to employment-related child care, with questions restricted to households where all carers in household were employed.
d Questions split by school-aged children and children not yet at school from Wave 5 onwards.
e Collected as part of the HF in Wave 1.
f These expenditure items are collected in the SCQ from Wave 6 onwards.

CONTINUING PERSON QUESTIONNAIRE	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Country of birth & language						
Country of birth	X			X		
Year of arrival	X			X		
English as first language	X					
Aboriginality	X					
Australian citizenship				X		
Permanent residence				X		
NZ citizen prior to arrival				X		
Refugee				X		
Visa category [recent arrivals only]				X		
Family background						
Lived with parents at age 14	X					
Why not living with parents	X					
Parents ever separated / divorced	X					
Age at time of separation	X					
Age left home	X					
Siblings	X					
Whether eldest sibling or not	X					
Father's / mother's country of birth	X					
Father's / mother's occupation	X					
Father's unemployment experience	X					
Father's education					X	
Mother's education					X	
Education						
Study status	X	X	X	X	X	X
Year left school	X	X	X	X	X	X
Type of school last attended	X	X	X	X	X	X
Qualifications studying for		X	X	X	X	X
Qualifications completed	X	X	X	X	X	X
Date completed qualification		X	X	X	X	X
Country in which completed qualification	X	X	X	X	X	X
Employment history & status						
Years since left FT education	X					
Years in paid work	X					
Years unemployed	X					
Years out of labour force	X					
Main activity when not in labour force	X					

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Employment status – ABS definition (9 questions)	X [g]	X	X	X	X	X
Current employment						
Usual weekly hours of work – all jobs	X	X	X	X	X	X
Preferred weekly hours of work – all jobs	X	X	X	X	X	X
Reasons for working part-time hours	X	X	X	X	X	X
Multiple job holding	X	X	X	X	X	X
Usual weekly hours of work – main job	X	X	X	X	X	X
Days of the week worked [h]	X	X	X	X	X	X
Shiftwork	X	X	X	X	X	X
Occupation	X	X	X	X	X	X
Occupation change		X	X	X	X	X
Occupation experience	X	X	X	X	X	X
Job tenure	X	X	X	X	X	X
Industry	X	X	X	X	X	X
Working at home (3 questions)	X	X	X	X	X	X
Trade union membership	X	X	X	X	X	X
Paid holiday leave	X	X	X	X	X	X
Paid sick leave	X	X	X	X	X	X
Employment contract type	X	X	X	X	X	X
Expectation of contract renewal	X					
Labour hire	X	X	X	X	X	X
Expected quit probability	X	X	X	X	X	X
Expected dismissal probability	X	X	X	X	X	X
Expected probability of finding another job	X	X	X	X	X	X
Work-related training (3 questions)			X	X	X	X
PAYE tax status	X	X	X	X	X	X
Supervisory responsibilities	X	X	X	X	X	X
Employer type	X	X	X	X	X	X
Workplace size	X	X	X	X	X	X
Firm size	X	X	X	X	X	X
Job satisfaction (6 items)	X	X	X	X	X	X
Job search while employed	X	X	X	X	X	X
Intended age of retirement	X		X			X
Reason for ceasing last job		X	X	X	X	X
Characteristics of a previous job (5 items)		X	X	X	X	X

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Persons not in paid employment						
Job search activity	X	X	X	X	X	X
Looking for work – When began looking for work	X	X	X	X	X	X
Looking for work – Hours spent in job search in last week		X	X	X	X	X
Looking for work – Intensive Assistance	X	X	X	X	X	X
Looking for work – Availability to start work	X	X	X	X	X	X
Looking for work – Difficulties finding a job	X	X	X	X	X	X
Looking for work – Number of job offers	X	X	X	X	X	X
Not looking for work – Main activity	X	X	X	X	X	X
Not looking for work – Preference to work	X	X	X	X	X	X
Not looking for work – Reasons for not looking	X	X	X	X	X	X
Not looking for work – Availability to start work	X	X	X	X	X	X
Reservation wage	X	X	X	X	X	X
Desired hours of work	X	X	X	X	X	X
Expected probability of finding a job	X	X	X	X	X	X
Reason for ceasing last job	X	X	X	X	X	X
Characteristics of a previous job (5 items)	X	X	X	X	X	X
Work-related training (3 questions)						X
Whether retired	X	X	X	X	X	X
Year / age retired	X	X	X	X	X	X
Age plan to retire	X	X	X	X	X	X
Labour market calendar	X	X	X	X	X	X
Leave taking					X	X
Mutual Obligation activity	X	X	X	X	X	X
Income						
Current wage and salary income	X	X	X	X	X	X
Current income from government benefits	X	X	X	X	X	X
Financial year income by source	X	X	X	X	X	X
Credit card use and payment strategy	X	X	X	X	X	X
Wealth *[special module]*						
Bank accounts		X				X
Credit card debt		X				X
Other debts		X				X
Superannuation		X				X
Home and property ownership history						X
Unpaid personal bills						X

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Family formation						
Number of children	X	X	X	X	X	X
Non-resident children characteristics	X	X	X	X	X	X
Financial support for non-resident children	X	X	X	X	X	X
Amount of contact with youngest non-resident child	X	X	X	X	X	X
Employment status of other parent			X	X	X	X
Resident children characteristics	X	X	X	X	X	X
Financial support from other parent	X	X	X	X	X	X
Amount of contact other parent has with youngest child	X	X	X	X	X	X
Employment status of other parent			X	X	X	X
Desire to have another child	X	X	X	X	X	X
Likelihood of having another child	X	X	X	X	X	X
Number of additional children intend to have	X	X	X	X	X	X
Year intend to have next child		X			X	
Fertility *[special module]*						
Partner/self currently pregnant					X	
Time stopped/started work pre/post birth of baby					X	
Use of birth control					X	
Partnering / relationships						
Changes in marital status		X	X	X	X	X
Current marital status	X	X	X	X	X	X
Current living circumstances		X	X	X	X	X
Single persons – Likelihood of marriage	X	X	X	X	X	X
De facto relationships – Year relationship started	X	X	X	X	X	X
Number of other de facto relationships		X	X	X	X	X
Retirement *[special module]*			X			
Health / disability						
Whether has disability / health condition	X	X	X	X	X	X
Type of disability			X	X	X	X
Whether disability commenced in last year	X	X		X	X	X
Year of onset of disability			X			
Impact of disability on work (2 questions)	X	X	X	X	X	X
Difficulties as a result of disability (3 questions)				X		
Need for help / supervision (4 questions)				X		
Use of aids				X		
Home modifications				X		

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Employment difficulties				X		
Education difficulties				X		
DVA Treatment Entitlements Card				X		
Private health insurance (8 questions)				X		
Hospital visits in past 12 months (7 questions)				X		
Caring for others					X	X
Whether respondent is a carer in household					X	X
Whether respondent is a carer outside household					X	X
Youth *[special module]*				X		
Other						
Life satisfaction (9 items)	X	X	X	X	X	X
Importance of life domains (8 items)	X					
Attitudes to life in Australia (3 items)	X					
English language speaking (2 questions)	X	X	X	X	X	X
Movers – Date moved to current address		X	X	X	X	X
Movers – Date left previous address			X	X	X	X
Movers – Reasons for moving		X	X	X	X	X
Intentions / plans for next 3 years						
Move house					X	
Move to where					X	
Stop/start working					X	
Change Employment					X	

g In Wave 1 a shorter series of questions was used.
h From Wave 4 onwards, an additional sub-question was included to better enable weekend workers to be identified.

SELF-COMPLETION QUESTIONNAIRE	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Health / Lifestyle / Living Situation						
Health and well-being – SF36 (36 items)	X	X	X	X	X	X
Serious health conditions (8 items)			X			
Exercise (1 item)	X	X	X	X	X	X
Smoking incidence	X	X	X	X	X	X
Smoking frequency		X	X	X	X	X
Tobacco expenditure	X					
Alcohol consumption (2 items)	X	X	X	X	X	X
Height / weight						X
Time stress (2 items)	X	X	X	X	X	X
Preferences to live in area	X	X	X	X		X
Neighbourhood characteristics (10 items)	X	X	X	X		X[i]
Housing adequacy (6 items)	X	X				
Satisfaction with family life (8 items)	X	X	X	X	X	X
Satisfaction with h'h div of labour (2 items)					X	X
Fairness of housework	X	X	X	X	X	X
Marital relationship quality (6 items)			X			X
Membership of clubs etc	X	X	X	X	X	X[j]
Social interaction with friends etc	X	X	X	X	X	X
Community participation (12 items)						X
Social support (10 items)	X	X	X	X	X	X
Trust (2 or 7 items)					X	X[k]
Self-efficacy (7 items)			X	X		
Religion (3 questions)				X		
Life events in past 12 months		X	X	X	X	X
Time use	X	X	X	X	X	X
Responsibility for h'h tasks (6 items)					X	
Use of domestic help (2 items)					X	
Finances						
Self-assessed prosperity	X	X	X	X	X	X
Stressful financial events (7 items)	X	X	X	X	X	X
Response to financial emergency (2 items)	X	X	X	X	X	X
Savings habits	X	X	X	X		X
Savings time horizon	X	X	X	X		X
Reasons for saving		X				X
Risk preference	X	X	X	X		X

SELF-COMPLETION QUESTIONNAIRE (c'td)	*Wave*					
	1	*2*	*3*	*4*	*5*	*6*
Attitudes to borrowing (5 items)	X	X				
Intra-household decision-making (3 items)		X	X	X	X	X
Household expenditure [l]					X	X
Employment						
Job characteristics [m]	X	X	X	X	X	X
Family friendly workplace (3 items)	X					
Access to family friendly benefits (7 items)	X	X	X	X	X	X
Parenting						
Parenting stress (4 items)	X	X	X	X	X	X
Fairness of childcare	X	X	X	X	X	X
Work family gains and strains [m]	X	X	X	X	X	X
Other						
Attitudes about work and gender roles [m]	X				X	
Attitudes to marriage/ children (10 items)					X	
Benefits of employment (14 items)					X	
Personality (36 items)					X	
Sex		X	X	X	X	X
Age group		X	X	X	X	X

i 5 additional items included in this wave.
j Additional question on the number of clubs a member of asked in these waves.
k 7-item version included in Wave 6.
l List of items expanded to include consumer durables from Wave 6.
m List of items changed and extended in Wave 5.

APPENDIX 1b: SURVEY INSTRUMENT DEVELOPMENT AND SOURCES

The following provides a summary of the origin behind many of the questions and data items included in the instruments for the HILDA Survey. If an item is not listed it can be assumed that the question was either a generic item (such as the date of birth or sex of an individual) or was developed specifically for the HILDA Survey with no, or minimal, reference to previous survey instruments.

Household Form (HF)

Note on overall structure:

The HF essentially comprises three components:
(i) a record of calls made and outcomes;
(ii) a household grid; and
(iii) questions about all dwellings and refusal information.

The Household Grid was largely inspired by the Household Grid concept used in the BHPS and in the family composition section (Section A) of the Canadian Survey of Financial Security.

First Wave Qstn #	Data item / Topic	Notes on origin / Source
W1: X5a/b	Fraction of time spent living at address	Based on question B7 in the FaCS General Customer Survey (GCS), 2000.
W1: X6a	English language use at home	Based on question asked in the ABS, Population Census.
W1: X6b	English language speaking ability	Response categories identical to those used in the ABS, Population Census.
W1: X7	Long-term disability / chronic health condition	Concepts underlying this question (and the accompanying showcard) based on questions asked in the FaCS GCS and in the ABS Survey of Training and Education.
W1: X12	Intra-household relationships	Many other surveys (e.g., the British Household Panel Survey [BHPS] and the US Panel Study of Income Dynamics [PSID]) ask how each household member is related to a specific reference person in the household. The HILDA Survey, however, may well be the first survey of its type to directly code the relationships between all household members.
W1: Y1	Type of residence	Categories based on ABS, Survey of Income and Housing Costs. The question was moved into the HQ in Wave 2.
W1: Y3	Security features of premises	Adapted from US National Survey of Health and Stress (see Groves and Couper 1998, p. 75).

Household Questionnaire (HQ)

HOUSEHOLD QUESTIONNAIRE		
Note on overall structure: Each year the HQ comprises three main sections, covering: (i) childcare arrangements; (ii) housing and housing mortgages; and (iii) other miscellaneous household characteristics. In addition, the HQ in Waves 2 and 6 included a section on household wealth.		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
CHILDCARE		
W1: Q4	Problems or difficulties with childcare arrangements	Adapted from a comparable question included in the Negotiating the Life Course Study.
W1: Q7 / Q8 / Q10	Type, cost and hours of child care	The structure used is unique to the HILDA Survey, but the types of care identified draw heavily from the Negotiating the Life Course Study. The question sequence was substantially modified in Wave 2. In Wave 1 the scope of questions was restricted to households where all of the carers were in paid employment and only related to employment-related care. In Wave 2 the restriction to persons in paid employment was removed and employment-related care and non-employment-related care separately distinguished. Further changes to the layout of the questions for non-employment related care were introduced in Wave 5.
HOUSING		
W1: R1	Number of bedrooms	Based on questions included in the ABS 1999 Survey of Living Standards pilot (q. D4) and in the BHPS (q. H1a, Wave 1, HQ).
W1: R2	Residence ownership status	Adapted from a question included in the ABS Population Census.
W1: R3	Landlord type	Adapted from a question included in the ABS Population Census.
W1: R4	Rent	Based on q. D9 and q. D10, ABS 1999 Survey of Living Standards pilot.
W1: R10	Value of residence	Adapted from questions asked in the PSID and the BHPS.
W1: R11-R21	Mortgages / Home loans	While the structure is quite different, a number of the questions included here are quite similar (especially R15) to questions included in the US Survey of Consumer Finances (SCF).

HOUSEHOLD QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
OTHER		
W1: R27-R29	Household expenditure on groceries and meals out	Questions on expenditure on food and groceries and on meals out are asked each year in the BHPS. The question format, however, is markedly different (e.g., food is not separated from other groceries, use of banded response options, data on meals out are collected from individuals rather than households). The questions were not included in Wave 2, and have been omitted permanently since Wave 6 (given the inclusion of extensive expenditure questions in the SCQ).
W1: R30	Subjective income adequacy	European Community Household Panel Study. This question was only included in Wave 1.
W2: Y1	Dwelling type	Categories based on ABS, Survey of Income and Housing Costs. In Wave 1 this item was included as part of the HF.
HOUSEHOLD WEALTH (Waves 2 and 6)		
Waves 2 and 6 included special modules on household wealth, which were split across the PQ and HQ. While the HILDA Survey questions are distinct, their development was informed by questions included in previous surveys, most notably the SCF, but also the PSID, BHPS and GSOEP. The questions were designed in collaboration with staff from the Reserve Bank of Australia. The household component covered housing and property, business assets and liabilities, equity-type investments (e.g., shares, managed funds) and cash-type investments (e.g., bonds, debentures), vehicles and collectibles (e.g., art works). In answering all questions, respondents were asked to provide exact dollar amounts. In Wave 6 most questions were modified to enable those who were unsure of the value of the asset to select a pre-coded banded category. Wave 6 also saw the inclusion of additional questions on home loan refinancing, investment properties and unpaid overdue household bills. Data on the value of the primary residence are collected every wave.		

Person Questionnaire (PQ)

PERSON QUESTIONNAIRE		
Note on overall structure The PQ is administered to every person aged 15 years and over (on 30 June) in the household. There are two versions of the PQ: one for persons who have not previously responded (NPQ) and for previous wave respondents (CPQ).		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
BIOGRAPHICAL HISTORY		
W1: A2	Year of arrival	Based on a question asked in the BHPS, but with the addition of the 6-month residency requirement.
W1: A3	English language	Adapted from ABS 1993 Survey of Training and Education.
W1: A4	Indigenous origin	Question text based on a comparable question in the Population Census. Response options are as used in the ABS Monthly Population Survey (i.e., the Labour Force Survey).
W1: B1	Parental presence at age 14	International Social Science Survey, Australia (IsssA) 1999.
W1: B2	Reason for not living with both own parents at age 14	Re-worded version of question asked in IsssA 1999.
W1: B3a	Parents ever separated / divorced	IsssA 1999.
W1: B3b	Parents ever reunited after separation / divorce	IsssA 1999.
W1: B7-B9	Siblings	Based on similar questions asked in the PSID and the 1998 SCF.
W1: B12	Employment status of father at age 14	Similar questions asked in both the BHPS and PSID.
W1: B13	Occupation of father	Basic approach to measuring occupation follows standard ABS practice.
W1: B15	Employment status of mother at age 14	Similar questions asked in both the BHPS and PSID.
W1: B16	Occupation of mother	Basic approach to measuring occupation follows standard ABS practice.
W1: C1	Age left school	Adapted from FaCS GCS.
W1: C2	Highest year of school completed	Revised version of question in ABS 1993 Survey of Training and Education. Showcard based on information provided in ABS, *How to Complete Your Census Form*, p. 10 (ABS, Canberra, 2001).

PERSON QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W1: C6	Post-school qualifications	Based on a question included in the ABS 1999 Living Standards Survey pilot.
W1: C7a	Type and number of post-school qualifications	Response categories used are based on those used in various ABS surveys (e.g., the 1993 Survey of Training and Education and the 1999 Survey of Living Standards pilot). The list of categories, however, was extended to distinguish different levels of Certificate qualifications.
W1: C7c	Type of nursing qualification	Categories based on those used in the ABS 1999 Survey of Living Standards pilot.
W1: C7d	Type of teaching qualification	Categories based on those used in the ABS 1999 Survey of Living Standards pilot.
W1: C10A	Current education enrolment	Based on a question included in the ABS 1999 Living Standards Survey pilot.
W1: C11a	Type of qualification being studied	Response categories used are based on those used in various ABS surveys (e.g., the 1993 Survey of Training and Education and the 1999 Survey of Living Standards pilot). The list of categories, however, was extended to distinguish different levels of Certificate qualifications.
W1: C11c	Type of nursing qualification being studied	Categories based on those used in the ABS 1999 Survey of Living Standards pilot.
W1: C11d	Type of teaching qualification being studied	Categories based on those used in the ABS 1999 Survey of Living Standards pilot.
W1: D3a	Years in paid work	Modified version of a question included in the ABS Survey of Employment and Unemployment Patterns (SEUP).
W1: D3b	Years unemployed	Modified version of a question included in the ABS SEUP.
W1: D3c	Years not in labour force	Modified version of a question included in the ABS SEUP.
W1: D5	Main activity during years out of labour force	Modified version of a question included in the ABS SEUP.
W1:D12	Time since last worked for pay	Modified version of question asked in the ABS Monthly Population Survey.
W1: D13-D19	Characteristics of last job (persons not currently in paid work)	These items are essentially duplicates of questions listed below about characteristics of the current job.
W1: D20	Reason ceased last job	Based on questions asked in the ABS Monthly Population Survey, February 2000 (Labour Mobility supplement) and the Second Longitudinal Survey of Immigrants to Australia.
W1: J2	Marriage history grid	Based on AIFS Family Formation Project 1990.

PERSON QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W4: AA6-AA12	Visa category (for recent arrivals)	Designed in collaboration with officers from the Department of Immigration and Multicultural Affairs. The question sequence closely follows a similar sequence included in the ABS Monthly Population Survey, November 1999 (Characteristics of Migrants supplement).
EMPLOYMENT STATUS		
W1: D6-D7	Employment status in last week	ABS Monthly Population Survey, with the concept of "last week" replaced by "the last 7 days".
W1: D8	Employment status – main job	ABS Monthly Population Survey (prior to changes introduced in April 2001).
W1: D9	Business incorporation	ABS Monthly Population Survey.
CURRENT EMPLOYMENT		
W1: E1	Hour worked per week – all jobs	ABS 1993 Survey of Training and Education. Question modified in Wave 2 to better measure hours for persons with variable working hours.
W1: E5	Reason for working part-time	Modified version of a question asked in the Canadian Survey of Labour and Income Dynamics (SLID).
W1: E9	Hour worked per week – main job	ABS 1993 Survey of Training and Education. Question modified in Wave 2 to better measure hours for persons with variable working hours.
W1: E10	Days of the week usually worked	ABS, Working Arrangements Survey (Supplement to the LFS).
W1: E11	Number of days usually worked in 4-week period	ABS, Working Arrangements Survey (Supplement to the LFS).
W1: E12	Shift work arrangements	SLID.
W1: E13	Occupation in main job	Based on standard ABS item.
W1: E14	Years in current occupation	Based on question included in ABS 1993 Survey of Training and Education
W1: E15	Current job tenure	ABS Monthly Population Survey, February 2000 (Labour Mobility module).
W1: E16	Industry	Based closely on standard ABS question (but unlike the ABS we do not precede this question with one asking respondents to nominate the name of the business that employs them).

PERSON QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W1: E22	Annual leave entitlements	ABS Monthly Population Survey, August 2000 (Employment Benefits module).
W1: E23	Paid sick leave entitlement	ABS Monthly Population Survey, August 2000 (Employment Benefits module).
W1: E24	Type of employer / business	Based loosely on question used in the 1995 Australian Workplace Industrial Relations Survey (AWIRS).
W1: E26	Contract renewal expectations	ABS Monthly Population Survey, August 1998 (Non Standard Employment module).
W1: E28	Likelihood of losing job in next 12 months	Wisconsin Survey of Economic Expectations (see Manski and Straub 2000).
W1: E29	Likelihood of finding replacement job	Wisconsin Survey of Economic Expectations (see Manski and Straub 2000).
W1: E30	Likelihood of quitting job	Wisconsin Survey of Economic Expectations (see Manski and Straub 2000).
W1: E31	PAYE status	VandenHeuvel and Wooden (1995).
W1: E32	Supervisory responsibilities	BHPS / SLID.
W1: E33	Workplace size	Based on question asked in BHPS.
W1: E35	Firm size	ABS 1993 Survey of Training and Education. Response categories based on those provided in similar question asked of managers in the 1995 AWIRS.
W1: E36	Job satisfaction	Based on question in the BHPS, but with one item added and an 11-point scale used instead of a 7-point scale.
W1: E39	Intended retirement age	FaCS GCS.
W1: C27a-C27c	Work-related training	Adapted from suggestions by Alison Booth (ANU).
W5: C31b	Gender composition of workplace	Expanded version of question included in UN Generations and Gender Survey (GGS), wave 1 (q. 841).
PERSONS NOT IN PAID EMPLOYMENT		
W1: F1	Looking for work	Modified version of question in the ABS Monthly Population Survey.
W1: F2	Job search methods	ABS Monthly Population Survey.
W1: F3	When began looking for work	Modified version of question in ABS Monthly Population Survey.

PERSON QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W1: F5	Availability to start work (unemployed)	ABS Monthly Population Survey.
W1: F6 / F7	Reasons had trouble getting a job	Based on ABS Monthly Population Survey, July 2000 (Job Search Experience of Unemployed Persons module).
W1: F8	Number of job offers	ABS SEUP (Wave 2, q. S122).
W1: F10	Main activity since last worked or looked for work	Modified version of a question included in the ABS SEUP.
W1: F11	Work intentions	Based on ABS Monthly Population Survey, September 2000 (Persons Not in the Labour Force module).
W1: F12 / F13	Reasons for not looking for work in the last 4 weeks	Based on ABS Monthly Population Survey, September 2000 (Persons Not in the Labour Force module).
W1: F16	Availability to start work (not looking for work))	Based on ABS Monthly Population Survey, September 2000 (Persons Not in the Labour Force module).
W1: F17	Preparedness to start work	ABS Monthly Population Survey, September 2000 (Persons Not in the Labour Force module).
W1: F23b	Intended age of retirement	FaCS GCS.
W5: D12	Main reason for preferring not to work.	FaCS.
W5: D16	Factors influencing decision to start looking for work	FaCS.
INCOME		
W1: G1-G33	Income	All of the income questions are taken directly from, or based on, the ABS Survey of Income and Housing Costs, 1999/2000.
W1: G34	Credit card ownership / payment strategy	Canadian Survey of Financial Security.
FAMILY FORMATION		
W1: H3	Non-resident child grid	Based on the AIFS Family Formation Project 1990 and the AIFS Australian Divorce Transitions Project 1997. The grid used in Wave 1 (and hence in the NPQ) is slightly different from that used in the CPQ in subsequent waves. The grid was modified for Wave 5 to explicitly identify deceased children (similar to what was done in the GGS).

PERSON QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W1: H5	Child support payments	AIFS Australian Divorce Transitions Project 1997.
W1: H15	Children with parent living elsewhere grid	Based on the AIFS Family Formation Project 1990 and the AIFS Australian Divorce Transitions Project 1997.
W1: H18	Child support received	AIFS Australian Divorce Transitions Project 1997.
W1: H26	Desire to have children	Modified version of question asked in the Negotiating the Life Course Study.
W1: H29	Intended number of children	Modified version of question asked in the National Survey of Families and Households.
W5: G28	Responsibility for child care tasks	GGS, wave 1, q. 201.
W5: G36a-G56	Pregnancy and fecundity	Most of the questions in this sequence are drawn directly from, or based on, questions included in the GGS, wave 1.
W5: G63	Factors influencing the decision to have a child	Adapted from a question asked in the 1987-88 National (US) Survey of Families and Households (and analysed in Schoen et al. 1997).
PARTNERING / RELATIONSHIPS		
W1: J4	Duration of current de facto relationship	Modified version of a question asked in the AIFS Life Course Study.
W1: J5	Likelihood of marriage	AIFS Life Course Study.
W1: J6	De facto relationships history	Based on a question asked in the National Survey of Families and Households.
W1: J7	Number of de facto relationships	National Survey of Families and Households.
W1: J8 / J9	Duration of first de facto relationship	Based on a question asked in the AIFS Life Course Study.
W5: H10-H19	Non co-residential relationships	Peter McDonald (ANU).
LIVING IN AUSTRALIA		
W1: K1	Current health status	SF-36 Health Survey (Ware et al. 2000).
W1: K2	Health condition or disability status	Question text comes from FaCS GSC. The list of activities used to define disability, however, comes from the ABS Survey of Training and Education.
W1: K3	Impact of disability or condition on work	A similar question is asked in many surveys, including the BHPS and the PSID.

PERSON QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W1: K4	How much condition limits work	Conceptually similar question questions asked in many surveys (e.g., the BHPS and the PSID), but do not employ the 11-point scale that is used here.
W1: K6-K7	Life satisfaction	The format of the question is based on one included in the GSOEP, but the content is largely driven by the work of Cummins (1996).
W1: K9	Views about life in Australia	ACNielsen
W1: K10	Date began living at current address	Combination of questions from the BHPS and the US SCF.
W1: K14	Reasons for moving in last year	Mostly based on a question included in the PSID, but extensively revised. Also draws on questions included in the BHPS and the ABS SEUP.
W4: K5-K13	Disability	Most of the questions in this sequence are drawn directly from, or closely based on, questions included in the ABS 2002 General Social Survey (sequence 6.2).
W5: K5-K12	Caring	Most of the questions in this sequence are drawn directly from, or closely based on, questions included in the ABS 1998 Survey of Disability, Ageing and Caring (sequence 3.1).
WEALTH (Waves 2 and 6)		
As discussed earlier, while the HILDA Survey questions on wealth are unique, their development was informed by questions included in previous surveys, most notably the SCF, but also the PSID, BHPS and GSOEP (and designed in collaboration with staff from the Reserve Bank of Australia). The person component covered bank accounts, superannuation, credit cards, and personal debts. In Wave 6 the key question on personal debt (W2: J23) was significantly expanded. Two new questions on outstanding personal bills were also added.		
RETIREMENT (Wave 3)		
W3: L2a	Retirement status	US Health and Retirement Study (HRS), Wave 1.
W3: L4	Whether retirement voluntary or involuntary	HRS, Wave 1.
W3: L6a	Reason for retirement	English Longitudinal Survey of Ageing (ELSA). List of response options has been extended and modified.
W3: L18	Desired retirement age	Adapted from LaTrobe University, Healthy Retirement Project.
W3: L19-L20	Expected probability of working past age 65 / 75	Adapted from ELSA.
W3: L21	Influences on the decision to retire	WA Public Service Retirement Intentions Study (plus FACS Work and Retirement Study).

PERSON QUESTIONNAIRE (c'td)		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W3: L22	Expected sources of retirement income	FACS Work and Retirement Study.
W3: L28	Expected changes in work hours	LaTrobe University, Healthy Retirement Project.
W3: L29	Expected financial situation in retirement.	LaTrobe University, Healthy Retirement Project.
W3: L41, L74	Reasons for changing employer	Based on question asked in FACS Work and Retirement Study.
W3: L61	How life has changed since retirement	Adapted from questions asked in the National Survey of Families and Households.
W3: L62	Attitudes about life in retirement	LaTrobe University, Healthy Retirement Project (but with one additional item).
PRIVATE HEALTH INSURANCE (Wave 4)		
Developed collaboratively with staff from the Centre for Health Economics Research Evaluation, University of Technology Sydney, and from the Melbourne Institute of Applied Economic and Social Research (Applied Microeconomics program).		
YOUTH ISSUES (Wave 4)		
W4: L12	Employment intentions at age 35	NLSY79.
W4: L13	Desired occupation at age 35	NLSY79.
W4: L17	Performance at school	Adapted from questions included in LSAY95.
INTENTIONS AND PLANS (Wave 5)		
The question sequence here is based on a proposal designed by Peter McDonald (ANU).		
TRACKING		
W1: T4	Likelihood of moving in next 12 months	Adapted from question asked in the BHPS.
INTERVIEWER OBSERVATIONS		
W1: Z1	Presence of others during interview	BHPS.
W1: Z2	Influence exerted by others on respondent	BHPS.
W1: Z3	Understanding of questions	1998 SCF.
W1: Z4	Suspicion about study	1998 SCF.
W1: Z5	Frequency respondent referred to documentation	1998 SCF.
W1: Z6	Degree of cooperation	BHPS.
W1: Z7	Presence of problems	BHPS.

Self-Completion Questionnaire (SCQ)

SELF-COMPLETION QUESTIONNAIRE		
First Wave Qstn #	**Data item / Topic**	**Notes on origin / Source**
W1: A1-A11d	General health and well-being	SF-36 Health Survey (Ware et al. 2000). The Standard English (Australia / New Zealand) Version 1.0 is employed.
W1: B1	Frequency of moderate / intensive physical activity	Based on a question used in the ABS 1995 National Health and Attrition Survey. The wording of the two questions, however, is very different and, unlike the ABS survey, pre-coded categories are used.
W2:B2	Smoking frequency	Developed with advice from staff at the Australian Institute of Welfare (AIHW). In Wave 1 a simpler version of this question (three response categories instead of five) was included.
W2: B3	Number of cigarettes smoked per week	Developed with advice from staff at the AIHW.
W1: B4	Frequency of alcohol consumption	Based on a question included in the AIHW 1998 National Drug Strategy Household Survey. The question was amended slightly in Wave 2 to provide for one additional response category. The order of the response categories was also reversed, bringing it more in line with AIHW practice.
W1: B5	Daily consumption of alcohol when drinking	AIHW 1998 National Drug Strategy Household Survey (q. H14). The question was amended slightly in Wave 2 to provide for one additional response category.
W1: B6	Frequency of feeling pressed for time	ABS 1999 Survey of Living Standards pilot (q. L1).
W1: B7	Frequency of spare time	ABS 1999 Survey of Living Standards pilot (q. L3).
W1: B9	Neighbourhood characteristics	Based on a question occasionally used in the IsssA and the British Social Attitudes Survey. Four items are taken directly from the BSA Survey, one is a modified version of a BSA item, three are direct from IsssA and two are new. Like the IsssA, a 5-point scale is used (the BSA uses a 4-point scale), but the bottom category has been relabelled and the lead-in question is different.
W6: B12	Neighbourhood characteristics	In wave 6 an additional battery of items seeking respondents' views about the neighbourhood was included. These new items were taken from Sampson et al. (1997).
W1: B10	Housing adequacy	Based on a question used in the Tasmanian Healthy Communities Survey (HCS). The HILDA question, however, only uses six items (not 11), one of which is not from the HCS. The categories are also labelled differently.
W1: B11	Satisfaction with family life	Taken from AIFS Australian Living Standards Study (Part 4, q. 103), but asked on an 11-point scale rather than a 9-point scale.

SELF COMPLETION QUESTIONNAIRE (c'td)		
First wave: qstn #	***Data item / topic***	***Notes on origin / source***
W1: B12	Perception of whether doing fair share of the housework	Negotiating the Life Course Study.
W1: B14	Frequency of social interaction	Based on a question asked in the Tasmanian HCS.
W1: B15	Social support	The first seven items come from Henderson et al. (1978), while the last three items are from Marshall and Barnett (1993).
W1: B16	Time use	Based loosely on a question included in the GSOEP. An extended version of the final question used was piloted as part of the IsssA 2000. In wave 2 two additional categories were added (for paid employment and looking after other people's children), and the response categories amended to seek both hours and part hours (i.e., minutes) data.
W2: B16	Life events	The list of life events was informed by the list originally used by Holmes and Rahe (1967) in their development of a stressful life events measure.
W3: B10	Self-efficacy	Pearlin and Schooler (1978).
W3: B20	Marital quality	Hendrick (1988). The Hendrick scale comprised 7-items – the 6 used here, as well as one item on satisfaction with relationship which HILDA asks every wave as part of its battery on satisfaction with relationships. Hendrick also labelled the mid-point on the scale whereas in the HILDA Survey only the extreme points are labelled.
W4: B18	Religious denomination	Pre-coded categories selected on the basis of the most frequent responses to the 2001 Census.
W5: B10	Satisfaction with division of household tasks	Adapted from two questions asked in the GGS (wave 1, q. 202 and q. 402).
W5: B16	Responsibility for household tasks	GGS (wave 1, q. 401). The list of response options was slightly extended in the HILDA Survey while one item ('organising joint social activities') was omitted.
W5: B17-B18	Use of domestic help	B17 came from GGS (wave 1, q. 404).
W5: B19	Personality traits	Closely based on measure developed by Saucier (1994). The final list of 36 items includes 30 items taken directly from Saucier's original list of 40.
W6: B6-B7	Height and weight	Generic questions, but format influenced by like questions included in the 1994 AYS.
W6: B21	Community participation	Helen Berry, National Centre for Epidemiology and Population Health, ANU.
W1: C1	Financial well-being (self-assessed prosperity)	Tested as part of IsssA 2000 (q. 5, p. 84).
W1: C2	Stressful financial events	Based closely on ABS 1999 Survey of Living Standards pilot (q. H6).

SELF COMPLETION QUESTIONNAIRE (c'td)		
First wave: qstn #	***Data item / topic***	***Notes on origin / source***
W1: C3a	Ability to raise $2000 in an emergency	Inspired by ABS 1999 Survey of Living Standards pilot (q. H4). The ABS survey, however, did not seek to identify how difficult it would be to raise the money, only whether it was possible or not.
W1: C3b	Source of money in an emergency	Categories based on those used in Canadian Survey of Financial Security (q. L14).
W1: C4	Family's savings habits	1998 SCF (X3015-3020).
W1: C5	Savings time horizon	1998 SCF (X3008).
W1: C6	Risk preference	1998 SCF (X3014), but with addition of option: "I never have any spare cash". This question was substantially modified in Wave 6.
W1: C7	Attitudes to borrowing	Based closely on 1998 SCF (X402-406).
W2: C9	Intra-household decision-making	The first version of this question included just three items and was developed with little external input. In Wave 5 the question was modified to bring it more in line with the format of a like question included in the GGS (wave 1, q. 405). The list of items was thus increased to seven (four of which were taken directly from the GGS) and the number of response options expended.
W5: C5-C8	Household expenditure	This section was largely developed specifically for the HILDA Survey, but drawing in some small part on the evidence reported in Browning et al. (2003). The structure of the question set was significantly modified in Wave 6.
W1: D1	Attitudes about work and gender roles	The original question comprised 14 items: five come from Galinsky (1999), two from the Negotiating the Life Course Study, and four from the PSID / NSFH, leaving three items that were created specifically for the HILDA Survey. In Wave 5 two items from the original list of 14 were removed and five new items added. The new items were all drawn from the GGS (wave 1, q. 1113).
W5: D1	Trust	In Wave 5 two items used in the GGS, but originally drawn from the World Values Survey were included. The questions were re-formatted using the standard 7-point agree / disagree scale which is widely used in the HILDA Survey SCQ. In Wave 6 an additional five items were included. These items were suggested by Helen Berry (ANU) and are based on the Organisational Trust Inventory used in Berry and Rodgers (2003).
W5: D2	Attitudes about marriage and children	GGS (wave 1, q. 1107). For the HILDA Survey 7-point scales were used rather than the 5-point scale proposed in the GGS.
W5: D4	Perceived benefits of paid employment	Based on a proposal from FaCS.

SELF COMPLETION QUESTIONNAIRE (c'td)		
First wave: qstn #	***Data item / topic***	***Notes on origin / source***
E1	Job characteristics	The 12-items used here, or variants of them, have been included in a great number of surveys about job characteristics. Four of the items, however, are taken directly from the IsssA. In wave 5 a further nine items were added. All of these items were drawn from the "PATH Through Life Project" (run by the Centre for Mental Health Research, ANU).
E2	Family friendliness of workplace.	Inspired by work of Marshall and Barnett (1993).
F2	Parenting stress	PSID Child Development Supplement 1997, Primary Caregiver of Target Child – Household Questionnaire (q. A29).
F3	Perception of whether doing fair share of the child care	Negotiating the Life Course Study.
F4	Work-family gains and strains	Marshall and Barnett (1993). In Wave 1 the question only included 12 of the 26 original items used by Marshall and Barnett. In Wave 5 the list of items used was expanded to 16.

List of Acronyms

ABS	Australian Bureau of Statistics
AIFS	Australian Institute of Family Studies
AIHW	Australian Institute of Health and Welfare
AWIRS	Australian Workplace Industrial Relations Survey
BHPS	British Household Panel Survey
BSA	British Social Attitudes
FaCS	Family and Community Services (Department of), now known as the Department of Families, Housing, Communities and Indigenous Affairs (FaHCSIA)
GCS	General Customer Survey
GSOEP	German Socio-Economic Panel
HCS	Healthy Communities Survey
HILDA	Household, Income and Labour Dynamics in Australia
IsssA	International Social Science Survey, Australia
LFS	Labour Force Survey
PSID	Panel Study of Income Dynamics
SCF	Survey of Consumer Finances
SEUP	Survey of Employment and Unemployment Patterns
SLID	Survey of Labour and Income Dynamics

APPENDIX 2: CHANGES TO THE DATA FROM RELEASE 5.1

Note that changes to the data between earlier releases are listed on the HILDA website: http://www.melbourneinstitute.com/hilda/manual/userman_previous_releases.html.

Changes to master file between releases 5.1 (March 2007) and 6.0 (January 2008)

- Sampling strata variable (ahhstrat) created from wave 1 major statistical region variable (ahhmsr) and serpentine ordering of CDs across Australia. The same stratification is given to new household members when they join the sample at a later wave. The previous name given to this stratification variable on the master file (ahhmsr) has a different specification on other files (i.e. it would not be applicable if person was not enumerated in wave 1).

- *[In-confidence master file only]. Minor corrections (about 30 cases) to most detailed levels of geography variables as now using the new updated national address file (GNAF). Minor changes to SEIFA raw scores which are matched by geography.*

- *[In-confidence master file only]. Date of birth updated for 94 cases. This is usually a single correction to day of birth or month of birth or year of birth. Or the date of birth was unknown to the person answering the Household Form at a prior wave and has been supplied in the current wave.*

Changes to wave 5 files between releases 5.1 (March 2007) and 6.0 (January 2008)

Cleaning

- ebncoth2 F16 "Current pensions and benefits - other (specify)". Removed 13 cases as contained other income, not pensions or benefits.

- Interviewer observations. 23 cases corrected in Z2 - Z10 due to skip issue. CPQ and NPQ codes at Z2 altered from wave 7 to prevent this reoccurring.

- ehgage Age at 30 June 2005. Due to revisions in date of birth, 45 ages changed (usually by plus or minus 1 year). Age changes effected for prior waves.

Corrections in calculations

- eoifrnta Rental income - 15 cases that were imputed set to zero. The “Already reported (as part of business income)” code in the questionnaire was one generally used to indicate “no answer” so these cases were inappropriately set to missing and imputed. As already reported, rental income set to zero.

- ejhljii2 3 cases corrected to ANZSIC ‘no further information’ codes.

- eehtuj eehtjb eehto History variables for employment activities (time working, unemployed or other activity) since leaving full time education. The algorithm was updated which more accurately calculates the calendar periods to be added to previously known activities. The algorithm was revised for individuals who were not interviewed for one or more waves, resulting in an increase in the number of ‘unable to be determined’ cases.

- emrn History variable for number of marriages. Algorithm corrected from not asked (-1) to zero marriages for never married individuals (who are not asked the number of marriages).

- eedhists and eedhigh – For continuing persons no direct question was made of school year attended in current year (a new question has been added for wave 7). For each year continuing persons haven't left school and are enrolled, the history variable 'Highest year of school completed/currently attending' (_edhists) is incremented by one year until year 12 is achieved. This caused some movements between 'Year 11 and below' and 'Year 12' in _edhigh.

New variables

- ehgdli1 to ehgdli14 Date last interviewed (HF)

- ehgwli1 to ehgwli14 Wave last interviewed (HF)

- ehgdli date last interviewed (PQ)

- ehgwsli Weeks since last interviewed (for continuing persons)

- 1 digit version of occupation and industry variables

- Replicate weights 31 thru 45

Deleted variables

- History variables : Previous job

Documentation corrections

- Household form marked up questionnaire (both u&c). At page 2 said DOB at June 30 2004, corrected to June 30 2005.

- In “Derived Variable Coding Framework e51c.pdf” page D80 it incorrectly states that orderly (epnorder) is reversed in the construction of epnconsc ‘Personality scale – Conscientiousness’. Orderly is not reversed in the construction of this scale.

Changes to wave 4 files between releases 5.1 (March 2007) and 6.0 (January 2008)

Cleaning

- 38 cases corrected in Z2 - Z10 (interviewer observations) due to skip issue. CPQ and NPQ codes at Z2 altered from wave 7 to prevent this reoccurring.

Corrections in calculations

- doifrnta Rental income - 14 cases that were imputed set to zero. The “Already reported (as part of business income)” code in the questionnaire was one generally used to indicate “no answer” so these cases were inappropriately set to missing and imputed. As already reported, rental income set to zero.

- Corrected family number and income unit numbers for household ‘24822’ as relationship had been updated.

- dehtuj dehtjb dehto History variables for employment activity since leaving full time education. The algorithm was updated which more accurately calculates the calendar periods to be added to previously known activities. The algorithm was revised for individuals who were not interviewed for one or more waves, resulting in an increase in the number of ‘unable to be determined’ cases.

- dmrn History variable for number of marriages. Algorithm corrected from not asked (-1) to zero marriages for never married individuals (who are not asked the number of marriages).

- dedhists and dedhigh – For continuing persons no direct question was made of school year attended in current year (a new question has been added for wave 7). For each year continuing persons haven't left school and are enrolled, the history variable 'Highest year of school completed/currently attending' (_edhists) is incremented by one year until year 12 is achieved. This caused some movements between 'Year 11 and below' and 'Year 12' in _edhigh.

New variables

- dhgdli1 to dhgdli14 Date last interviewed (HF)

- dhgwli1 to dhgwli14 Wave last interviewed (HF)

- dhgdli date last interviewed (Eperson)

- dhgwsli Weeks since last interviewed (for continuing persons)

- 1 digit version of occupation and industry variables

- Replicate weights 31 thru 45

Deleted variables

- History variables : Previous job

Documentation corrections

- None

Changes to wave 3 files between releases 5.1 (March 2007) and 6.0 (January 2008)

Cleaning

- Section Z: 30 cases corrected Z2 - Z10.

- Corrected family number and income units for 76571.

Corrections in calculations

- coifrnta Rental income - 14 cases that were imputed set to zero. The “Already reported (as part of business income)” code in the questionnaire was one

generally used to indicate “no answer” so these cases were inappropriately set to missing and imputed. As already reported, rental income set to zero.

- cehtuj cehtjb cehto History variables for employment activity since leaving full time education. The algorithm was updated which more accurately calculates the calendar periods to be added to previously known activities. The algorithm was revised for individuals who were not interviewed for one or more waves, resulting in an increase in the number of ‘unable to be determined’ cases.

- cmrn History variable for number of marriages. Algorithm corrected from not asked (-1) to zero marriages for never married individuals (who are not asked the number of marriages).

- cedhists and cedhigh – For continuing persons no direct question was made of school year attended in current year (a new question has been added for wave 7). For each year continuing persons haven't left school and are enrolled, the history variable 'Highest year of school completed/currently attending' (_edhists) is incremented by one year until year 12 is achieved. This caused some movements between 'Year 11 and below' and 'Year 12' in _edhigh.

New variables

- chgdli1 to chgdli14 Date last interviewed (HF)

- chgwli1 to chgwli14 Wave last interviewed (HF)

- chgdli date last interviewed (Eperson)

- chgwsli Weeks since last interviewed (for continuing persons)

- 1 digit version of occupation and industry variables

- Replicate weights 31 thru 45

Deleted variables

- History variables : Previous job

Documentation corrections

- None

Changes to wave 2 files between releases 5.1 (March 2007) and 6.0 (January 2008)

Cleaning

- Corrected family number and income units for households 52331 and 35712.

Corrections in calculations

- Rental income - 10 cases that were imputed set to zero. The “Already reported (as part of business income)” code in the questionnaire was one generally used to indicate “no answer” so these cases were inappropriately set to missing and imputed. As already reported, rental income set to zero.

- bmrn History variable for number of marriages. Algorithm corrected from not asked (-1) to zero marriages for never married individuals (who are not asked the number of marriages).

- Since wave 2 wealth was first released. Superannuation for retirees (bpwsuprt) 9 cases set to zero as had incorrectly been set to refused. Superannuation for non-retirees (bpwsupwk) 135 cases set to zero as had incorrectly been set to to refused or don't know.Superannuation imputation flags at person level corrected. [Imputed variables and flags at household level had been constructed correctly.]

- bedhists and bedhigh – For continuing persons no direct question was made of school year attended in current year (a new question has been added for wave 7). For each year continuing persons haven't left school and are enrolled, the history variable 'Highest year of school completed/currently attending' (_edhists) is incremented by one year until year 12 is achieved. This caused some movements between 'Year 11 and below' and 'Year 12' in _edhigh.

New variables

- bhgdli1 to bhgdli14 Date last interviewed (HF)

- bhgwli1 to bhgwli14 Wave last interviewed (HF)

- bhgdli date last interviewed (Eperson)

- bhgwsli Weeks since last interviewed (for continuing persons)

- 1 digit version of occupation and industry variables

- Replicate weights 31 thru 45

- Household level joint and own credit card debt supplied for completeness of wealth model. The total credit card debt at household level supplied in previous releases is still provided.

Deleted variables

- History variables : Previous job

Documentation corrections

- bw1dt1 to bw1dt14 (date last interviewed) renamed to bhgdli1 to bhgdli for conformity with later waves.

- bpwsupr renamed to bpwsuprt . Wealth: retireees superannuation. For consistency with household level wealth variables

Changes to wave 1 files between releases 5.1 (March 2007) and 6.0 (January 2008)

Cleaning

- aoifrnta Rental income - 9 cases that were imputed set to zero. The “Already reported (as part of business income)” code in the questionnaire was one

generally used to indicate “no answer” so these cases were inappropriately set to missing and imputed. As already reported, rental income set to zero.

- Corrected family number and income units for 010319 and 003910.
- Where it was unknown whether the respondent’s mother or father was working when the respondent was aged 14, the occupation was set to not asked. Removes 2 father occupations and 4 mother occupations.
- Minor corrections to skip pathways for 3 variables (resets a handful of cases from no answer to not asked).

Corrections in calculations

- ajbmday Main job - days usually worked - code 5 "Other - specify days below" changed to code 8 "Other - specify days below" in accord with code values in subsequent waves.

New variables

- 1 digit version of occupation and industry variables

Deleted variables

- None

Documentation corrections

- None

APPENDIX 3: Little and Su Method

Formulae

The Little and Su method was implemented as follows:

(a) Column (wave) effects of the form

$$c_{hj} = \frac{\bar{Y}_{hj}}{\bar{Y}_{h}}$$

where $\bar{Y}_{h} = \frac{1}{m}\sum_{j}\bar{Y}_{hj}$

were computed for each wave j = 1, …, m, and for each age group h = 1, …, c, where $\bar{Y}_{hj}$ is the sample mean of variable Y for wave j, age group h based on complete cases and $\bar{Y}_{h}$ is the global mean of variable Y for age group h based on complete cases.

(b) Row (person) effects of the form

$$\bar{Y}_{h}^{(i)} = \frac{1}{m_i}\sum_{j}\frac{Y_{hij}}{c_{hj}}$$

were computed for both complete and incomplete cases. Here the summation is over recorded waves for case i; m_i is the number of recorded waves; Y_{hij} is the variable of interest for case i, wave j, age group h; and c_{hj} is the simple wave correction from (a).

(c) Cases were ordered by $\bar{Y}_{h}^{(i)}$, and incomplete case i is matched to the closest complete case, say *l* within age group h.

(d) Missing value Y_{hij} was imputed by

$$\hat{Y}_{hij} = \left[\bar{Y}_{h}^{(i)}\right]\left[c_{hj}\right]\left[\frac{Y_{hlj}}{\bar{Y}_{h}^{(l)}c_{hj}}\right]$$

$$= Y_{hlj}\frac{\bar{Y}_{h}^{(i)}}{\bar{Y}_{h}^{(l)}}$$

where the three terms in square parentheses represent the row, column, and residual effects, the first two terms estimate the predicted mean, and the last term is the stochastic component of the imputation from the matched case.

Example

Suppose we have the following small sample of fictitious responses to current wages and salaries.

All cases

OBS	Wages & Salaries		
	Wave 1	Wave 2	Wave 3
1		400	420
2	675	235	700
3	345	690	800
4	200	480	210
5	200		
6	350	370	
7	400	450	470
8	0	790	790
9	360	450	600
10	135	130	200

From this example, we see that observation 1 did not respond to the current wages and salaries question in wave 1, but provided responses in subsequent waves. Observations 5 and 6 also partially responded and wages and salaries information are not provided in all 3 waves.

The first step in the Little and Su method is to calculate the column effects based on complete cases only. Complete cases were defined as individuals that were interviewed in all 3 waves and responded in all 3 waves for the variable of interest. In this example, the complete cases are:

Complete cases

OBS	Wages & Salaries		
	Wave 1	Wave 2	Wave 3
2	675	235	700
3	345	690	800
4	200	480	210
7	400	450	470
8	0	790	790
9	360	450	600
10	135	130	200

The column effects are calculated using formula (a) above and are computed to be:

Column effects

OBS	Wages & Salaries		
	Wave 1	Wave 2	Wave 3
1		400	420
2	675	235	700
3	345	690	800
4	200	480	210
5	200		
6	350	370	
7	400	450	470
8	0	790	790
9	360	450	600
10	135	130	200
	0.70	**1.06**	**1.24**

The Little and Su method incorporates trend information into the imputed amounts via the column effects. In this example, the wave 1 column effect of 0.70 indicates that the mean current wages and salaries in wave 1 is 30% lower than the overall mean current wages and salaries, and the means in waves 2 and 3 are 6% and 24% higher than the overall mean, respectively.

Next, the row effects are calculated using formula (b) above and are computed to be:

Row effects

OBS	Wages & Salaries			
	Wave 1	Wave 2	Wave 3	
1		400	420	**357**
2	675	235	700	**585**
3	345	690	800	**596**
4	200	480	210	**303**
5	200			**287**
6	350	370		**425**
7	400	450	470	**459**
8	0	790	790	**460**
9	360	450	600	**475**
10	135	130	200	**159**
	0.70	**1.06**	**1.24**	

The sample is then ordered by the row effects, and the closest donor is identified.

Sorted by row effects

OBS	Wages & Salaries			
	Wave 1	Wave 2	Wave 3	
10	135	130	200	**159**
5	200			**287**
4	200	480	210	**303**
1		400	420	**357**
6	350	370		**425**
7	400	450	470	**459**
8	0	790	790	**460**
9	360	450	600	**475**
2	675	235	700	**585**
3	345	690	800	**596**

Once the closest donor has been identified, the missing value is imputed by multiplying the actual value for the variable of interest of the donor with the row effect of the recipient divided by the row effect of the donor.

In this example, the imputed current wages and salary amounts using the Little and Su method are highlighted below.

Impute missing values

OBS	Wages & Salaries		
	Wave 1	Wave 2	Wave 3
10	135	130	200
5	200	455	199
4	200	480	210
1	236	400	420
6	350	370	436
7	400	450	470
8	0	790	790
9	360	450	600
2	675	235	700
3	345	690	800