HILDA User Manual – Release 5

Last modified 23/04/07

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

Acknowledgements

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

Simon Freidin, in consultation with Mark Wooden, Bruce Headey and Nicole Watson, prepared the HILDA Data Files and developed the associated data processing system with the support of Diana Warren and Paul Agius. In addition, Nicole Watson, 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:

> Goode, A. and N. Watson (eds) (2007) *HILDA User Manual – Release 5.0*, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Updates to This Manual

Date	*Update*
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 5.0.
22/05/06	Added supplementary programme in “Creating Longitudinal Files – Balanced Panel” (SPSS, SAS and STATA) to create longitudinal weights variables for Responding and Enumerated Persons specifically to be used in a balanced file of 4 waves (*_lnwte, _lnwtrp*).
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, Tables and Examples. Renumbered Table 3a to 4 and all subsequent Tables. Updated sample figures in Tables now numbered 37, 38 and 39.
16/01/06	Updated Manual for Release 4.0.
14/04/05	Renumbered Appendix 1 to 1A Added Appendix 1B on data item sources.
25/01/05	Added Appendix 4 on Little and Su method.
13/01/05	Original version of manual for Release 3.0.

Contents

LIST OF FIGURES ..... **VI**

LIST OF TABLES ..... **VII**

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 *4*

THE HILDA DATA **6**

ORDERING THE DATA 6
CROSS-NATIONAL EQUIVALENT FILE (CNEF) 7
A REMINDER OF THE SECURITY REQUIREMENTS FOR THE DATA 7
HOW THE DATA FILES ARE PROVIDED 8
STRUCTURE OF THE DATA FILES 8
IDENTIFIERS AND USEFUL VARIABLES 9
MATCHING FILES 11
PANELWHIZ 12
CREATING LONGITUDINAL FILES 13
VARIABLE NAME CONVENTIONS 24
MISSING VALUE CONVENTIONS 29
Numeric Variables *29*
Text Variables *29*
DATA WITH NEGATIVE VALUES 29
CONFIDENTIALISATION 30

DERIVED VARIABLES **31**

AGE AND SEX 31
HISTORY VARIABLES 32
SPATIAL VARIABLES 37
CURRENT EDUCATION VARIABLES 38
CURRENT MARRIAGE AND DEFACTO RELATIONSHIP VARIABLES 41
CHILDREN VARIABLES 41
CHILD CARE VARIABLES 42
EMPLOYMENT VARIABLES 50
CALCULATING HOURLY WAGE RATES 51
CALENDAR VARIABLES 52
FAMILY VARIABLES 52
HEALTH VARIABLES 56
HOUSING VARIABLES 56
TIME USE VARIABLES 56
HOUSEHOLD EXPENDITURE VARIABLES 57
LIFE – PERSONALITY VARIABLES 58
INCOME VARIABLES AND INCOME IMPUTATION 59
Income, Tax and Family Benefits Model *59*
Imputation Method *66*
Imputed Income Variables *68*
WEALTH VARIABLES AND WEALTH IMPUTATION (WAVE 2 SPECIAL TOPIC) 70
Wealth Model *70*
Imputation Method *72*
Imputed Wealth Variables *72*
WEIGHTS 74
Cross-Sectional Weights *74*
Longitudinal Weights *76*

Replicate Weights *77*
Weights Provided on the Data Files *77*
Advice on Using Weights *78*

DOCUMENTATION **82**

DOCUMENTATION CHOICES 82
MARKED-UP QUESTIONNAIRES 82
VARIABLE LISTINGS 83
Derived Variable Listing *83*
File-Based Listing *84*
Subject Listing *84*
Cross-Wave Variable Listing *85*
FREQUENCIES 85

DATA QUALITY ISSUES **86**

MISSING INCOME DATA AND THE EXTENT OF INCOME IMPUTATION 86

THE HILDA SAMPLE **91**

SAMPLE DESIGN 91
Overview *91*
Reference Population *91*
Sampling Units *92*
Sample Selection *92*
FOLLOWING RULES 93

DATA COLLECTION **94**

PILOT TESTING 94
QUESTIONNAIRE LENGTH 94
INTERVIEWERS 95
FIELDWORK PROCESS 95
Data Collection Mode *95*
Timeline *96*
Survey Notification Material *96*
Respondent Incentives *96*
Call routine, Follow-Up and Refusal Aversion *97*
Foreign Language Interviews *97*
Interviewer Monitoring *97*
RESPONSE RATES 98
ATTRITION BIAS 103
DATA PROCESSING 105
Data Entry *105*
Coding Responses *105*

HILDA USER TRAINING **106**

GETTING MORE INFORMATION **107**

REFERENCES **108**

APPENDIX 1A: SUMMARY OF HILDA SURVEY CONTENT, WAVES 1 – 5 ..... **110**

APPENDIX 1B: SURVEY INSTRUMENT DEVELOPMENT AND SOURCES **119**

APPENDIX 2A: CHANGES TO THE DATA FROM RELEASE 5.0 ..... **135**

APPENDIX 2B: CHANGES TO THE DATA FROM RELEASE 4.1 ..... **139**

APPENDIX 3: LITTLE AND SU METHOD **147**

FORMULAE 147
EXAMPLE 148

List of Figures

Figure 1: The Evolution of the HILDA Survey Sample 3

Figure 2: Family where new defacto relationship is formed 53

Figure 3: Family where new child is born 54

Figure 4: Family with child under 15 and non-dependent children 54

Figure 5: Income units in a family with child under 15 and non-dependent children 55

Figure 6: Financial Year Income Model: Household 61

Figure 7: Financial Year Income Model: Enumerated Person 62

Figure 8: Financial Year Income Model: Responding Person 63

Figure 9: Wealth Model Diagram, Wave 2 Household-level 71

Figure 10: Example of the marked-up questionnaires 83

Figure 11: Example of the derived variable listing 83

Figure 12: Example of the file-based listing 84

Figure 13: Example of the subject listing 84

Figure 14: Example of the cross-wave variable listing 85

Figure 15: Example of the frequencies 85

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, 2, 3 and 4 6

Table 4: HILDA Survey data users by type: Release 1, 2, 3 and 4 6

Table 5: Broad subject area naming conventions, characters 2 and 3 (sorted by code) 24

Table 6a: Unchanged variable name and differing codes for Household Response Status 26

Table 6b: Unchanged variable name and differing codes for Person Response Status 27

Table 6c: Unchanged variable name and differing codes for SCQ Field Response Status 28

Table 6d: Unchanged variable name and differing codes for Household Membership 28

Table 6e: Unchanged variable name and differing codes for New Location of mover 28

Table 7: History Variables 32

Table 8: Derived spatial variables 38

Table 9: Derived current education variables 39

Table 10: Derived current marriage and defacto relationship variables 41

Table 11: Derived children variables 41

Table 12: Derived child care variables 43

Table 13: Derived employment variables 50

Table 14: Derived calendar variables 52

Table 15: Derived family variables 55

Table 16: Derived health variables 56

Table 17: Derived housing variables 56

Table 18: Derived time use variables 57

Table 19: Derived household expenditure variables 57

Table 20: Derived personality variables 58

Table 21: Australian Resident Income Tax Rates, Waves 1- 5 60

Table 22: Derived income variables 65

Table 23a: Number of missing cases imputed using Little and Su method 67

Table 23b: Proportion of missing cases imputed using Little and Su method 68

Table 24: Person imputed income variables 69

Table 25: Derived wealth variables, person-level 72

Table 26: Imputed wealth variables 73

Table 27: Weights 77

Table 28: Sample design variables 81

Table 29a: Number of cases with missing income data, waves 1 – 5 87

Table 29b: Proportion of cases with missing income data, waves 1 – 5 88

Table 30a: Mean income ($) (including imputed values), waves 1 – 5 (unweighted) 89

Table 30b: Proportion of mean income ($) imputed, waves 1 – 5 (unweighted) 90

Table 31: Average time (minutes) taken to complete questionnaires, waves 1 to 5 94

Table 32: Number of interviewers and percentage of new interviewers each wave 95

Table 33: Proportion of respondents interviewed by telephone (%) 95

Table 34: Fieldwork dates and proportion of interviews post Dec 2005 96

Table 35: Wave 1 Household Outcomes 98

Table 36: Wave 1 Person Outcomes 98

Table 37: Wave 2 Household Outcomes 100

Table 38: Wave 3 Household Outcomes 100

Table 39: Wave 4 Household Outcomes 101

Table 40: Wave 5 Household Outcomes 101

Table 41: Wave 2 - 5 Person Outcomes Against Wave 1 Person Outcomes 102

Table 42: Response rates (%) for the HILDA Survey, waves 2, 3, 4 and 5 compared 102

Table 43: Self Completion Questionnaire response rate by wave 103

Table 44: Proportion of Wave 1 Respondents Re-interviewed by Selected Sample Characteristics . 104

USING THIS MANUAL

The HILDA Survey is more formally 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 all of 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 third version of the manual, and as such we welcome any feedback you have. It is an evolving manual and is being 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 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 – Household wealth;
- Wave 3 – Retirement and plans for retirement;
- Wave 4 – Private health insurance, and youth;
- Wave 5 – Fertility and Partnering, Personality, Household Expenditure, Intentions and Plans.

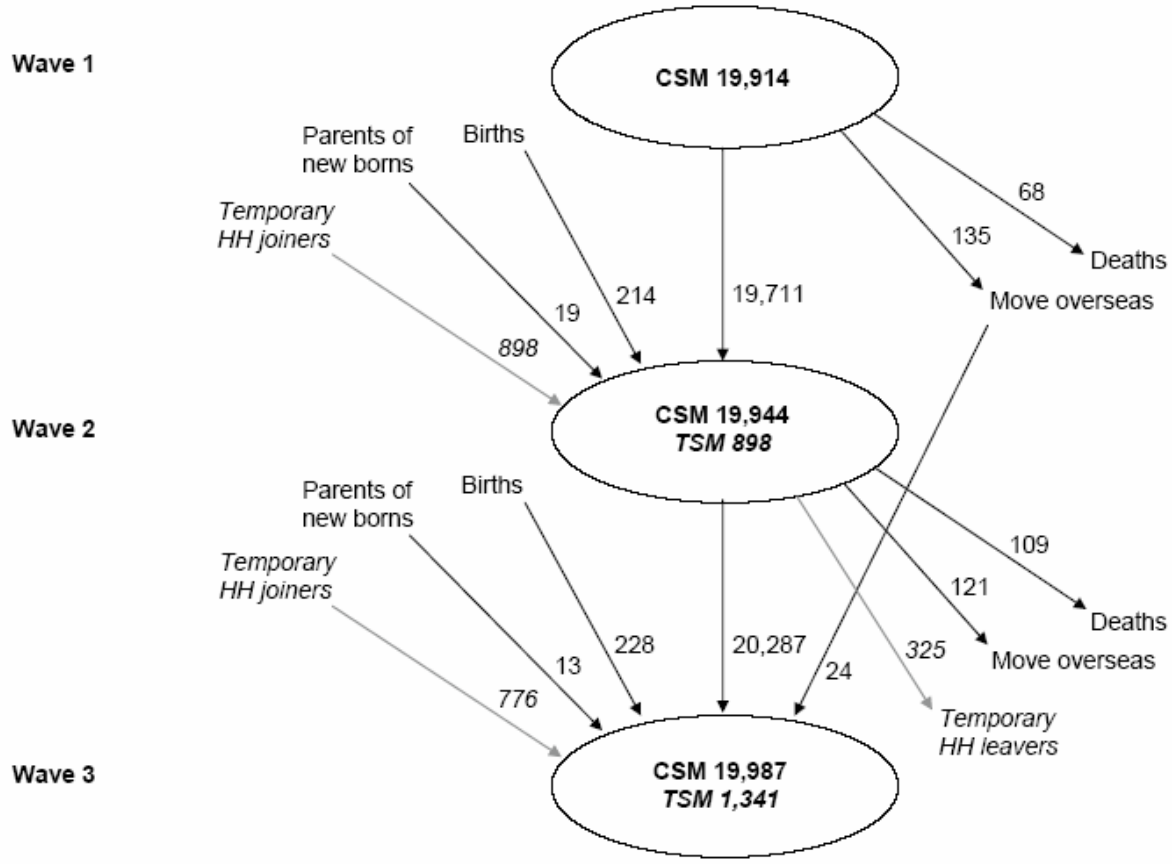
The HILDA Sample and Following Rules: A Summary

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



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. Background history is elicited for each NPQ.

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 and, in wave 2, 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.

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 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 only)	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	

*Note: Immigration Status asked in wave 4 in section AA

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 persons who have been approved access to at least one of the last four data releases and the composition of our user community.

Table 3: HILDA Survey total number of data users: Release 1, 2, 3 and 4

Release	*Total data orders*	*Orders by new users*	*Cumulative no. of users*
Release 1	204	204	204
Release 2	266	167	371
Release 3	280	154	527
Release 4	314	163	690

Table 4: HILDA Survey data users by type: Release 1, 2, 3 and 4

Type of user	*Release 1 (final)*	*Release 2 (final)*	*Release 3 (final)*	*Release 4 (final)*
Academic – Australia	84	105	118	138
Academic – Overseas	5	13	20	18
Students – Honours year	5	14	18	14
Students – Postgraduate	9	23	24	33
Government – Australian	87	90	85	96
Government – State/Local	7	13	5	10
Other	7	8	10	5
Total	204	266	280	314

Ordering the Data

The data can be obtained by:

- visiting the 'data and documentation' page on the HILDA website: www.melbourneinstitute.com/hilda/data.html;
- completing the appropriate order form and deed of licence – 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);
- sending these completed documents to the address indicated on the order form, together with your payment of $77 for administration costs ($132 for overseas users);
- waiting approximately two to three weeks for the delivery of the data.

When your order form and deed of licence reaches the Department of Families, Community Services and Indigenous Affairs (FaCSIA), you will receive an email confirmation that it has arrived. If you do not receive this, then email: hilda@facs.gov.au as soon as possible to avoid delays in your application.

Employees of FaCSIA have different access arrangements and should contact hilda@facs.gov.au.

Up until Release 4.0 the HILDA data files have been referred to as the "confidentialised" and "unconfidentialised" files. From Release 5.0 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 NCEF 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, you MUST contact FaCSIA prior to doing so to dicuss 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.
- If you change your research project you MUST seek permission for the new project from FaCSIA
- The HILDA CD-rom MUST be kept secure in a locked filing cabinet or other secure container when not in use.
- 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 8 characters long, and include some non-alphanumeric characters such as #, ;,*, etc.
- Any printed unit record output MUST 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 completed the order form and have the deed of licence signed by both yourself and the FaCSIA delegate).

- 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 CD-Rom 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 5.0 and between Release 5.0 and Release 4.1 are provided in Appendix 2a and 2b, respectively. Changes to the data files, including those made for earlier releases, can be found on the HILDA website: Changes to data files between Releases

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, one master file is provided with the files for the most recent wave (for Release 5.0, this is wave 5). The master file contains all persons enumerated at any wave, their interview status in each wave and limited information about the individual.

[1] SPSS portable files provided on previous releases are no longer provided on the CD 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.
[3] A trial copy of StatTransfer Version 8 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 *_hhrpid*. In wave 1, the first six characters of the person identifier 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 in *_hhprtrid* on the enumerated and responding person files using the information in the HF relationship grid. Partners are either married or de-facto and includes same sex couples. *_hhprtid* is the person number within 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. Alternatively, you can use the partner's crosswave identifier (*-hhpxid*).

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

Note that *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 these 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*.

Listed below are some common socio-demographic variables which are useful and are intended to provide a guide to the new HILDA user to get started using the 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.

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	Cross-wave 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 Stata
_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*	FT 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 icnome (positive and negative)	*_edhigh*	Highest education level achieved
_hhda(1-10)	SEIFA decile of socio-economic disadvantage	*_hglth*	Long term health condition/disability/impairment

Matching Files

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

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

```
/* ************************************************************ */
/* Created by: Simon Freidin                                    */
/* ************************************************************ */
libname w1 'e:\release5.0\a50c';
data w1p;
  set w1.rperson_a50c;
data w1h;
  set w1.household_a50c (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;
```

Example 2: SPSS program to match wave 1 household and responding person files

```
/* ************************************************************ */
/* Created by: Simon Freidin                                    */
/* ************************************************************ */

file handle w1p /name='e:\release5.0\a50c\Rperson a50c.sav'.
file handle w1h /name='e:\release5.0\a50c\Household a50c.sav'.
file handle w1psort /name='e:\release5.0\a50c\Rperson sorted a50c.sav'.
file handle w1hsort /name='e:\release5.0\a50c\Household sorted a50c.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.
```

Example 3 : STATA program to match wave 1 household, enumerated and responding person files

```
/* ************************************************************ */
/* Created by: Alison Goode                                     */
/* ************************************************************ */

*Enumerated Person File
use "C:\eperson a50c.dta",clear
sort ahhrpid
save "C:\ eperson_2001.dta"
clear

*Responding Person File
use "C:\ rperson a50c.dta",clear
sort ahhrpid
save "C:\ rperson_2001.dta"
```

Example 3 (c'td)

```
clear

*Household File
use "C:\household a50c.dta",clear
sort ahhrhid
save" C:\household_2001.dta"
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"                                /*3: now save cross-section of combined files*/

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

PanelWhiz

PanelWhiz is a collection of STATA/SE Add-On scripts to make using panel datasets easier and allows the use to select vectors of variables at once and the matching and merging is done automatically. More information and how to order PanelWhiz can be found at: http://www.melbourneinstitute.com/hilda/doc/doc_panelwhiz.htm

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 Examples 4 and 5 below. The wide files are created by matching the responding or enumerated files for each wave together using *xwaveid*.

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 Example 6 below.

Example SAS, SPSS and STATA programs to create wide files are provided in Examples 7, 8 and 9, respectively.

The longitudinal weights only support the full balanced panel of respondents and enumerated persons from wave 1 (i.e., across the first two, three, four or five waves). Out of scopes are treated as acceptable outcomes, so these people have weights applied as well.

Example 4: SAS program to create long longitudinal files

```
/* Example SAS program to create balanced enumerated person     */
/* (combined) long longitudinal file.              */
/*Created by: Simon Freidin */
/* ************************************************************ */
/*  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\a50\SAS a50c';
  libname wave2 'e:\b\written datasets\b50\SAS b50c';
  libname wave3 'e:\c\written datasets\c50\SAS c50c';
```

Example 4: (c'td)

```
libname wave4 'e:\d\written datasets\d50\SAS d50c';
 libname wave5 'e:\d\written datasets\e50\SAS e50c';

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

* Master file;
 data master;
 set wave5.master_d50c (keep=xwaveid ahhrpid afstatus bhhrpid bfstatus chhrpid cfstatus dhhrpid dfstatus
ehhrpid efstatus);
 * Wave 1 enumerated persons;
 data w1c;
   set wave1.Combined_a50c;
 * Wave 2 enumerated persons;
 data w2c;
   set wave2.Combined_b50c;
 * Wave 3 enumerated persons;
 data w3c;
   set wave3.Combined_c50c;
 * Wave 4 enumerated persons;
 data w4c;
   set wave4.Combined_d50c;
 * Wave 5 enumerated persons;
 data w5c;
   set wave5.Combined_d50c;
 run;

* (1.2) Identify cases responding in all waves from the master file;
*      status codes: 1=face to face interview, 2=telephone interview;
*      Release 5 balanced panel, n = 9311 ;
data masters (keep=xwaveid ahhrpid bhhrpid chhrpid dhhrpid ehhrpid);
 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);
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;
   if _n_=1 then put " rename";
   if _name_ not in
("XWAVEID","AHHRHID","AHHRPID","BHHRPID","BHHRHID","CHHRPID","CHHRHID","DHHRPID","DHHRHID",
"EHHRPID","EHHRHID")
   then do;
     _name2_=_name_ || '=' || substr(_name_,2);
     put " " _name2_;
   end;
   if eof then put " ;";
 run;
 %mend;
run;
```

Example 4: (c'td)

```
/* ************************************************************** */
/* 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;
  %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';
```

Example 4: (c'td)

```
proc sort data=masters; by xwaveid;
proc sort data=longcw5; by xwaveid;
data longcw5;
  merge masters (in=inc) longcw5;
  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);
  * 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;
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;

/* Supplementary programme to create longitudinal weights variables for Responding and Enumerated
Persons specifically to be used in a balanced file of 5 waves. */
/* Simon Freidin 16/05/2006 */

data weights;
  set wave5.Combined_e50c;
  keep xwaveid elnwtrp elnwte;
  rename elnwtrp=lnwtrp elnwte=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;
```

Example 4: (c'td)

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

Example 5: 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              */

/* ************************************************************** */
/*   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 - THEY DESCRIBE WHERE THE INPUT FILES RESIDE ON
YOUR COMPUTER).
file handle master  /name='e:\d\written datasets\e50\SPSS e50c\Master e50c.sav' .
file handle w1c     /name='e:\a\written datasets\a50\SPSS a50c\Combined a50c.sav'.
file handle w2c     /name='e:\b\written datasets\b50\SPSS b50c\Combined b50c.sav'.
file handle w3c     /name='e:\c\written datasets\c50\SPSS c50c\Combined c50c.sav'.
file handle w4c     /name='e:\d\written datasets\d50\SPSS d50c\Combined d50c.sav'.
file handle w5c     /name='e:\d\written datasets\e50\SPSS e50c\Combined e50c.sav'.

* (1.1.2) Output: generated renaming program, balanced master file.
* (YOU MUST EDIT THE NAME PATHS - THEY DESCRIBE WHERE INTERMEDIATE FILE WILL RESIDE
ON YOUR COMPUTER).
file handle renprog  /name='e:\xwave\rename.sps'.
file handle balanced /name='e:\xwave\balanced.sav'.
* (1.1.3) Output: intermediate longitudinal files.
* (YOU MUST EDIT THE NAME PATHS - THEY DESCRIBE WHERE INTERMEDIATE FILES WILL
RESIDE ON YOUR COMPUTER).
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'.
file handle w5l     /name='e:\xwave\w5l.sav'.

* (1.1.4) Output: final balanced long responding person file.
* (YOU MUST EDIT THE NAME PATH - IT DESCRIBES WHERE THE FINAL FILE WILL RESIDE ON
YOUR COMPUTER).

file handle w1w5long /name='e:\xwave\w1w5long.sav'.

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

Example 5: (c'td)

```
*       Release 5 balanced panel, master n = 9311.
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).
sort cases by xwaveid.
save outfile=balanced/keep=xwaveid ahhrpid bhhrpid chhrpid dhhrpid ehhrpid.

* (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).
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.
```

Example 5: (c'td)

```
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.


* (2.3) Add datasets together to create the balanced long longitudinal
*       responding person file. Add a year variable. Balanced panel long n= 9311
add files file=w1l/in=w1/file=w2l/in=w2/file=w3l/in=w3/file=w4l/in=w4/file=w5l/in=w5.
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.

save outfile=w1w5long.

freq vars=year.

/*(2.4) Supplementary programme to create longitudinal weights variables for Responding and Enumerated
Persons specifically to be used in a balanced file of 4 waves. */

get file=w1w5long./drop=lnwtrp lnwte.
sort cases by xwaveid.
match files file=*table=w5c/rename=(elnwtrp elnwte= lnwtrp lnwte)/keep=xwaveid to year lnwtrp lnwte/by
xwaveid.
means vars=lnwtrp by year.
save outfile=w1w5long_1.
```

Example 6: 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
*************************************************************************************************/
/*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
use "C:\RELEASE 5.0\Master_e50c.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=5 if (afstatus==1|afstatus==2)&(bfstatus==1|bfstatus==2)
                    &  (cfstatus==1|cfstatus==2)&(dfstatus==1|dfstatus==2)&(efstatus==1|efstatus==2);
#delimit cr
label variable totwave "present-5waves"
keep xwaveid totwave ahhrpid bhhrpid chhrpid dhhrpid ehhrpid afstatus bfstatus cfstatus dfstatus efstatus
sort xwaveid
save "C:\RELEASE 5.0\Masterfile_e50c.dta"
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 5.0\Combined_a50c.dta"
renpfix a
gen wave=1
sort xwaveid
save "C:\RELEASE 5.0\Combined_wave1.dta",replace
clear
use "C:\RELEASE 5.0\Combined_b50c.dta"
renpfix b
gen wave=2
sort xwaveid
save "C:\RELEASE 5.0\Combined_wave2.dta"
clear
use "C:\RELEASE 5.0\Combined_c50c.dta"
renpfix c
gen wave=3
sort xwaveid
save "C:\RELEASE 5.0\Combined_wave3.dta"
clear
use "C:\RELEASE 5.0\Combined_d50c.dta"
renpfix d
gen wave=4
sort xwaveid
save "C:\RELEASE 5.0\Combined_wave4.dta"
clear
use "C:\RELEASE 5.0\Combined_e50c.dta"
renpfix e
gen wave=5
sort xwaveid
save "C:\RELEASE 5.0\Combined_wave5.dta"
clear
```

Example 6: (c'td)

```
*Now create a single long file
use "C:\RELEASE 5.0\Masterfile_e50c.dta"
merge xwaveid using "C:\RELEASE 5.0 \combined_wave1.dta"
sort xwaveid
drop _merge
append using "C:\RELEASE 5.0\combined_wave2.dta"
sort xwaveid
append using "C:\RELEASE 5.0\combined_wave3.dta"
sort xwaveid
append using "C:\RELEASE 5.0\combined_wave4.dta"
sort xwaveid wave
append using "C:\RELEASE 5.0\combined_wave5.dta"
sort xwaveid wave
tab wave
tab totwave
by xwaveid:egen totwave1=sum(totwave)

save "C:\RELEASE 5.0\Masterfile_Unbalanced.dta"
clear

*/Section 2: To convert to a balanced panel
* Release 5 balanced panel, n = 9311/*
use "C:\RELEASE 5.0\Masterfile_Unbalanced.dta"
sort xwaveid wave
by xwaveid:keep if totwave1==5
tab wave

*/* Supplementary programme to create longitudinal weights variables for Responding and Enumerated
Persons specifically to be used in a balanced file of all 5 waves. */*

destring xwaveid, gen(id)
sort id wave
by id:egen elnwte_1=max(lnwte) if wave==5
by id:egen elnwte_2=max(elnwte_1)
drop elnwte_1 lnwte
rename elnwte_2 lnwte

sort id wave
by id:egen elnwtrp_1=max(lnwtrp) if wave==5
by id:egen elnwtrp_2=max(elnwtrp_1)
drop elnwtrp_1 lnwtrp
rename elnwtrp_2 lnwtrp
sort wave
save "C:\RELEASE 5.0\Masterfile_Balanced.dta"
clear
```

Example 7: SAS program to create wide longitudinal file

```
/*Example SAS program to create unbalanced and balanced wide files*/
/*Created by Paul Agius 2/11/2006*/

/*Setting libraries*/
libname wave1 'E:\Release 5.0\a50c'; run;
libname wave2 'E:\Release 5.0\b50c'; run;
libname wave3 'E:\Release 5.0\c50c'; run;
libname wave4 'E:\Release 5.0\d50c'; run;
libname wave5 'E:\Release 5.0\e50c'; run;

/*Load the library of formats required for the datasets*/
libname library 'E:\Release 5.0\formats'; run;

/*Sorting files by xwaveid for merging*/
proc sort data = wave1.combined_a50c out=w1;
by xwaveid;
proc sort data = wave2.combined_b50c out=w2;
by xwaveid;
proc sort data = wave3.combined_c50c out=w3;
by xwaveid;
proc sort data = wave4.combined_d50c out=w4;
by xwaveid;
proc sort data = wave5.combined_e50c out=w5;
by xwaveid;
proc sort data = wave5.master_e50c out=master;
by xwaveid;
run;

/*Creating unbalanced data file*/
data wide_UNBALANCED;
  merge w1 w2 w3 w4 w5 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);

run;
```

Example 8: SPSS program to create wide longitudinal file

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

match files
  file='E:\e\written datasets\e50\SPSS e50c\Master e50c.sav'  /in=inmaster
 /file='E:\a\written datasets\a50\SPSS a50c\Combined a50c.sav'/in=inwave1
 /file='E:\b\written datasets\b50\SPSS b50c\Combined b50c.sav'/in=inwave2
 /file='E:\c\written datasets\c50\SPSS c50c\Combined c50c.sav'/in=inwave3
 /file='E:\d\written datasets\d50\SPSS d50c\Combined d50c.sav'/in=inwave4
```

Example 8: (c'td)

```
/file='E:\e\written datasets\e50\SPSS e50c\Combined e50c.sav'/in=inwave5
 /by xwaveid.

/* choose one selection option */
/* no 'select if' gives an unbalanced (enumerated) panel: 23904 cases */
/*  interviewed all 5 waves: 9311 cases */
*select if (ahgint) and (bhgint) and (chgint) and (dhgint) and (ehgint).
/* enumerated all 5 waves: 13446 cases */
*select if (ahgenum) and (bhgenum) and (chgenum) and (dhgenum) and (ehgenum).


save outfile='e:\xwave\w1w5c.sav'.
```

Example 9: STATA program to create wide longitudinal file

```
Example STATA program to create Unbalanced and Balanced Wide File
Created by: Alison Goode 27/09/06
*/**********************************************************************************************************/*
*/*Please note - uses all variables. If not all variables required, select and save as a seperate 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 5.0\Masterfile_e50c.dta"
merge xwaveid using "E:\Release 5.0\Combined_a50c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 5.0\Combined_b50c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 5.0\Combined_ c50c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 5.0\Combined_ d50c.dta"
sort xwaveid
drop _merge
merge xwaveid using "E:\Release 5.0\Combined_ e50c.dta"
sort xwaveid
drop _merge
save "E:\Release 5.0\Wide-UNBALANCED_5waves.dta"
*/*************************************************************************
*Section 2: To convert to a balanced wide file
 n==9311
*************************************************************************/*
use "E:\Release 5.0\Wide-UNBALANCED_5waves.dta"
sum afstatus bfstatus cfstatus dfstatus
sort xwaveid
by xwaveid:keep if
(afstatus==1|afstatus==2)&(bfstatus==1|bfstatus==2)&(cfstatus==1|cfstatus==2)&(dfstatus==1|dfstatus==2)
sum afstatus bfstatus cfstatus dfstatus
save "E:\Release 5.0\Wide-BALANCED_5waves.dta"
drop _all
```

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 5) 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 6a, 6b, 6c, 6d and 6e show how the response categories differ for these variables across the first four waves.

Table 5: Broad subject area naming conventions, characters 2 and 3 (sorted by code)

Code	Broad Subject Area	Code	Broad Subject Area	Code	Broad Subject Area
AN	Ancestry	HG	Household enumeration grid	PH	Private Health Insurance
AT	Attitudes and values	HH	Household information	PJ	Previous job
BA	Bank accounts	HI	Household Income	PI	Personal information
BI	Business income	HS	Housing	PN	Personality
BF	Business	HW	Household wealth	PR	Partnering /relationships
BN	Benefits	IC	Intentions to have children	PW	Personal wealth
CA	Calendar	IO	Interviewer observations	RE	Religion

Table 5: (c'td)

Code	Broad Subject Area	Code	Broad Subject Area	Code	Broad Subject Area
CC	Child care general	IP	Intentions and Plans	RC	Resident children
CH	Child care during school holidays	JB	Job characteristics of employed	RG	Relationship grid
CN	Non employment related childcare	JO	Opinions about job	RT	Retirement intentions
CP	Child care for pre-school children	JS	Job search of those not employed	RW	Replicate weight
		LE	Major life events	SA	Superannuation
CR	Credit cards	LN	Longitudinal weights	TC	Total children
CS	Child care during school	LO	Life opinions	TI	Total income
DB	Debt	LS	Lifestyle	TS	Time stamps
DO	Dwelling observations	MH	Moving house	TX	Taxes
DT	Personal debt	MO	Mutual obligations	UJ	Job history of those not in paid employment
		MR	Marital relationships		
ED	Education	MV	Motor vehicles		
EH	Employment history	NC	Non-resident children	WS	Wage and salaries
ES	Employment status	NL	Not in labour force	XP, XY	Household expenditure
FA	Financial assets	OA	Other assets	YE	Youth - employment
FI	Attitudes to finances	OI	Other income	YH	Youth – education
FM	Family background	OP	Other property	YI	Youth - importance
GH	General health and well-being	OR	Other relationships	YP	Youth - property
HE	Health	PA	Parenting	YS	Youth – life satisfaction

Table 6a: Unchanged variable name and differing codes for Household Response Status

Variable and associated response category: 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 ACNielsen 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

Note 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 6b: Unchanged variable name and differing codes for Person Response Status

Variable:_fstatus, initial*_hgri and _hgri1 to hgri14;* follow-up*_hgrf and hgrf1 to hgrf14;* final*_hgivw and _hgivw1 to _hgivw14*[1]			
	Codes used		
Description	*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

Note 1: *_hgri, _hgri1 to hgri14, hgrf and hgrf1 to hgrf14* are only on the In-confidence Release (unconfidentialised files).

Table 6c: Unchanged variable name and differing codes for SCQ Field Response Status

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

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

Table 6d: Unchanged variable name and differing codes for Household Membership

Variable: _hghhm, _hghhm1 to _hghhm14[1]			
Description	*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

Note 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 6e: Unchanged variable name and differing codes for New Location of mover

Variable: _hgnlc1 to _hgnlc1			
Description	*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.

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)

Text Variables

Text variables with missing values will typically contain the following text.

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

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).

From release 3 onwards, 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 weighted means of the top-coded variable will be the same as the original variable. (In previous releases, the cut-off value was used which failed to preserve the weighted means.)

An example of top-coding is _*wscg* (main job, current gross wages per week). All cases whose gross wages are equal to or exceed $4800 have the real value substituted with a value which has been calculated as the weighted average of all those cases whose income is equal to or exceeds $4800.The weighted average wage of all those cases (22 in total) is, let us say, $8450. Another example is the variable for age (_*hgage*) which is top-coded 90 years. The weighted average age for all those who are >89years (38 cases in total) is, let us say, 92 years. $8450 or 92 years are used to substitute for the real wages or real age of all those cases whose wages >=$4800 or age is >89years. The reasons for doing this are twofold. Firstly, it maximizes confidentiality and secondly, it preserves the weighted distribution means. If the distribution of wages or age had been simply cut off at $4800 or 90 years, when the relevant weights are applied, the value would be too low.

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 short description of how the variable was derived is supplied in the coding framework and in this user manual.

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* has also been 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 *_hgage14,* where *_hgage1* is for person 01, etc.

For the small number of cases where age was not provided, it has been imputed, *_hgagef* flags which cases have been imputed.In the General Release (confidentialised) datasets, age is top-coded at 90 years (see chapter on Confidentialisation).

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 that 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 7 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).

Table 7: History Variables

Variable	Description
Ancestry	
_ANCOB	History: Country of Birth
_ANBCOB	History: DV: 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
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

Table 7: (c'td)

Variable	Description
_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: DV: 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 six months
_FMMEMP	History: Was mother in paid employment when you were 14
_FMMOCC[1]	History: Mothers occupation - 4 digit ASCO
_FMMOCC2	History: DV: 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
_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
_EDQ110	History: ASCED 110 Doctoral degree - Number of qualifications
_EDQ120	History: ASCED 120 Master degree - Number of qualifications
_EDQ200	History: ASCED 200 Grad Dip and Grad Cert - Number of qualifications
_EDQ221	History: ASCED 221 Graduate certificate - Number of qualifications
_EDQ211	History: ASCED 211 Graduate diploma - Number of qualifications
_EDQ310	History: ASCED 310 Bachelor degree - Number of qualifications
_EDQ311	History: ASCED 311 Bachelor (Honours) degree - Number of qualifications
_EDQ312	History: ASCED 312 Bachelor (Pass) degree - Number of qualifications
_EDQ400	History: ASCED 400 Advanced diploma and diploma - Number of qualifications
_EDQ413	History: ASCED 413 Associate degree - Number of qualifications
_EDQ421	History: ASCED 421 Diploma - Number of qualifications

Table 7: (c'td)

Variable	Description
_EDQ411	History: ASCED 411 Advanced diploma - Number of qualifications
_EDQ500	History: ASCED 500 Certificate - dont know level - Number of qualifications
_EDQ524	History: ASCED 524 Certificate level I - Number of qualifications
_EDQ521	History: ASCED 521 Certificate level II - Number of qualifications
_EDQ514	History: ASCED 514 Certificate level III - Number of qualifications
_EDQ511	History: ASCED 511 Certificate level IV - Number of qualifications
_EDQ600	History: ASCED 600 Secondary education - Number of qualifications
_EDQ611	History: ASCED 611 Year 12 - Number of qualifications
_EDQUNK	History: Unknown - not enough information - Number of qualifications
_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
_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

Table 7: (c'td)

Variable	Description
_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
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: DV: 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
_JHTSJHA	History: How long since last worked for pay answered weeks or years (not currently in paid work)
_JHWKU	History: How long since last worked for pay – less than 1 year - weeks (not currently in paid work)
_JHLHRU	History: How long since last worked for pay - years (not currently in paid work)
_JHLHRUW	History: Hours a week usually worked in last job (not currently in paid work)
_JHLJIND	History: Industry, last job (not currently in paid work)
_JHLJTYP	History: Employment status, last job (not currently in paid work)
_JHLJCNT	History: Contract of employment, last job (not currently in paid work)
_JHLJTHA	History: Employment length – answered weeks or years, last job (not currently in paid work)

Table 7: (c'td)

Variable	Description
_JHLJTWK	History: weeks worked for last employer (not currently in paid work)
_JHLJTYR	History: Years worked for last employer (not currently in paid work)
_JHLJOCC	History: Occupation, last job (not currently in paid work)
_JHLJREA	History: Main reason stopped working (not currently in paid work)
_JHLJOCS	History: ANU4 occupational status scale: Last job (not currently in paid work)
_JHLJI88[1]	History: ISCO_88: Last job (not currently in paid work)
_JHLJ182	History: ISCO_88 2 digit: Last job (not currently in paid work)
_JHLJOC2	History: Occupation last job (not currently in paid work) 2-digit ASCO
_JHLJII2	History: isic 3.1 Last job ISCO_88: Last job (not currently in paid work) International industry classification 2-digit
Health[3]	
_HESPNCY	History: Year condition first developed – Sight problems
_HEHEARY	History: Year condition first developed – Hearing problems
_HESPCHY	History: Year condition first developed – Speech problems
_HEBFLCY	History: Year condition first developed – Blackouts/fits/loss of consciousness
_HESLUY	History: Year condition first developed – Difficulty learning
_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 – Nervous/emotional condition requiring treatment
_HECRYPAY	History: Year condition first developed – Condition restricting physical activity/work
_HEDISFY	History: Year condition first developed – Any disfigurement or deformity
_HEMIRHY	History: Year condition first developed – Any mental illness requiring help/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 from head injury/stroke/other brain damage
_HEMEDY	History: Year condition first developed – Long term condition/ailment still restrictive despite current treatment or current medication taken
_HEOTHY	History: Year condition first developed –Other long term condition (eg arthritis, asthma, heart disease, Alzeiheimers 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 5.0 onwards, 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.

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.

Table 8: 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
_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
_HHMOVE	HF Household moved address since previous wave
_HHMOVEK	DV: Distance moved since last wave (kilometers), available from wave 2 onwards
_HHMOVEM	DV: Distance moved since last wave (miles), available from wave 2 onwards
AHHCD96[1]	DV: 1996 ABS Census Collection District
_HHCD01[1]	DV: 2001 ABS 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 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 9: Derived current education variables

Variable	Description
_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
_EDQ421N	DV: ASCED 421 Diploma - Number of qualifications
_EDQ413N	DV: ASCED 413 Associate Degree - 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
_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 9: (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
_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
_EDRQUNK	DV: ASCED Unknown - Not enough information - Recent qualifications
_EDFTS	DV: Full-time student
AEDQ613	DV: ASCED 613 Year 11 – Number of qualifications obtained since leaving school
AEDQ620	DV: ASCED 620 Year 7 to 10 - Number of qualifications obtained since leaving school
AEDCQ613	DV: ASCED 613 Year 11 - Currently studying
AEDCQ620	DV: ASCED 620 Years 7 to 10 - Currently studying
BEDRQ600	DV: ASCED 600 Secondary school - Lower level – 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 10.

Table 10: 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
_ORDFRLT[2]	DV: Years living together, most recent defacto

1. Waves 1,4 & 5 (NPQ) only
2. Waves 2 and 3 only

Children Variables

Table 11 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 11: Derived children variables

Variable	Description
_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)
_TCNR	DV: Number of own non-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)
_TCYNG	DV: Age youngest own child (excl. foster/step). Weighted topcode.
_TCR	DV: Count of own resident children
_TCNR	DV: Count of own non-resident children
_NCNGT	DV: Overnight stays of non-resident child (Days per annum)

Table 11: (c'td)

Variable	Description
_NCDAY	DV: Day visits of non-resident child (Days per annum)
_NCEFSPY[1]	DV: Non-resident child maintenance annual - paid - all children ($)
_NCEFSRY[1]	DV: Non-resident child maintenance received - annual - all children ($)
_NCYNG	DV: Age youngest non-resident own child. Weighted topcode.
_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 ($)
_RCEFSRY[1]	DV: Resident child maintenance received - annual - all children ($)
_RCYNG	DV: Age youngest resident own child (excl. foster/step).
ANCEFSY[1]	DV: Child maintenance paid - annual - all children ($)
ARCEFSY[1]	DV: Child maintenance received - annual - all children ($)

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 12). The children referred to in this section of the HQ are those living in the household aged under 15. The summary variables provide a measure of:

- whether a particular type of child care is used;
- how many hours of each care type is used (summed across the relevant children);
- the total cost of child care across all care types (for wave 1 only as the summary variable is collected directly from respondents in later waves).

'School-aged' children are those who are of an age to attend school (that is, from aged 4 or 5, depending on the State, to 15). 'Pre-school' children are those aged 0 to 3 or 4, depending on the State, that have ***not yet started school*** (not just those who aged 3 or 4 who are attending pre-school). The shorter name for those children not yet at school was used in the questionnaire for space reasons and the interviewers were briefed about this issue. The variable labels relating to the children not yet at school have been revised to remove the 'pre-school" reference.

Please note that in wave 5, more detailed questions are asked in the section on non-employment related child care than asked in previous waves. The derived variables at the bottom of Table 12, from variable *_nsctc* onwards, have been generated in order to maintain consistency with waves 2, 3 and 4.

Table 12: Derived child care variables

Variable	Description
_CSU_ME	DV: Childcare used: Me or my partner. Any school-age child
_CSU_BS	DV: Childcare used: The childs brother or sister. Any school-age child
_CSU_SF	DV: Childcare used: Child looks after self. Any school-age child
_CSU_WP	DV: Childcare used: Child comes to my workplace. Any school-age child
_CSU_OS	DV: Childcare used: Out of hours pw care at childs school. Any school-age child
_CSU_OE	DV: Childcare used: Out of hours pw care elsewhere. Any school-age child
_CSU_RU	DV: Childcare used: A relative who lives with us. Any school-age child
_CSU_RE	DV: Childcare used: A relative who lives elsewhere. Any school-age child
_CSU_FO	DV: Childcare used: A friend or neighbour coming to our home. Any school-age child
_CSU_FT	DV: Childcare used: A friend or neighbour in their home. Any school-age child
_CSU_PS	DV: Childcare used: A paid sitter or nanny. Any school-age child
_CSU_FD	DV: Childcare used: Family day care. Any school-age child
_CSU_O1	DV: Childcare used: Other 1 (Specify). Any school-age child
_CSU_O2	DV: Childcare used: Other 2 (Specify). Any school-age child
_CSU_OP	DV: Childcare used: Other parent not living in household/expartner. Any school-age child
_CSU_BR	DV: Childcare used: Not applicable. Boarding school. Any school-age child
_CSU_NA	DV: Childcare used: Not Answered. Any school-age child
_CSH_BS	DV: Total hours pw: The childs brother or sister. All school-age children
_CSH_SF	DV: Total hours pw: Child looks after self. All school-age children
_CSH_WP	DV: Total hours pw: Child comes to my workplace. All school-age children
_CSH_OS	DV: Total hours pw: Out of hours care at childs school. All school-age children
_CSH_OE	DV: Total hours pw: Out of hours care elsewhere. All school-age children
_CSH_RU	DV: Total hours pw: A relative who lives with us. All school-age children
_CSH_RE	DV: Total hours pw: A relative who lives elsewhere. All school-age children
_CSH_FO	DV: Total hours pw: A friend or neighbour coming to our home. All school-age children
_CSH_FT	DV: Total hours pw: A friend or neighbour in their home. All school-age children
_CSH_PS	DV: Total hours pw: A paid sitter or nanny. All school-age children
_CSH_FD	DV: Total hours pw: Family day care. All school-age children
_CSH_O1	DV: Total hours pw: Other specify 1. All school-age children
_CSH_O2	DV: Total hours pw: Other specify 2. All school-age children
_CSCTC	DV: Childcare total cost ($) for all school age children during term, across all types of care
_CHU_ME	DV: Childcare used: Me or my partner. Any school-age child
_CHU_BS	DV: Childcare used: The childs brother or sister. Any school-age child
_CHU_SF	DV: Childcare used: Child looks after self. Any school-age child

Table 12: (c'td)

Variable	Description
_CHU_WP	DV: Childcare used: Child comes to my workplace. Any school-age child
_CHU_VS	DV: Childcare used: Vacation care at childs school. Any school-age child
_CHU_VE	DV: Childcare used: Vacation care elsewhere. Any school-age child
_CHU_RU	DV: Childcare used: A relative who lives with us. Any school-age child
_CHU_RE	DV: Childcare used: A relative who lives elsewhere. Any school-age child
_CHU_FO	DV: Childcare used: A friend or neighbour coming to our home. Any school-age child
_CHU_FT	DV: Childcare used: A friend or neighbour in their home. Any school-age child
_CHU_PS	DV: Childcare used: A paid sitter or nanny. Any school-age child
_CHU_FD	DV: Childcare used: Family day care. Any school-age child
_CHU_O1	DV: Childcare used: Other 1 (Specify). Any school-age child
_CHU_O2	DV: Childcare used: Other 2 (Specify). Any school-age child
_CHU_OP	DV: Childcare used: Other parent not living in household/expartner. Any school-age child
_CHU_NA	DV: Childcare used: Not Answered. Any school-age child
_CHH_BS	DV: Total hours pw: The childs brother or sister. All school-age children
_CHH_SF	DV: Total hours pw: Child looks after self. All school-age children
_CHH_WP	DV: Total hours pw: Child comes to my workplace. All school-age children
_CHH_VS	DV: Total hours pw: Vacation care at childs school. All school-age children
_CHH_VE	DV: Total hours pw: Vacation care elsewhere. All school-age children
_CHH_RU	DV: Total hours pw: A relative who lives with us. All school-age children
_CHH_RE	DV: Total hours pw: A relative who lives elsewhere. All school-age children
_CHH_FO	DV: Total hours pw: A friend or neighbour coming to our home. All school-age children
_CHH_FT	DV: Total hours pw: A friend or neighbour in their home. All school-age children
_CHH_PS	DV: Total hours pw: A paid sitter or nanny. All school-age children
_CHH_FD	DV: Total hours pw: Family day care. All school-age children
_CHH_O1	DV: Total hours pw: Other specify 1. All school-age children
_CHH_O2	DV: Total hours pw: Other specify 2. All school-age children
_CHCTC	DV: Childcare total cost ($) for all school age children during school holidays, across all types of care
_CPU_ME	DV: Childcare used: Me or my partner. Any pre-school child
_CPU_BS	DV: Childcare used: The childs brother or sister. Any pre-school child
_CPU_RU	DV: Childcare used: A relative who lives with us. Any pre-school child
_CPU_RE	DV: Childcare used: A relative who lives elsewhere. Any pre-school child
_CPU_FO	DV: Childcare used: A friend or neighbour coming to our home. Any pre-school child
_CPU_FT	DV: Childcare used: A friend or neighbour in their home. Any pre-school child

Table 12: (c'td)

Variable	Description
_CPU_PS	DV: Childcare used: A paid sitter or nanny. Any pre-school child
_CPU_FD	DV: Childcare used: Family day care. Any pre-school child
_CPU_WD	DV: Childcare used: Long day care centre at workplace. Any pre-school child
_CPU_PD	DV: Childcare used: Private or community long day care centre. Any pre-school child
_CPU_KP	DV: Childcare used: Kindergarten/pre-school. Any pre-school child
_CPU_O1	DV: Childcare used: Other 1 (Specify). Any pre-school child
_CPU_O2	DV: Childcare used: Other 2 (Specify). Any pre-school child
_CPU_OP	DV: Childcare used: Other parent not living in household/expartner. Any pre-school child
_CPU_NA	DV: Childcare used: Not Answered. Any pre-school child
_CPH_BS	DV: Total hours pw: The childs brother or sister. All pre-school children
_CPH_RU	DV: Total hours pw: A relative who lives with us. All pre-school children
_CPH_RE	DV: Total hours pw: A relative who lives elsewhere. All pre-school children
_CPH_FO	DV: Total hours pw: A friend or neighbour coming to our home. All pre-school children
_CPH_FT	DV: Total hours pw: A friend or neighbour in their home. All pre-school children
_CPH_PS	DV: Total hours pw: A paid sitter or nanny. All pre-school children
_CPH_FD	DV: Total hours pw: Family day care. All pre-school children
_CPH_WD	DV: Total hours pw: Long day care centre at workplace. All pre-school children
_CPH_PD	DV: Total hours pw: Private or community long day care centre. All pre-school children
_CPH_KP	DV: Total hours pw: Kindergarten / preschool. All pre-school children
_CPH_O1	DV: Total hours pw: Other specify 1. All pre-school children
_CPH_O2	DV: Total hours pw: Other specify 2. All pre-school children
_CPCTC	DV: Childcare total cost ($) for all preschool children across all types of care
ACSC_WP	DV: Total cost pw: Child comes to my workplace. All school-age children
ACSC_OS	DV: Total cost pw: Out of hours care at childs school. All school-age children
ACSC_OE	DV: Total cost pw: Out of hours care elsewhere. All school-age children
ACSC_RU	DV: Total cost pw: A relative who lives with us. All school-age children
ACSC_RE	DV: Total cost pw: A relative who lives elsewhere. All school-age children
ACSC_FO	DV: Total cost pw: A friend or neighbour coming to our home. All school-age children
ACSC_FT	DV: Total cost pw: A friend or neighbour in their home. All school-age children
ACSC_PS	DV: Total cost pw: A paid sitter or nanny. All school-age children
ACSC_FD	DV: Total cost pw: Family day care. All school-age children
ACSC_O1	DV: Total cost pw: Other specify 1. All school-age children
ACSC_O2	DV: Total cost pw: Other specify 2. All school-age children
ACHC_WP	DV: Total cost pw: Child comes to my workplace. All school-age children
ACHC_VS	DV: Total cost pw: Vacation care at childs school. All school-age children
ACHC_VE	DV: Total cost pw: Vacation care elsewhere. All school-age children
ACHC_RU	DV: Total cost pw: A relative who lives with us. All school-age children
ACHC_RE	DV: Total cost pw: A relative who lives elsewhere. All school-age children

Table 12: (c'td)

Variable	Description
ACHC_FO	DV: Total cost pw: A friend or neighbour coming to our home. All school-age children
ACHC_FT	DV: Total cost pw: A friend or neighbour in their home. All school-age children
ACHC_PS	DV: Total cost pw: A paid sitter or nanny. All school-age children
ACHC_FD	DV: Total cost pw: Family day care. All school-age children
ACHC_O1	DV: Total cost pw: Other specify 1. All school-age children
ACHC_O2	DV: Total cost pw: Other specify 2. All school-age children
ACPC_RU	DV: Total cost pw: A relative who lives with us. All pre-school children
ACPC_RE	DV: Total cost pw: A relative who lives elsewhere. All pre-school children
ACPC_FO	DV: Total cost pw: A friend or neighbour coming to our home. All pre-school children
ACPC_FT	DV: Total cost pw: A friend or neighbour in their home. All pre-school children
ACPC_PS	DV: Total cost pw: A paid sitter or nanny. All pre-school children
ACPC_FD	DV: Total cost pw: Family day care. All pre-school children
ACPC_WD	DV: Total cost pw: Long day care centre at workplace. All pre-school children
ACPC_PD	DV: Total cost pw: Private or community long day care centre. All pre-school children
ACPC_KP	DV: Total cost pw: Kindergarten / preschool. All pre-school children
ACPC_O1	DV: Total cost pw: Other specify 1. All pre-school children
ACPC_O2	DV: Total cost pw: Other specify 2. All pre-school children
DCSU_GU	DV: Childcare used: Childs grandparent who lives with us. Any school-age child. While parents work
DCSU_GE	DV: Childcare used: Childs grandparent who lives elsewhere. Any school-age child. While parents work
DCSU_AU	DV: Childcare used: Other relative who lives with us. Any school-age child. While parents work
DCSU_AE	DV: Childcare used: Other relative who lives elsewhere. Any school-age child. While parents work
DCSU_FC	DV: Childcare used: Formal outside of school hours care. Any school-age child. While parents work
DCSH_GU	DV: Total hours pw: Childs grandparent who lives with us. All school-age children. While parents work
DCSH_GE	DV: Total hours pw: Childs grandparent who lives elsewhere. All school-age children. While parents work
DCSH_AU	DV: Total hours pw: Other relative who lives with us. All school-age children. While parents work
DCSH_AE	DV: Total hours pw: Other relative who lives elsewhere. All school-age children. While parents work
DCSH_FC	DV: Total hours pw: Formal outside of school hours care. All school-age children. While parents work
DCHU_GU	DV: Q9 Childcare used: Childs grandparent who lives with us. Any school-age child. While parents work
DCHU_GE	DV: Childcare used: Childs grandparent who lives elsewhere. Any school-age child. While parents work

Table 12: (c'td)

Variable	Description
DCHU_AU	DV: Childcare used: Other relative who lives with us. Any school-age child. While parents work
DCHU_AE	DV: Childcare used: Other relative who lives elsewhere. Any school-age child. While parents work
DCHU_VC	DV: Childcare used: Vacation care. Any school-age child. While parents work
DCHH_GU	DV: Total hours pw: Childs grandparent who lives with us. All school-age children. While parents work
DCHH_GE	DV: Total hours pw: Childs grandparent who lives elsewhere. All school-age children. While parents work
DCHH_AU	DV: Total hours pw: Other relative who lives with us. All school-age children. While parents work
DCHH_AE	DV: Total hours pw: Other relative who lives elsewhere. All school-age children. While parents work
DCHH_VC	DV: Total hours pw: Vacation care. All school-age children. While parents work
DCPU_GU	DV: Childcare used: Childs grandparent who lives with us. Any not yet at school child. While parents work
DCPU_GE	DV: Childcare used: Childs grandparent who lives elsewhere. Any not yet at school child. While parents work
DCPU_AU	DV: Childcare used: Other relative who lives with us. Any not yet at school child. While parents work
DCPU_AE	DV: Childcare used: Other relative who lives elsewhere. Any not yet at school child. While parents work
DCPH_GU	DV: Total hours pw: Childs grandparent who lives with us. All not yet at school children. While parents work
DCPH_GE	DV: Total hours pw: Childs grandparent who lives elsewhere. All not yet at school children. While parents work
DCPH_AU	DV: Total hours pw: Other relative who lives with us. All not yet at school children. While parents work
DCPH_AE	DV: Total hours pw: Other relative who lives elsewhere. All not yet at school children. While parents work
_NSCTC[1]	DV: Childcare total cost ($) for all school age children across all types of care. Not employment related
_NSU_BS[1]	DV: Childcare used – The childs brother or sister. Any children school aged – not employment related
_NSU_GU[1]	DV: Childcare used – Childs grandparent who lives with us. Any children school aged – not employment related
_NSU_GE[1]	DV: Childcare used – Childs grandparent who elsewhere. Any children school aged – not employment related
_NSU_AU[1]	DV: Childcare used – Other relative who lives with us. Any children school aged – not employment related
_NSU_AE[1]	DV: Childcare used – Other relative who lives elsewhere. Any children school aged – not employment related

Table 12: (c'td)

Variable	Description
_NSU_FO[1]	DV: Childcare used – A friend or neighbour coming to our home. Any children school aged – not employment related
_NSU_FT[1]	DV: Childcare used – A friend or neighbour coming in their home. Any children school aged – not employment related
_NSU_PS[1]	DV: Childcare used – A paid sitter/nanny. Any children school aged – not employment related
_NSU_FD[1]	DV: Childcare used – Family day care. Any children school aged – not employment related
_NSU_PD[1]	DV: Childcare used – Private/community long day care centre. Any children school aged – not employment related
_NSU_FC[1]	DV: Childcare used – Formal outside of school hours care. Any children school aged – not employment related
_NSU_01[1]	DV: Childcare used – Other 1 (Specify). All children school aged – not employment related
_NSU_02[1]	DV: Childcare used – Other 2 (Specify). All children school aged – not employment related
_NSU_NA[1]	DV: Childcare used – No answer. All children school aged – not employment related
_NSH_BS[1]	DV: Total hours pw. – The childs brother or sister. All children school aged – not employment related
_NSH_GU[1]	DV: Total hours pw. – Childs grandparent who lives with us. All children school aged – not employment related
_NSH_GE[1]	DV: Total hours pw. – Childs grandparent who elsewhere. All children school aged – not employment related
_NSH_AU[1]	DV: Total hours pw. – Other relative who lives with us. All children school aged – not employment related
_NSH_AE[1]	DV: Total hours pw. – Other relative who lives elsewhere. All children school aged – not employment related
_NSH_FO[1]	DV: Total hours pw. – A friend or neighbour coming to our home. All children school aged – not employment related
_NSH_FT[1]	DV: Total hours pw. – A friend or neighbour in their home. All children school aged – not employment related
_NSH_PS[1]	DV: Total hours pw. – A paid sitter/nanny. All children school aged – not employment related
_NSH_FD[1]	DV: Total hours pw. – Family day care. All children school aged – not employment related
_NSH_PD[1]	DV: Total hours pw. – Private/community long day care centre. All children school aged – not employment related
_NSH_FC[1]	DV: Total hours pw. – Formal outside of school hours care. All children school aged – not employment related
_NSH_01[1]	DV: Total hours pw.– Other 1. All children school aged – not employment related
_NSH_02[1]	DV: Total hours pw.– Other 2. All children school aged – not employment related
_NPCTC[1]	DV: Childcare total cost ($) for all preschool children across all types of care – not employment related.
_NPU_BS[1]	DV: Childcare used – The childs brother or sister. Any children not yet at school – not employment related.
_NPU_GU[1]	DV: Childcare used – Childs grandparent who lives with us. Any children not yet at school – not employment related.
_NPU_GE[1]	DV: Childcare used – Childs grandparent who lives elsewhere. Any children not yet at school – not employment related.

Table 12: (c'td)

Variable	Description
_NPU_AU[1]	DV: Childcare used – Other relative who lives with us. Any children not yet at school – not employment related.
_NPU_AE[1]	DV: Childcare used – Other relative who lives elsewhere. Any children not yet at school – not employment related.
_NPU_FO[1]	DV: Childcare used – A friend or neighbour coming to our home. Any children not yet at school – not employment related.
_NPU_FT[1]	DV: Childcare used – A friend or neighbour in their home. Any children not yet at school – not employment related.
_NPU_PS[1]	DV: Childcare used – A paid sitter/nanny. Any children not yet at school – not employment related.
_NPU_FD[1]	DV: Childcare used – Family day care. Any children not yet at school – not employment related.
_NPU_PD[1]	DV: Childcare used – Private/community long day care centre. Any children not yet at school – not employment related.
_NPU_KP[1]	DV: Childcare used – Kindergarten/pre-school. Any children not yet at school – not employment related.
_NPU_01[1]	DV: Childcare used – Other 1 (Specify). Any children not yet at school – not employment related.
_NPU_02[1]	DV: Childcare used – Other 2 (Specify) . Any children not yet at school – not employment related.
_NPU_NA[1]	DV: Childcare used – No answer. Any children not yet at school – not employment related.
_NPH_BS[1]	DV: Total hours pw. The childs brother or sister. All children not yet at school – not employment related.
_NPH_GU[1]	DV: Total hours pw. Childs grandparent who lives with us. All children not yet at school – not employment related.
_NPH_GE[1]	DV: Total hours pw. Childs grandparent who lives elsewhere. All children not yet at school – not employment related.
_NPH_AU[1]	DV: Total hours pw. Other relative who lives with us. All children not yet at school – not employment related.
_NPH_AE[1]	DV: Total hours pw. Other releative who lives elsewhere. All children not yet at school – not employment related.
_NPH_FO[1]	DV: Total hours pw. A friend or neighbour coming to our home. All children not yet at school – not employment related.
_NPH_FT[1]	DV: Total hours pw. A friend or neighbour in their home. All children not yet at school – not employment related.
_NPH_PS[1]	DV: Total hours pw. A paid sitter/nanny. All children not yet at school – not employment related.
_NPH_FD[1]	DV: Total hours pw. Family day care. All children not yet at school – not employment related.
_NPH_PD[1]	DV: Total hours pw. Private/community long day care centre. All children not yet at school – not employment related.
_NPH_KP[1]	DV: Total hours pw. Kindergarten/pre-school. All children not yet at school – not employment related.
_NPH_01[1]	DV: Total hours pw. Other 1.
_NPH_02[1]	DV: Total hours pw. Other 2.

1. Wave 5 onwards

Employment Variables

The employment related derived variables are listed in Table 13. The occupation variables are coded to 4-digit Australian Standard Classification of Occupation (ASCO) which is then used to produce:

- the 2-digit ASCO codes;
- ANU4 occupational status scale (which ranges from 0 to 100); and
- the 2-digit International Standard Classification of Occupation-88 (ISCO-88) codes.

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

- the 2-digit ANZSIC codes; and
- the 2-digit International Standard Industry Classification 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 for you.

Table 13: Derived employment variables

Variable	Description
_ESDTL	DV: Labour force status – detail
_ESBRD	DV: Labour force status – broad
_JBHRUC	DV: Combined hrs per week usually worked in all jobs
_JBMHRUC	DV: Combined hrs 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
_JBMOCC2	DV: Occupation 2-digit ASCO
_JBMOCCS	DV: ANU4 occupational status scale, main job
_JBMI82	DV: ISCO-88 2-digit, Occupation main job
_JBMIND2	DV: C14 Current job industry. 2-digit ANZSIC
_JBMII2	DV: ISIC 3.1 Main job: International industry classification 2-digit
_WCPD[1]	DV: Days of paid workers compensation in last 12 months
_ALPD[1]	DV: Days of paid annual leave in last 12 months
_ALSK[1]	DV: days of paid sick leave in last 12 months
_ALOP[1]	DV: days of paid (maternity, paternity, bereavement, family, carers) leave in last 12 months
_ALUP[1]	DV: Days of unpaid leave in last 12 months

Table 13: (c'td)

Variable	Description
_ES	DV: Employment status in main job if currently employed
_JBMUABS	DV: ABS definition C19 Trade union membership
_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
_UJLJOC2	DV: Unemployed: Occupation last job. 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
_UJLJIN2	DV: NPQ:D25 Unemployed last job industry. 2-digit ANZSIC
_UJLJII2	DV: ISIC 3.1 Unemployed, previous job (NPQ): International industry classification 2-digit
_UJLHRUC	DV: Combined 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
_PJOOCC2	DV: CPQ:C46 Occupation in previous job 2-digit ASCO
_PJOOCCS	DV: ANU4 occupational status scale, CPQ: C46 in previous job if job changed
_PJOI82	DV: ISCO-88 2-digit, CPQ: Occupation previous job if job changed
_PJOTOC2	DV: CPQ:D30 Occupation in previous job 2-digit ASCO
_PJOTOCS	DV: ANU4 occupational status scale, CPQ: D30 in previous job if unemployed
_PJOTI82	DV: ISCO-88 2-digit, CPQ: Occupation previous job if unemployed
_PJOTIN2	DV: CPQ:D27 Unemployed previous job industry. 2-digit ANZSIC
_PJOTII2	DV: ISIC 3.1 Unemployed, previous job: International industry classification 2-digit
_PJOIND2	DV: CPQ:C43 Previous job industry. 2-digit ANZSIC
_PJOII2	DV: ISIC 3.1 Previous job: International industry classification 2-digit
AJBPERM	DV: Permanently unable to work
AUJOCCS	DV: ANU4 occupational status scale, last job, not currently in paid work

1. Wave 5 onwards

Calculating hourly wage rates

The following is aimed at pointing you in the right direction if you want to calculate hourly wage rates. By using 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 (*cesbrd*=1 and *cjbhruc*>0 and *cwscei*>0) hwr03 = rnd(*cwscei/cjbhruc*).
if (*desbrd*=1 and *djbhruc*>0 and *dwscei*>0) hwr04 = rnd(*dwscei/djbhruc*).

The above only calculates the hourly wage rate if the respondent:

(i) is employed;
(ii) has current wages and salaries;
(iii) has usual hours worked in all jobs.

If you wish to look at full and part time employed separately, use_*esdtl* (DV: Labour force status – detail). The unimputed cases can be identified using the flag _*wscef* =0.

Please note:
(i) that the hours worked and income questions are asked in separate sections of the person questionnaire.
(ii) due to some respondents reporting low wages/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 14 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 14: 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.

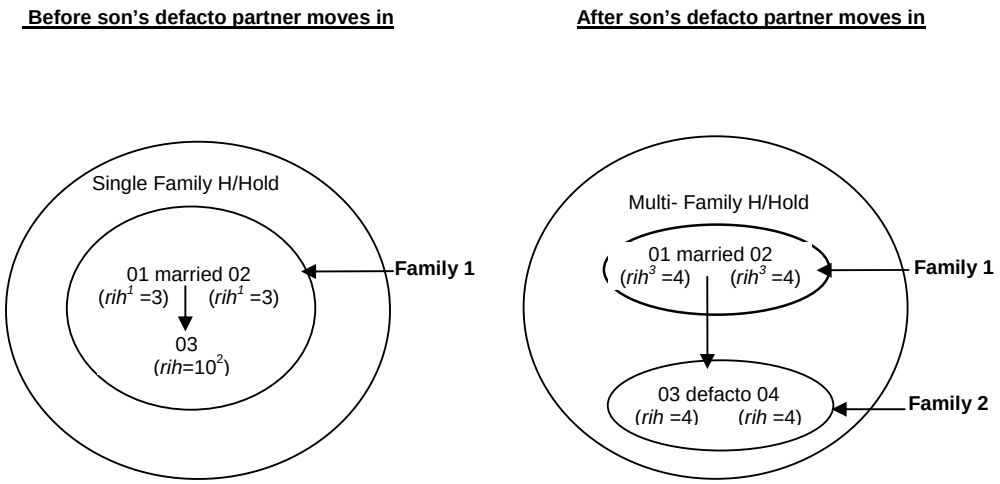
The core relationships needed to make a family are a couple relationship or a parent-child relationship. There are several key points to note about how the families are defined:

- A couple relationship takes precedence over a parent-child relationship (see Figure 2). In a household with mum, dad, 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 mum and dad 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 3 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).

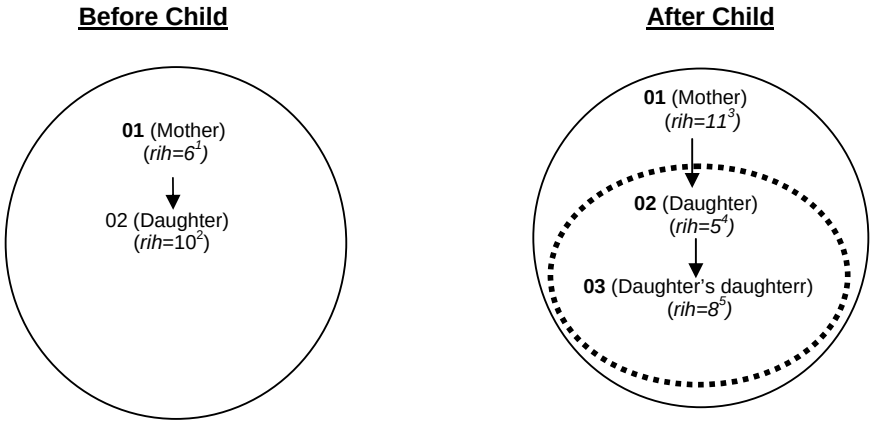
- Children aged under 15 living in a household without natural, adopted, step or foster parents are attached to their closest relative. If they are without relatives, then they are attached to the person thought most likely to form a parent-child relationship with that child.

Figure 2: 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 3: Family where new child is born

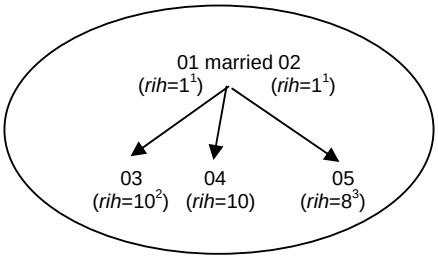


1. *rih* = 6: lone parent with non-dependent child
2. *rih* = 10: Non-dependent child
3. *rih* = 11: Other family member
4. *rih* = 5: Lone parent with dependent child
5. *rih* = 8: Dependent child < 15

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.[5]

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 4 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, grandparents or nannies respectively).

Figure 4: Family with child under 15 and non-dependent children



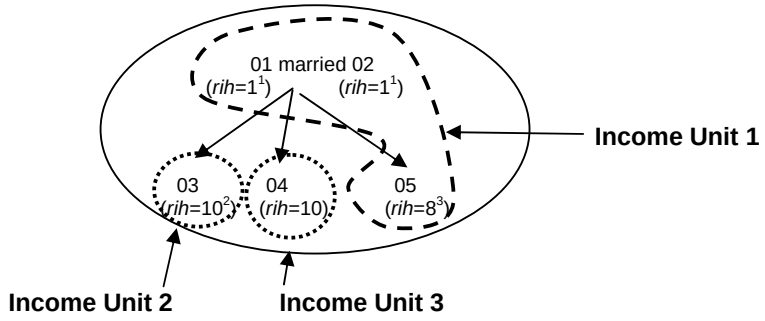
1. *rih* = 1: Couple with children < 15)
2. *rih* = 10: Non-dependent child
3. *rih* = 8: Dependent child < 15

The income units are derived from the family units and separate out the non-dependant children and other relatives from rest of the family. Following this, using the previous

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

scenario from Figure 4 this household is broken into 3 income units. Figure 5 shows how this is done. The first income unit (**1**) includes the couple and the child < 15. Each non-dependent child forms their own income unit (Income units 2 and 3).

Figure 5: 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

A number of other variables are listed with the variables based on the relationship grid in Table 15, including identifiers for various people in the household and counts of the number of people in certain age groups. The partner, father and mother identifiers were discussed in a preceding section on identifiers.

Table 15: 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
_HHFXID	DV: Father's crosswave person number
_HHMXID	DV: Mother's crosswave person number
_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
_HHSM	DV: Type of sample member
_HGINT	DV: Interviewed flag

1. On the Household File, these variables are listed for each person, that is _hhrih01 to _hhrih14, _hhfam01 to _hhfam14, _hhfty01 to _hhfty14, and _hhiu01 to _hhiu14.

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. The variable names are listed in Table 16.

Table 16: 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[1]	DV: SF-6D Health state Classification

Housing Variables

Table 17 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).

Table 17: 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

Time Use Variables

Table 18 lists derived time use variables which combine the hours and minutes spent in a week on various activities.

Table 18: 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 19 lists derived household-level expenditure, as collected in the Self-Completion Questionnaire for 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.

Table 19: Derived household expenditure variables

Variable	Description
_HXYGROC	DV: Household annual expenditure ($) - Groceries
_HXYALC	DV: Household annual expenditure ($) - Alcohol
_HXYCIG	DV: Household annual expenditure ($) - Cigarettes
_HXYPUBT	DV: Household annual expenditure ($) - Public transport and taxis
_HXYMEAL	DV: Household annual expenditure ($) - Meals eaten out
_HXYHSGE	DV: Household annual expenditure ($) - Hobbies, sports, gambling, entertainment
_HXYMVF	DV: Household annual expenditure ($) - Motor vehicle fuel
_HXYCLTH	DV: Household annual expenditure ($) - Clothing and footwear
_HXYTEL	DV: Household annual expenditure ($) - Telephone rent and calls (excl. internet charges)
_HXYHOL	DV: Household annual expenditure ($) - Holidays and holiday travel
_HXYPHI	DV: Household annual expenditure ($) - Private health and accident insurance
_HXYHLTH	DV: Household annual expenditure ($) - Health care
_HXYELEC	DV: Household annual expenditure ($) - Electricity bills
_HXYGAS	DV: Household annual expenditure ($) - Gas bills
_HXYOHF	DV: Household annual expenditure ($) - Other heating fuels
_HXYHMRN	DV: Household annual expenditure ($) - Home repairs/renovations/maintenance

Table 19 (c'td)

Variable	Description
_HXYMVR	DV: Household annual expenditure ($) - Motor vehicle repairs/maintenance
_HXYEDUC	DV: Household annual expenditure ($) - Education fees
_HXYUTIL	DV: Household annual expenditure ($) - Electricity, gas and other heating fuel bills
_XPYGROC	DV: Typical annual expenditure ($) - Groceries
_XPYALC	DV: Typical annual expenditure ($) - Alcohol
_XPYCIG	DV: Typical annual expenditure ($) - Cigarettes
_XPYPUBT	DV: Typical annual expenditure ($) - Public transport and taxis
_XPYMEAL	DV: Typical annual expenditure ($) - Meals eaten out
_XPYHSGE	DV: Typical annual expenditure ($) - Hobbies, sports, gambling, entertainment
_XPYMVF	DV: Typical annual expenditure ($) - Motor vehicle fuel
_XPYCLTH	DV: Typical annual expenditure ($) - Clothing and footwear
_XPYTEL	DV: Typical annual expenditure ($) - Telephone rent and calls (excl. internet charges)
_XPYHOL	DV: Typical annual expenditure ($) - Holidays and holiday travel
_XPYPHI	DV: Typical annual expenditure ($) - Private health and accident insurance
_XPYHLTH	DV: Typical annual expenditure ($) - Health care
_XPYELEC	DV: Typical annual expenditure ($) - Electricity bills
_XPYGAS	DV: Typical annual expenditure ($) - Gas bills
_XPYOHF	DV: Typical annual expenditure ($) - Other heating fuels
_XPYHMRN	DV: Typical annual expenditure ($) - Home repairs/renovations/maintenance
_XPYMVR	DV: Typical annual expenditure ($) - Motor vehicle repairs/maintenance
_XPYEDUC	DV: Typical annual expenditure ($) - Education fees
_XPYUTIL	DV: Typical annual expenditure ($) - Electricity, gas and other heating fuel bills

Life – Personality Variables

For the first time in Release 5.0 respondents were asked questions about their personality character traits based on the Big 5 Personality Inventory. Five scales have been created and are listed in Table 20 below.

Table 20: Derived personality variables

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

Income Variables and Income Imputation

Income, Tax and Family Benefits Model

Figure 6, Figure 7 and Figure 8 show how the numerous income questions in the Person Questionnaire are combined together to form several financial year income components and one 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.[6] 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).[7]

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

From Release 4.0 the income components have been imputed on the enumerated file (rather than providing total financial year income and windfall income alone). This has permitted the calculation of these components at the household level as detailed in Figure 6. 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).

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

[7] This is a change for 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 21). 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 21: Australian Resident Income Tax Rates, Waves 1- 5

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

Figure 6: Financial Year Income Model: Household

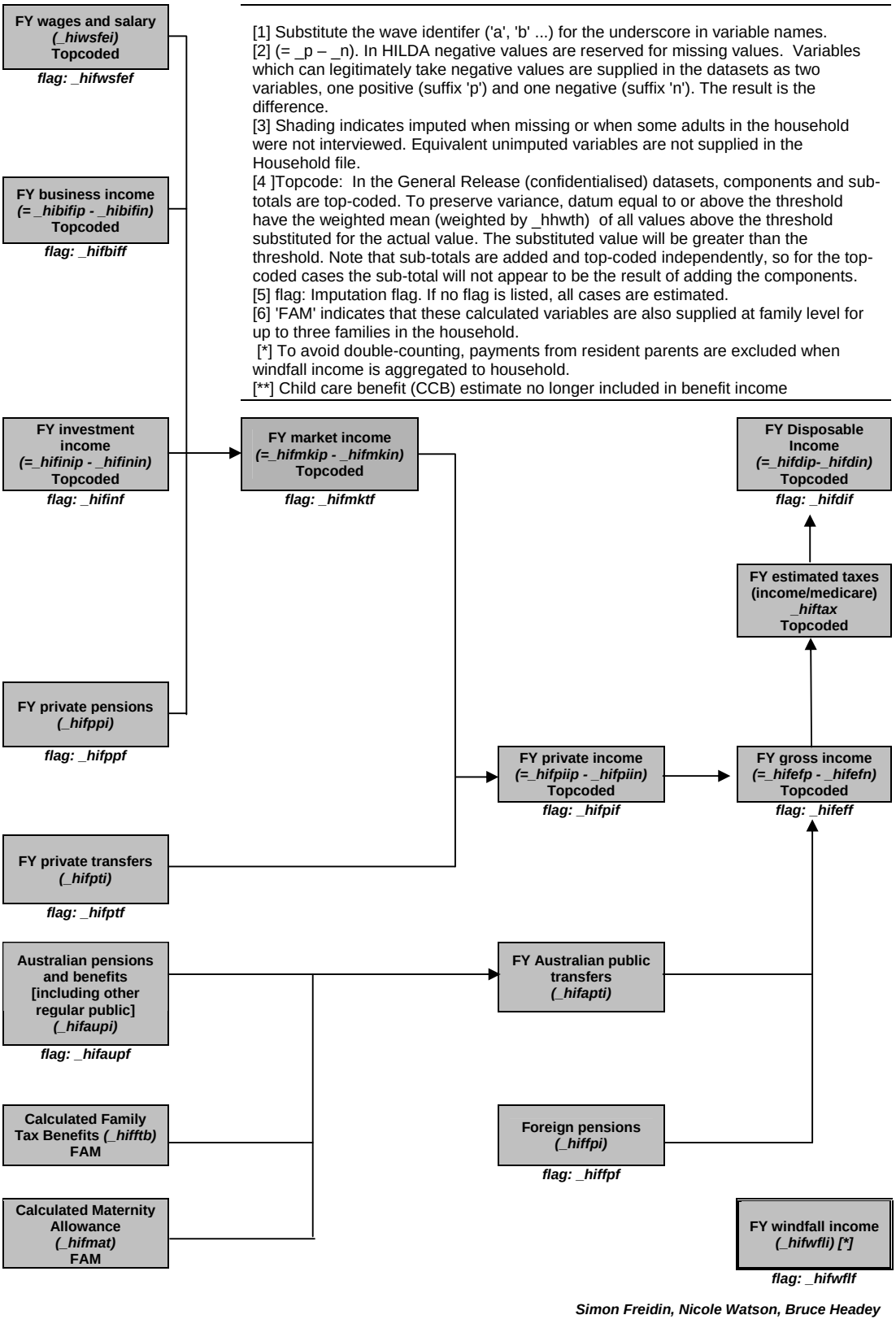


Figure 7: Financial Year Income Model: Enumerated Person

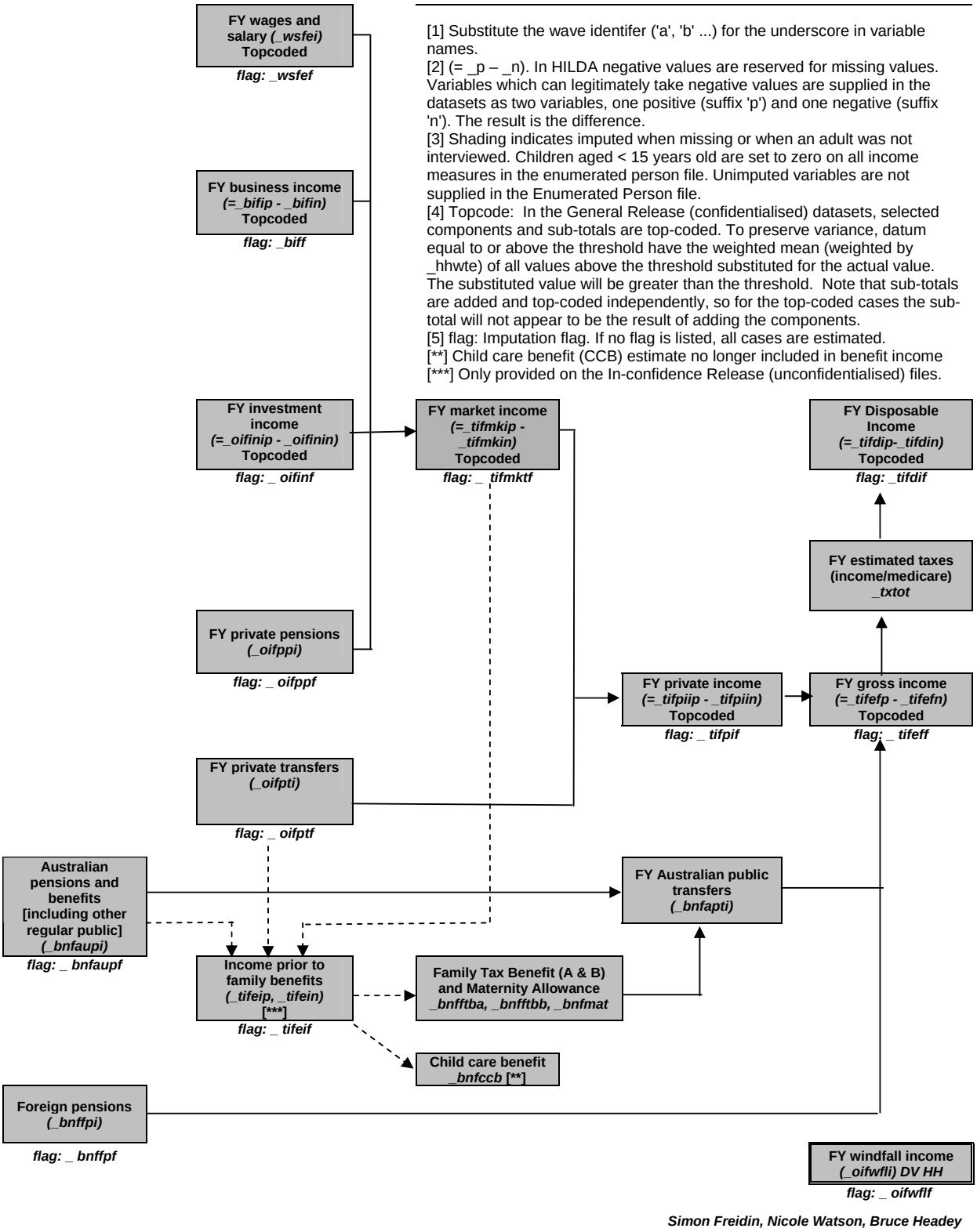


Figure 8: Financial Year Income Model: Responding Person

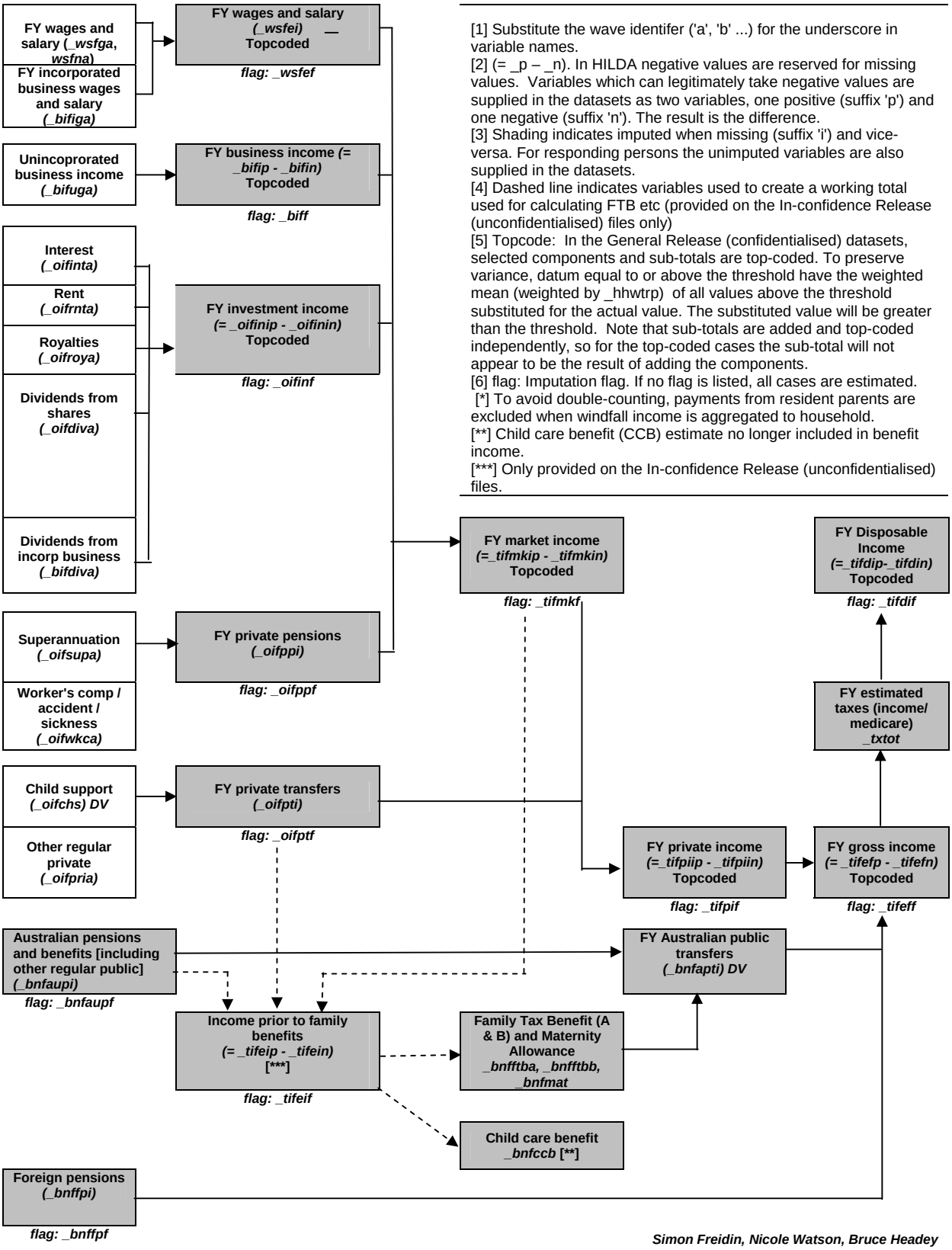
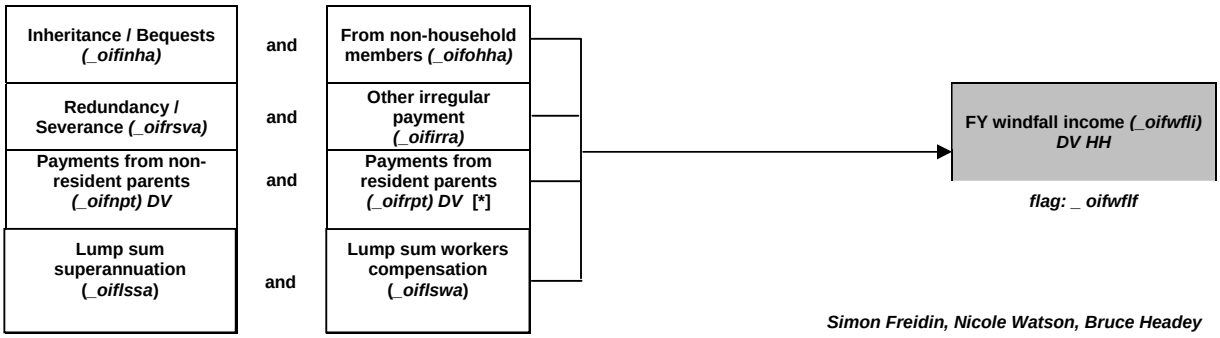


Figure 8: Financial Year Income Model: Responding Person (c'td)



A list of (unimputed) derived income variables are provided in Table 22. 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 chapter 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).

- In wave 2, we calculated current weekly payments for workers compensation, accident or sickness insurance. These questions were only asked in wave 2 (in F10a-e in the PQ).

- 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 22: Derived income variables

Variable	Description
_WSCMG	DV: Main job, current gross wages per week ($). Weighted topcode.
_WSCME	DV: Main job, current gross wages per week ($). Weighted topcode.
_WSCOG	DV: Other jobs, current gross wages per week ($).Weighted topcode.
_WSCOE	DV: Other jobs, current gross wages per week ($).Weighted topcode.
_WSCG	DV: All jobs, current gross wages per week ($).Weighted topcode.
_WSCE	DV: All jobs, current gross wages per week ($).Weighted topcode.
_BNCAUP	DV: Current Australian pensions & benefits per week ($)
_WSFG	DV: Financial year gross wages & salary ($).Weighted topcode.
_WSFE	DV: Financial year gross wages & salary (includes estimated). Weighted topcode.
_BIFP	DV: Financial year business income ($). Positive values. Weighted topcode.
_BIFN	DV: Financial year business income ($). Negative values
_OIFINVP	DV: Financial year investments ($). Positive values. Weighted topcode.
_OIFINVN	DV: Financial year investments ($). Negative values
_OIFPP	DV: Financial year private pensions ($)
_TIFMKTP	DV: Financial year market (factor) income ($) Positive values. Weighted topcode.
_TIFMKTN	DV: Financial year market (factor) income ($) Negative values
_OIFCHS	DV: Financial year other income - child support ($)
_OIFPRI	DV: F33a Sources of payments - other regular private
_OIFPRIA	DV: F33b Estimated value - other regular private ($)
_OIFPT	DV: Financial year private transfers ($)
_TIFPRIP	DV: Financial year private income ($). Positive values. Weighted topcode.
_TIFPRIN	DV: Financial year private income ($). Negative values
_BNFAUP	DV: Financial year Australian pensions and benefits ($) [includes other regular public]
_OIFPUB	DV: F33a Sources of payments received - Other regular public
_OIFPUBA	DV: F33b Estimated value - Other regular public ($)
_BNFFP	DV: Financial year foreign pensions ($)
_OIFIRR	DV: F33a Sources of payments received - Other irregular payment
_OIFIRRA	DV: F33b Estimated value - Other irregular ($)
_OIFNPT	DV: Financial year other income, non-resident parent transfers ($)
_OIFRPT	DV: Financial year other income, resident parent transfers ($) (inc pocket money)
_OIFWFL	DV: Financial year windfall income ($)
BWSCWC	DV: Gross Workers compensation, accident or sickness insurance per week ($)
BWSCWCA	DV: Net Workers compensation, accident or sickness insurance per week ($)
AWSLY	DV: Gross weekly current wages & salary (from all jobs) one year ago ($)

*(see chapter on Confidentialisation for explanation of top-coding)

Imputation Method

Since Release 3.0, 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, therefore, was extended to take into account the 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+.[8] The formulae for the Little and Su method are provided in Appendix 3, together with a worked example.

For some cases, for example 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 2.0 (Watson, 2004a).

Tables 23a and 23b show the number and proportion of missing cases that were imputed using the Little and Su method[9]. For income items, such as business, investments, private pensions and transfers, and windfall income there is a lower proportion of cases imputed using the Little and Su method.

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. Where a Little and Su imputation could be obtained for a component, this was used. 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 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.

The quality of the imputation for non-responding persons within responding households is slightly poorer than the imputation for responding persons for wages and salaries and

---

[8] 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.

[9] For the number and proportion of cases which are missing, see Tables 29a and 29b.

Australian government pensions. The quality of the other income components is poorer for non-responding persons when compared to responding persons. However, the income components are still provided to enable the components to be available at the household level.

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

Table 23a: Number of missing cases imputed using Little and Su method

	Wave				
Variable	1	2	3	4	5
Responding Persons					
Current income					
Wages and salaries	324	253	236	227	201
Benefits	65	62	53	51	44
Financial year income					
Wages and salaries	450	442	368	237	249
Aust govt pensions	62	79	46	64	46
Foreign govt pensions	0	5	0	1	5
Business income	115	146	158	125	117
Interest income	352	426	330	265	274
Dividends and royalties	351	432	345	253	264
Rent income	86	94	103	72	63
Private pensions	27	24	14	17	24
Private transfers	20	35	29	29	38
Windfall income					
Windfall income	11	16	17	18	14
Enumerated Persons (excluding children)					
Current income					
Wages and salaries	492	446	448	468	416
Benefits	119	144	123	127	119
Financial year income					
Wages and salaries	649	656	586	493	460
Aust govt pensions	131	175	130	154	137
Foreign govt pensions	0	8	3	5	9
Business income	132	159	169	150	141
Interest income	405	484	377	328	335
Dividends and royalties	418	501	406	317	331
Rent income	107	113	123	91	87
Private pensions	40	36	23	34	36
Private transfers	24	37	32	31	41
Windfall Income					
Windfall income	29	35	36	49	30

Table 23b: Proportion of missing cases imputed using Little and Su method

	Wave				
Variable	1	2	3	4	5
Responding Persons					
Current income					
Wages and salaries	70.0	81.6	85.8	84.7	79.8
Benefits	47.8	77.5	71.6	73.9	73.3
Financial year income					
Wages and salaries	67.6	80.4	84.8	81.4	68.8
Aust govt pensions	63.9	83.2	82.1	80.0	79.3
Foreign govt pensions	0.0	83.3	0.0	100.0	100.0
Business income	28.5	39.9	44.6	51.7	43.3
Interest income	53.3	71.5	77.8	80.3	77.2
Dividends and royalties	60.1	82.9	85.8	86.9	80.5
Rent income	35.8	49.7	56.9	50.0	43.4
Private pensions	45.8	58.5	48.3	48.6	54.5
Private transfers	62.5	39.3	40.3	48.3	35.5
Windfall income					
Windfall income	34.4	51.6	43.6	58.1	56.0
Enumerated Persons (excluding children)					
Current income					
Wages and salaries	30.4	34.6	39.0	39.7	39.1
Benefits	9.2	13.6	13.0	12.9	13.6
Financial year income					
Wages and salaries	35.6	42.9	44.8	41.0	39.1
Aust govt pensions	10.4	16.3	14.0	15.5	15.7
Foreign govt pensions	0.0	0.8	0.3	0.5	1.1
Business income	8.5	11.8	13.8	13.0	13.0
Interest income	22.3	30.8	29.0	26.4	28.7
Dividends and royalties	24.0	33.4	31.8	26.4	29.0
Rent income	7.7	9.7	11.7	8.6	9.1
Private pensions	3.3	3.5	2.5	3.6	4.2
Private transfers	2.0	3.5	3.4	3.2	4.5
Windfall income					
Windfall income	2.4	3.5	3.9	5.2	3.6

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 24. 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.

Table 24: Person imputed income variables

	Pre-imputed	*Post-imputed*	*Flag*
Responding person file			
Current income			
Wages and salaries	_wsce	_wscei	_wscef
Benefits	_bncaup	_bncaupi	_bncaupf
Financial year income a			
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 [b]	Not provided	_tifefn, _tifefp	_tifeff
Windfall income	_oifwfl	_oifwfli	_oifwflf
Enumerated person file			
Current income			
Wages and salaries	Not provided	_wscei	_wscef
Benefits	Not provided	_bncaupi	_bncaupf
Financial year income a			
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 [b]	Not provided	_tifefn, _tifefp	_tifeff
Windfall income	Not provided	_oifwfli	_oifwflf
Household file			
Current income			
Wages and salaries	Not provided	_hiwscei	_hifwscef
Benefits	Not provided	_hicaupi	_hicaupf
Financial year income c			
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

Notes:

a. Several sub-totals also provided on the datasets: Australian public transfers (_bnfapti including FTB and maternity allowance, use flag _bnfaupf), market income (_tifmkin, _tifmkip, with flag _tifmktf) and private income (_tifpiin, _tifpiip, with flag _tifpif). On the In-confidence Release (unconfidentialised) files datasets there is also a sub-total for income prior to FTB, MA and CCB (pre-imputation: _tifen, _tifep; post-imputation: _tifein, tifeip; and imputation flag __tifeif).

b. The following variables use total person financial year income (_tifefn, _tifefp) in their calculations: income tax (_txinc), medicare (_txmed), total taxes (_txtot). Use _tifeff as imputation flag for these variables. Also disposable income is provided (_tifdin, _tifdip, with flag _tifdif).

c. Household level information for income components are now provided on the dataset (by summing imputed person level information): household financial year gross income, family tax benefit (_bnftaf1 to _bnftaf3 for part A and _bnftbf1 to _bnftbf3 for part B for families 1 to 3 and household total _hifftb), maternity allowance (_bnfmat and household total _hifmat), child care benefit (_bnccbf1 to _bnccbf3 for families 1 to 3 and household total _hifccb), household total taxes (_hiftax, use flag _hifeff),and household disposable income (_hifdin, _hifdip, with flag _hifdif). Several sub-totals also provided on the dataset: Australian public transfers (_hifapti including family tax benefit and maternity allowance, use flag _hifaupf), market income (_hifmkin, _hifmkip, with flag _hifmktf), and private income (_hifpiin, _hifpiip, with flag _hifpif).

Wealth Variables and Wealth Imputation (Wave 2 Special Topic)

Wealth Model

In wave 2 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;[10]
- Collectables and vehicles.

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

- Bank accounts and credit card debt;
- Superannuation;
- HECS and other personal debts.

Figure 9 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. As the imputation for non-respondents was only conducted at the total assets and debts level, the household-level components in the broken boxes are just the sum of ***responding*** persons only. Please be aware of this when using these variables.

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

Figure 9: Wealth Model Diagram, Wave 2 Household-level

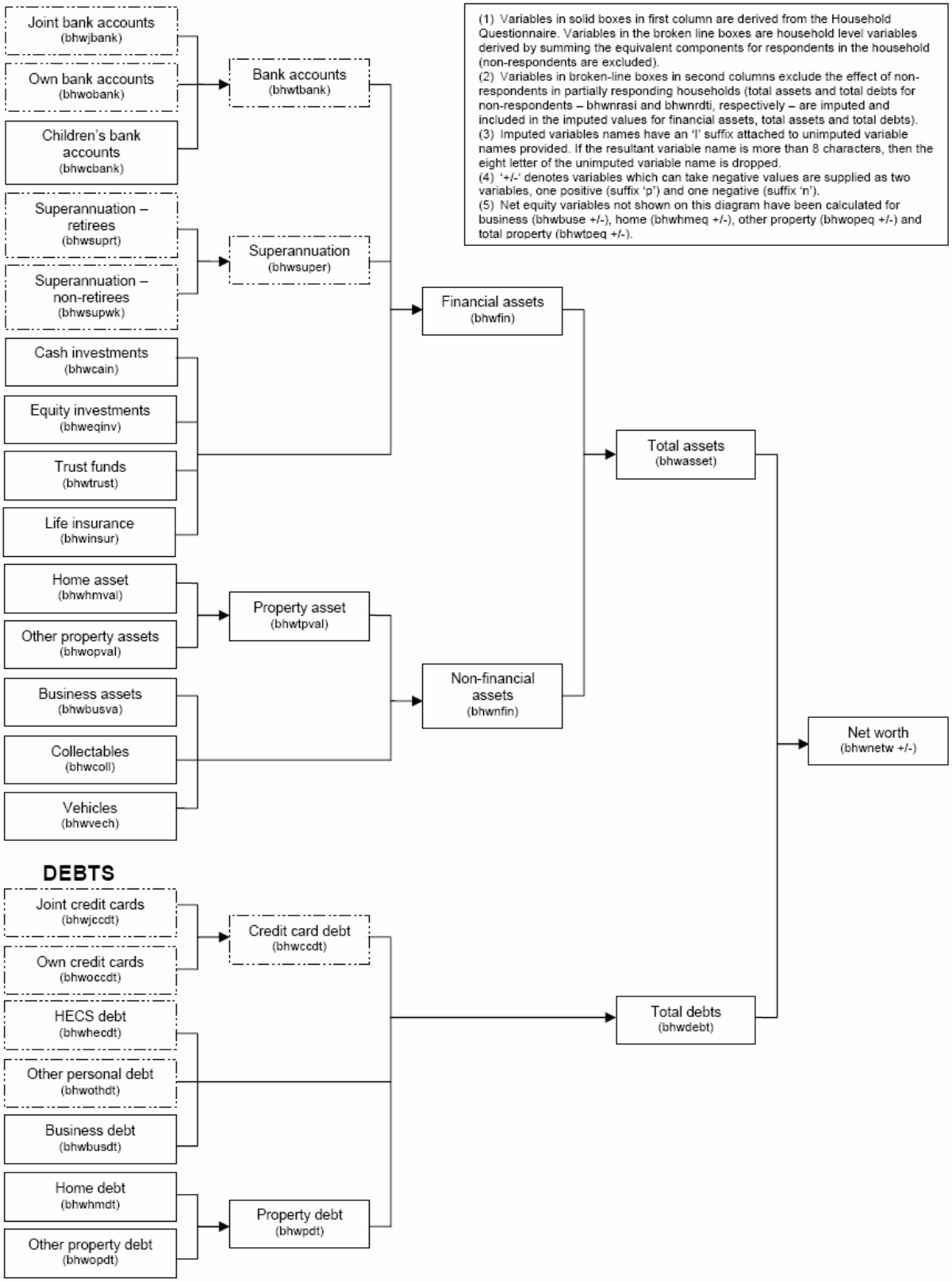


Table 25 shows the household-level and person-level wealth variables that are not shown in Figure 9.

Table 25: Derived wealth variables, person-level

Variable	Description
BHWHMHL	DV: Household wealth: Home: Home loans ($)
BHWHMOL	DV: Household wealth: Home: Other loans ($)
BHWHMEQL	DV: Household wealth: Home: Equity loan ($)
BPWOBANK	DV: Own bank accounts ($)
BPWJBANK	DV: Joint bank accounts ($)R
BPWSUPR	DV: superannuation, retirees ($)
BPWSUPWK	DV: superannuation, not retired ($)
BPWOCCDT	DV: Own credit card debt ($)
BPWJCCDT	DV: Joint credit card debt ($)
BPWHECDT	DV: HECS debt ($)
BPWOTHDT	DV: Other Debt: Car loan/Loans/Hire purchase/overdraft ($)

Imputation Method

The wealth imputation is unchanged from Release 2.0. The wealth imputation uses a nearest neighbour regression method. 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. Where we knew the recipient's 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 further details of the imputation method, see Watson (2004a).

As we only have one wave of wealth data, we are unable to use the Little and Su method currently being used for the income imputation.

Imputed Wealth Variables

Table 26 outlines the imputed wealth variables included on the wave 2 files.

Note that in addition to total household assets and debts, several sub-totals and totals are also provided on the household file (by summing imputed components). These are:

- business equity,
- investment equity,
- home equity,
- other property equity,
- total property equity,
- total credit card debt,
- total superannuation,
- total bank accounts,
- total property debt,
- total property value,

- household financial assets,
- household non-financial assets,
- net worth, and
- total assets and debts of non-respondents in responding households.

These subtotals (except for household financial assets) exclude wealth imputations on non-responding persons, for whom only totals for debt and asset were imputed.

Table 26: Imputed wealth variables

	Pre-imputed	*Post-imputed*	*Flag*
Responding person file			
Assets			
Joint bank accounts	bpwjbank	bpwjbani	bpwjbanf
Own bank accounts	bpwobank	bpwobani	bpwobanf
Superannuation – retirees	bpwsupr	bpwsupri	bpwsuprf
Superannuation – non-retirees	bpwsupwk	bpwsupwi	bpwsupwf
Debts			
HECS debt	bpwhecdt	bpwhecdi	bpwhecdf
Joint credit cards	bpwjccdt	bpwjccdi	bpwjccdf
Own credit cards	bpwoccdt	bpwoccdi	bpwoccdf
Other personal debt	bpwothdt	bpwothdi	bpwothdf
Enumerated person file			
Total person assets	Not provided	bpwassei	bpwassef
Total person debts	Not provided	bpwdebti	bpwdebtf
Household file			
Assets			
Joint bank accounts[1]	bhwjbank	bhwjbani	bhwjbanf
Own bank accounts[1]	bhwobank	bhwobani	bhwobanf
Children's bank accounts	bhwcbank	bhwcbani	bhwcbanf
Superannuation – retirees[1]	bhwsupr	bhwsupri	bhwsuprf
Superannuation – non-retirees[1]	bhwsuhwk	bhwsuhwi	bhwsuhwf
Business assets	bhwbusva	bhwbusvi	bhwbusvf
Cash investment	bhwcain	bhwcaini	bhwcainf
Equity investment	bhweqinv	bhweqini	bhweqinf
Collectables	bhwcoll	bhwcolli	bhwcollf
Home asset	bhwhmval	bhwhmvai	bhwhmvaf
Other property assets	bhwopval	bhwopvai	bhwopvaf
Life insurance	bhwinsur	bhwinsui	bhwinsuf
Trust funds	bhwtrust	bhwtrusi	bhwtrusf
Vehicles value	bhwvech	bhwvechi	bhwvechf
Total household assets	bhwasset	bhwassei	bhwassef

Table 26: (c'td)

	Pre-imputed	*Post-imputed*	*Flag*
Debts			
HECS debt[1]	bhwhecdt	bhwhecdi	bhwhecdf
Joint credit cards[1]	bhwjccdt	bhwjccdi	bhwjccdf
Own credit cards[1]	bhwoccdt	bhwoccdi	bhwoccdf
Other personal debt[1]	bhwothdt	bhwothdi	bhwothdf
Business debt	bhwbusdt	bhwbusdi	bhwbusdf
Home debt	bhwhmdt	bhwhmdti	bhwhmdtf
Other property debt	bhwopdt	bhwopdti	bhwopdtf
Total household debts	bhwdebt	bhwdebti	bhwdebtf

1. Care should be taken when using these variables at the household level. These household variables are calculated as the sum of the equivalent wealth component for responding persons only. If non-responding adults exist in these household, no attempt to apportion their imputed total assets and debts to the person level components has been made, resulting in an underestimate of these components at the household level.

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 4.0 and these changes have been carried over to Release 5.0. 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.[11]

[11] 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.

- 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.[12]

The enumerated person benchmarks have been revised to:

- Sex by broad age;
- State by part of State;
- Labour force status[13]; and
- Marital status.[14]

The responding person benchmarks have been revised to:

- 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.[15]
- 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.[16] The remaining benchmarks come from the ABS Labour Force Survey.

From Release 5.0 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

[12] Prior to Release 4.0, 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).
[13] The Wave 2 benchmarks do not include labour force status as this was not collected on the Household Form that wave.
[14] Prior to Release 4.0, 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 in Release 4.0.
[15] No benchmarks for the responding person weights have been dropped, but the marital status and occupation benchmarks have been included.
[16] In earlier Releases, only household estimates based on the 1996 Census were available.

run by the ABS.[17] 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 waves 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.

The benchmarks used in Release 4.0 and later have been amended as described above. Other aspects of the weighting process for wave 2 onwards are detailed in Watson (2004b).[18]

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 4.0 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.[19]

The longitudinal responding person benchmarks have been revised to:

- 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.[20]

[17] Prior to Release 5.0 only the sparsely settled parts of the Northern Territory were excluded.
[18] While this paper is written in relation to the wave 2 weighting, the process in later waves follows the same methodology.
[19] Prior to Release 4.0, 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 in Release 4.0.

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

Replicate Weights

Replicate weights have been provided for users of the GREGWT macro (see page 79), or the 'svy jackknife' commands in STATA. Thirty replicate groups have been provided.

Weights Provided on the Data Files

Table 27 provides a list of the weights provided on the data files together with a description of those weights. We have adopted the convention of adding the longitudinal weights only to the most recent wave undertaken to which they relate.

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 27: 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 4 is approximately 7.8 million.
	_hhwths	This is the cross-section household population weight rescaled to the sum of the sample size for the relevant wave (i.e. for wave 4, 6987 responding households). Use this weight when the statistical package requires the weights to sum to the sample size.
	_hhwte01 to _hhwte14	The enumerated person weights are provided on both the household file and the enumerated person file. See description below.
	_rwh01to _rwh30	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 4 is 19.7 million.
	blnwte	This longitudinal enumerated person weight is the longitudinal population weight for all people who were enumerated (i.e. in responding households) in both waves 1 and 2. This weight applies to children, non-respondents, intermittent respondents, and full respondents in responding households.
	clnwte	This longitudinal enumerated person weight is the longitudinal population weight for all people who were enumerated (i.e. in responding households) in each of waves 1, 2 and 3. This weight applies to children, non-respondents, intermittent respondents, and full respondents in responding households.
	dlnwte	This longitudinal enumerated person weight is the longitudinal population weight for all people who were enumerated (i.e. in responding households) in each of waves 1 through 4. This weight applies to children, non-respondents, intermittent respondents, and full respondents in responding households.

[20] No benchmarks for the longitudinal responding person weights have been dropped, but the marital status and occupation benchmarks have been included.

Table 27: (c'td)

File	*Weights*	*Description*
	elnwte	This longitudinal enumerated person weight is the longitudinal population weight for all people who were enumerated (i.e. in responding households) in each of waves 1 through 5. This weight applies to children, non-respondents, intermittent respondents, and full respondents in responding households.
	_rwe01 to _rwe30	Cross-section enumerated person population replicate weights.
	_rwlne01 to _rwlne30	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 4 is 15.8 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. 12,408 in wave 4). Use this weight when the statistical package requires the sum of the weights to be the sample size.
	blnwtrp	This longitudinal responding person weight is the longitudinal population weight for all people responding (i.e. provided an interview) in both of waves 1 and 2.
	clnwtrp	This longitudinal responding person weight is the longitudinal population weight for all people responding (i.e. provided an interview) in each of waves 1, 2 and 3.
	dlnwtrp	This longitudinal responding person weight is the longitudinal population weight for all people responding (i.e. provided an interview) in each of waves 1 through 4.
	elnwtrp	This longitudinal responding person weight is the longitudinal population weight for all people responding (i.e. provided an interview) in each of waves 1 through 5.
	_rwrp01 to _rwrp30	Cross-sectional responding person population replicate weights
	_rwlnr01 to _rwlnr30	Longitudinal responding person population replicate weights.

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 between waves 1 and 4, then you would use the longitudinal weights for waves 1 through 4.

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

- What proportion of households rent in 2004? We would use the cross-section household weight for wave 4 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.

- How many people have moved out of the poorest 10 per cent of households between 2001 and 2004? We might define the 'poorest' 10 per cent of households as having the lowest equivalised household incomes in each wave. We could then calculated how many people move out of the poorest decile between waves 1 and 4 by summing the longitudinal enumerated person weight (*dlnwte*) for those people.

- What proportion of people have changed their employment status between 2001 and 2004? This question can only be answered by considering the responding persons in both waves. We would use the longitudinal responding person weight (*dlnwtrp*) and construct a weighted cross-tabulation of the employment status of respondents in wave 1 against the employment status of respondents in wave 4.

It has been assumed that when using the longitudinal weights, researchers will be using a balanced panel from wave 1. If you wanted to just consider the movements between waves 2 and 3, then you could restrict the sample to the balanced wave 1, 2 and 3 panel and use the longitudinal weight provided. This will mean that some people who responded in waves 2 and 3 but not wave 1 will be excluded, so the estimates will be a little less efficient than if we constructed weights for your specific circumstances.

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

Calculating Standard Errors

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

- clustered – 488 areas were originally selected from which households were chosen and people are clustered within households;
- stratified – the 488 areas were selected from a frame of areas stratified by State and part of State; and
- unequally weighted – the households and individuals have unequal weights due to some irregularities in the selection of the sample in wave 1 and the non-random non-response in wave 1 and the non-random attrition in 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 the recently released complex survey commands in SPSS (available in version 12).
- 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.
- Use of GREGWT macro in SAS – Some users within FaCSIA, 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.

To assist you in the calculation of appropriate standard errors, the wave 1 area, replicate group, and stratification variables have been included on the master file. These are listed in Table 28. Any new entrants to the household are assigned to the same sample design information as the permanent sample member.

Table 28: Sample design variables

Variable	Description	Design element
AHHRAID	DV: randomised area id	Cluster
AHHMSR	DV: Major Statistical Region	Stratification
AHHREPGP	DV: Replicate group	Replicate group

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 as you use the variables.

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.[21]

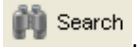
While frequencies of the variables have been provided, it is expected that you might only refer to these files for some simple queries with the variable name in mind (for example, how many employed people do we have in the sample, or what are the codes used for question R3).

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.

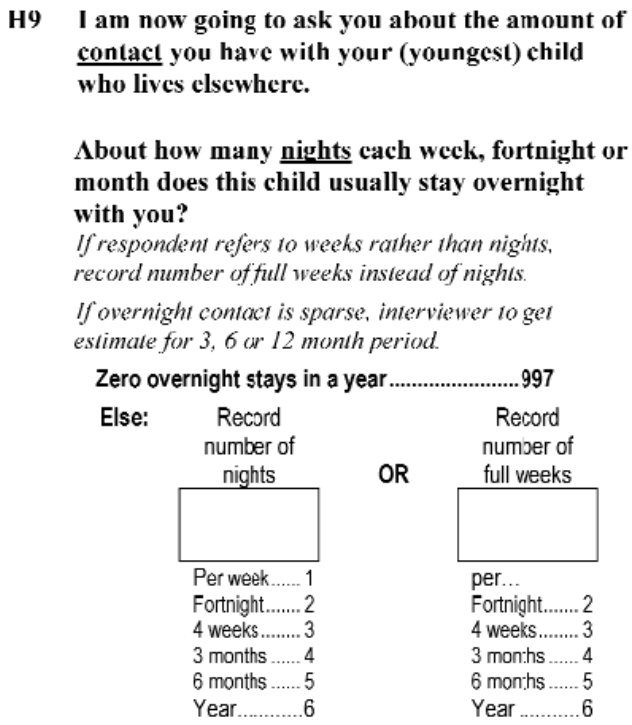
Marked-Up Questionnaires

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 10 for an example.

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

.

Figure 10: 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	OR	Record number of full weeks
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 11 shows the derived variable associated with the variables listed on the questionnaire in Figure 10.

Figure 11: 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 12 below:

Figure 12: 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 250 page document. See Figure 13 below:

Figure 13: 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 14, 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 14: Example of the cross-wave variable listing

			Wave			
File	**Variable**	**Data Item**	**1**	**2**	**3**	**4**
RP	_NCNGTH	Youngest non-resident child overnight stays - answered nights or weeks	PQ H9	PQ G9	PQ G9b	PQ G9b
RP	_NCNGTN	Youngest non-resident child overnight stays - number of nights	PQ H9	PQ G9	PQ G9b	PQ G9b
RP	_NCNGTNP	Youngest non-resident child overnight stays - nights - period	PQ H9	PQ G9	PQ G9b	PQ G9b
RP	_NCNGTW	Youngest non-resident child overnight stays - number of weeks	PQ H9	PQ G9	PQ G9b	PQ G9b
RP	_NCNGTWP	Youngest non-resident child overnight stays - weeks - period	PQ H9	PQ G9	PQ G9b	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 15 provides an example of the listing.

Figure 15: 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		

DATA QUALITY ISSUES

For issues of data quality, refer to the wave 1 and 2 data quality technical papers (Watson and Wooden, 2002a; Watson and Wooden, 2004a). These can be found on the HILDA website. As further research is carried out on a variety of data quality issues, this section will be added to.

Some 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 number and proportion of cases with missing income data are provided in Tables 29a and 29b. 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.

Tables 30a and 30b show how much of the mean income was imputed for each wave and income component. For responding people with item non-response, 5.5 percent of total financial year income was imputed in wave 5, compared to 7.8 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 14.5 percent in wave 5.

This shows that while approximately one in ten responding persons are missing some component of financial year income in wave 5, only one twentieth of the mean income comes from imputed values and the remainder is from reported values. At the household level, one in six households is missing some component of financial year income and one seventh of the mean income is from imputed values.

Table 29a: Number of cases with missing income data, waves 1 – 5

	Wave				
Variable	*1*	*2*	*3*	*4*	*5*
Responding Persons (non-zero cases only)					
Current income					
Wages and salaries	462	310	275	268	252
Benefits	136	80	74	67	60
Financial year income					
Wages and salaries	666	550	434	291	362
Aust govt pensions	97	95	56	80	58
Foreign govt pensions	1	6	0	1	5
Business income	404	366	354	242	268
Investments					
Interest income	661	596	424	330	355
Dividends and royalties	584	521	401	291	328
Rent income	240	189	181	144	145
Private pensions	59	40	29	35	43
Private transfers	32	88	70	60	104
Total FY income	2,072	1,847	1,463	1,138	1,300
Windfall income					
Windfall income	27	31	39	31	25
Enumerated Persons (zero and non-zero cases, excluding children)					
Current income					
Wages and salaries	1,621	1,288	1,149	1,180	1,065
Benefits	1,294	1,058	948	981	873
Financial year income					
Wages and salaries	1,824	1,528	1,308	1,203	1,175
Aust govt pensions	1,255	1,073	930	992	871
Foreign govt pensions	1,159	984	874	913	818
Business income	1,562	1,344	1,228	1,154	1,083
Investments					
Interest income	1,819	1,574	1,298	1,242	1,168
Dividends and Royalties	1,742	1,499	1,276	1,203	1,141
Rent income	1,398	1,167	1,055	1,056	958
Private pensions	1,217	1,019	903	947	857
Private transfers	1,190	1,067	946	972	920
Total FY income	3,230	2,825	2,338	2,051	2,115
Windfall income					
Windfall income	1,190	1,009	913	944	838
Households (zero and non-zero cases)					
Current income					
Wages and salaries	1,181	967	864	911	839
Benefits	928	769	699	744	684
Financial year income					
Wages and salaries	1,306	1,137	979	908	913
Aust govt pensions	894	785	683	751	678
Foreign govt pensions	813	707	633	684	634
Business income	1,103	966	898	861	832
Investments	1,628	1,442	1,197	1,140	1,120
Private pensions	867	740	659	715	668
Private transfers	841	783	697	739	716
Total FY income	2,256	2,032	1,705	1,531	1,590
Windfall income					
Windfall income	838	723	662	710	649

Table 29b: Proportion of cases with missing income data, waves 1 – 5

	Wave				
Variable	*1*	*2*	*3*	*4*	*5*
Responding Persons (non-zero cases only)					
Current income					
Wages and salaries	6.0	4.2	3.8	3.8	3.4
Benefits	3.2	2.0	2.0	1.8	1.6
Financial year income					
Wages and salaries	7.9	6.9	5.5	3.8	4.5
Aust govt pensions	2.1	2.1	1.3	2.0	1.4
Foreign govt pensions	0.5	2.7	0.0	0.5	2.4
Business income	29.1	28.6	27.4	19.4	21.6
Investments					
Interest income	19.5	18.6	13.9	11.0	11.3
Dividends and royalties	14.6	14.5	11.9	9.2	10.2
Rent income	20.3	15.3	14.9	12.4	11.6
Private pensions	6.3	4.6	3.2	4.1	4.8
Private transfers	8.0	22.9	15.5	14.4	20.7
Total FY income	15.7	14.9	12.1	9.7	10.7
Windfall income					
Windfall income	3.4	2.8	3.2	2.7	2.1
Enumerated Persons (zero and non-zero cases, excluding children					
Current income					
Wages and salaries	10.7	9.2	8.4	8.9	7.8
Benefits	8.6	7.5	7.0	7.4	6.4
Financial year income					
Wages and salaries	12.1	10.9	9.6	9.0	8.7
Aust govt pensions	8.3	7.7	6.8	7.4	6.4
Foreign govt pensions	7.7	7.0	6.4	6.9	6.0
Business income	10.3	9.6	9.0	8.7	8.0
Investments					
Interest income	12.0	11.2	9.5	9.3	8.6
Dividends and Royalties	11.5	10.7	9.4	9.0	8.4
Rent income	9.2	8.3	7.8	7.9	7.1
Private pensions	8.0	7.3	6.6	7.1	6.3
Private transfers	7.9	7.6	7.0	7.3	6.8
Total FY income	21.4	20.2	17.2	15.4	15.6
Windfall income					
Windfall income	7.9	7.2	6.7	7.1	6.2
Households (zero and non-zero cases)					
Current income					
Wages and salaries	15.4	13.3	12.2	13.0	11.8
Benefits	12.1	10.6	9.9	10.6	9.6
Financial year income					
Wages and salaries	17.0	15.7	13.8	13.0	12.8
Aust govt pensions	11.6	10.8	9.6	10.7	9.5
Foreign govt pensions	10.6	9.8	8.9	9.8	8.9
Business income	14.4	13.3	12.7	12.3	11.7
Investments	21.2	19.9	16.9	16.3	15.7
Private pensions	11.3	10.2	9.3	10.2	9.4
Private transfers	10.9	10.8	9.8	10.6	10.0
Total FY income	29.4	28.0	24.0	21.9	22.3
Windfall income					
Windfall income	10.9	10.0	9.3	10.2	9.1

Table 30a: Mean income ($) (including imputed values), waves 1 – 5 (unweighted)

	Wave				
Variable	*1*	*2*	*3*	*4*	*5*
Responding Persons					
Current income (per week)					
Wages and salaries	390	397	412	432	464
Benefits	50	52	54	56	56
Financial year income					
Wages and salaries	20,587	21,067	21,793	22,783	24,544
Aust govt pensions	2,441	2,691	2,741	2,842	2,806
Foreign govt pensions	63	89	73	61	80
Business income	1,527	1,877	1,603	1,807	1,969
Investments	1,604	1,698	1,695	1,988	2,497
Private pensions	964	1,098	1,184	1,195	1,270
Private transfers	122	130	151	164	165
Total FY income	27,307	28,650	29,240	30,839	33,330
Windfall income					
Windfall income	453	1,756	1,936	1,483	2,143
Enumerated Persons (excluding children)					
Current income					
Wages and salaries	398	411	424	438	468
Benefits	51	53	56	57	57
Financial year income					
Wages and salaries	21,004	21,796	22,435	23,267	24,754
Aust govt pensions	2,501	2,727	2,836	2,862	2,875
Foreign govt pensions	65	94	75	64	75
Business income	1,571	1,831	1,548	1,907	1,917
Investments	1,648	1,707	1,740	2,080	2,459
Private pensions	985	1,119	1,209	1,157	1,290
Private transfers	129	133	153	161	180
Total FY income	27,903	29,408	29,996	31,497	33,550
Windfall income					
Windfall income	490	1,859	1,944	1,768	2,282
Households					
Current income (per week)					
Wages and salaries	800	825	851	881	941
Benefits	103	106	112	115	115
Financial year income					
Wages and salaries	42,217	43,690	45,032	46,831	49,773
Aust govt pensions	5,028	5,467	5,693	5,760	5,780
Foreign govt pensions	130	189	150	130	151
Business income	3,158	3,670	3,107	3,839	3,855
Investments	3,313	3,422	3,493	4,186	4,945
Private pensions	1,981	2,242	2,427	2,328	2,593
Private transfers	260	266	306	324	361
Total FY income	56,085	58,946	60,209	63,397	67,458
Windfall income					
Windfall income	986	3,725	3,901	3,558	4,589

Table 30b: Proportion of mean income ($) imputed, waves 1 – 5 (unweighted)

	Wave				
Variable	*1*	*2*	*3*	*4*	*5*
Responding Persons					
Current income (per week)					
Wages and salaries	5.1	3.6	3.2	3.0	3.0
Benefits	2.7	2.0	1.6	1.9	1.5
Financial year income					
Wages and salaries	5.9	4.7	3.9	3.0	3.3
Aust govt pensions	1.8	2.0	0.8	1.5	1.6
Foreign govt pensions	0.4	3.0	.	0.4	1.8
Business income	27.5	30.0	28.8	18.3	22.8
Investments	24.6	30.7	21.5	19.5	17.0
Private pensions	5.2	3.9	3.2	4.1	5.3
Private transfers	6.6	25.2	16.7	10.6	18.3
Total FY income	7.8	7.7	6.0	4.9	5.5
Windfall income					
Windfall income	2.1	1.9	2.2	2.5	9.6
Enumerated Persons (excluding children)					
Current income					
Wages and salaries	14.7	14.3	13.7	13.5	11.9
Benefits	11.5	12.8	14.0	12.5	13.5
Financial year income					
Wages and salaries	14.8	15.2	14.5	13.7	12.3
Aust govt pensions	10.7	12.9	13.3	11.7	13.5
Foreign govt pensions	7.5	18.3	9.8	16.5	10.2
Business income	33.3	35.6	33.3	29.1	27.8
Investments	30.6	36.5	28.9	28.0	25.2
Private pensions	13.6	16.0	15.3	12.0	17.4
Private transfers	15.0	30.3	25.1	16.3	28.1
Total FY income	16.3	17.6	16.2	15.3	14.5
Windfall income					
Windfall income	18.8	17.6	11.3	27.8	25.4
Households					
Current income (per week)					
Wages and salaries	14.7	14.3	13.7	13.5	11.9
Benefits	11.5	12.8	14.0	12.5	13.5
Financial year income					
Wages and salaries	14.8	15.2	14.5	13.7	12.3
Aust govt pensions	10.7	12.9	13.3	11.7	13.5
Foreign govt pensions	7.5	18.3	9.8	16.5	10.2
Business income	33.3	35.6	33.3	29.1	27.8
Investments	30.6	36.5	28.9	28.0	25.2
Private pensions	13.6	16.0	15.3	12.0	17.4
Private transfers	15.0	30.3	25.1	16.3	28.1
Total FY income	16.3	17.6	16.2	15.3	14.5
Windfall income					
Windfall income	18.8	17.6	11.3	27.8	25.4

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.[22]

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.[23]

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.

---

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

23 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.[24] 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'[25]. 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

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

occupancy rates of the area. The selections were made after all dwellings within each of 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.[26] 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.

[26] 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 ACNielsen, 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) conducted in an office environment.
- Dress Rehearsal with a sample of Sydney and rural NSW households (approximately 180 for waves 1 -4 and 580 for wave 5).

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

Table 31: Average time (minutes) taken to complete questionnaires, waves 1 to 5

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

Notes:
a. Approximate minutes as not timed

Interviewers

The number of interviewers used for each wave of the fieldwork is given in Table 32, 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 32: Number of interviewers and percentage of new interviewers each wave

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

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 33 provides the proportion of people interviewed by telephone in each waves. 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 6.5 per cent in wave 5.

Table 33: Proportion of respondents interviewed by telephone (%)

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

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 34 provides details of the fieldwork dates.

Table 34: Fieldwork dates and proportion of interviews post Dec 2005

	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%

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/respondent.html.

Respondent Incentives

In wave 5 the cash incentive structure changed for the first time. 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).

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 this new 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.[27] 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 stages.[28] All households were issued into the field for the first stage, and where all the interviews had not been completed, they were reissued into the field in the next stage.[29] 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 stage if found quickly or more generally into a later stage. The third stage 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 stage (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.[30]

---

[27] 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.

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

[29] 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 stage of the fieldwork. When following up a household, the interviewer had a total of five calls to finalise the household.

[30] 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 35 and Table 36. Table 35 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 35: 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 36. 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 36: 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

Tables 37, 38, 39 and 40 show the household outcomes for waves 2, 3, 4 and 5 respectively. The household response rate (including fully and partially responding households) for wave 2, 3, 4 and 5 are 87.0 percent, 81.8 percent, 78.6 percent and 78.0 percent.

It is also constructive to consider the wave 4 and 5 outcome in two groups – those that responded in the previous wave and those that didn't. The wave 3, 4 and 5 household response rate for those households responding in the previous wave was 90.6 percent, 92.3 percent and 94.6 percent, respectively, which compares favorably to the wave 2 household response rate of 87.0 percent.[31]

In Tables 41 and 42 we report a summary of the person-level response in waves 2, 3, 4 and 5. 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 and 10,391 were re-interviewed in wave 5. The number interviewed in all five waves is 9311.

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 rate for waves 2, 3, 4 and 5 were 13.2 percent, 9.6 per cent, 8.4 percent and 5.6 percent, respectively.

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). Although the attrition rate in wave 4 of the HILDA Survey is higher than the BHPS by over two percentage points, in wave 5 the HILDA Survey attrition rate is now just below that of the BHPS at 0.8 percentage points.

[31] 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.

Table 37: 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 38: Wave 3 Household Outcomes

Sample Outcome	*All households*		*Wave 3 responding HH*		*Wave 3 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 39: Wave 4 Household Outcomes

Sample Outcome	*All households*		*Wave 4 responding HH*		*Wave 4 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 40: Wave 5 Household Outcomes

Sample Outcome	*All households*		*Wave 5 responding HH*		*Wave 5 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 and 4.
b Includes 370 households that were identified as untraceable in an earlier wave.

Table 41: Wave 2 - 5 Person Outcomes Against Wave 1 Person Outcomes

	Wave 1			*New Entrant* *Wave*				
	Resp.	*Non-resp*	*Child*	2	3	4	5	*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	2465	886	165	154	157	-	-	3827
Out-of-scope	314	49	37	334	-	-	-	734
Child	-	-	4123	287	363	-	-	4773
Wave 4								
Respondent	10565	209	666	284	287	397	-	12408
Non-respondent	2899	879	289	138	113	167	-	4485
Out-of-scope	505	70	51	442	317	-	-	1383
Child	-	-	3781	267	300	332	-	4680
Wave 5								
Respondent	10391	238	909	261	230	247	483	12759
Non-respondent	2959	844	354	98	78	81	121	4535
Out-of-scope	619	76	52	513	422	279	-	1961
Child	-	-	3472	259	287	289	362	4669
Total	13969	1158	4787	1131	1017	896	966	23904

Table 42: Response rates (%) for the HILDA Survey, waves 2, 3, 4 and 5 compared

	Wave			
	2	*3*	*4*	*5*
Previous wave respondents	86.8	90.4	91.6	94.4
Previous wave non-respondents	19.7	17.6	12.7	14.7
Previous wave children turning 15 years	80.4	71.3	70.7	74.6
New entrants to the wave	73.3	76.1	70.4	81.7

Table 43 shows the response rates for the Self Completion questionnaire, calculated as the percentage of respondents to which an SCQ could be matched.

Table 43: 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

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 44 provides figures on the proportion of Wave 1 respondents who were re-interviewed in Wave 5 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 attition over the five waves is particularly marked with respect to age, country of birth, labour force and occupation.

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 44: Proportion of Wave 1 Respondents Re-interviewed by Selected Sample Characteristics

Wave 1 characteristics	In all waves (%)	In wave 5 (%)	Wave 1 characteristics	In all waves (%)	In wave 5 (%)
Area			Indigenous status		
Sydney	65.9	75.5	Indigenous	56.0	73.6
Rest of NSW	73.5	81.1	Non-indigenous	70.5	77.9
Melbourne	67.8	76.4	Education attainment		
Rest of Victoria	67.4	73.7	Year 11 or below	66.8	74.8
Brisbane	73.9	79.8	Year 12	66.6	75.7
Rest of Qld	72.0	78.6	Certificate	70.2	77.5
Adelaide	74.0	80.4	Diploma	75.2	82.3
Rest of SA	69.4	76.3	Degree or higher	79.2	85.1
Perth	71.4	77.2	Dwelling type		
Rest of WA	68.6	78.2	House	70.5	78.1
Tasmania	73.6	81.7	Semi-detached	72.8	79.9
Northern Territory	81.5	87.7	Flat, unit, apartment	65.6	73.7
ACT	73.2	81.2	Other	63.9	70.5
Sex			Labour force status		
Male	68.6	76.5	Employed full-time	69.5	77.4
Female	71.7	79.0	Employed part-time	71.7	79.5
Age group (years)			Unemployed	58.5	69.9
15–19	57.3	71.3	Not in labour force	71.8	78.5
20–24	57.9	70.2	Employment status in main job*		
25–34	68.2	76.8	Employee	70.0	77.8
35–44	72.3	79.0	Employer	70.1	79.3
45–54	72.5	79.0	Own account worker	72.6	80.1
55–64	75.6	80.6	Contributing family worker	67.2	75.4
65–74	80.1	84.9	Occupation*		
75+	71.4	76.6	Managers / administrators	73.2	81.2
Marital status			Professionals	77.7	84.7
Married	73.0	79.4	Associate professionals	69.3	78.2
De facto	66.1	75.4	Tradespersons	66.6	75.3
Separated	71.4	78.5	Advanced clerical / service	68.4	76.2
Divorced	76.7	83.5	Intermediate clerical / sales / service	70.9	77.5
Widowed	78.8	82.3	Intermediate production / transport	64.1	71.2
Single	61.9	72.8	Elementary clerical / sales / service	67.4	76.8
Country of birth			Labourers	62.9	71.5
Australia	72.3	79.7			
Overseas			Total	70.2	77.8
Main	71.4	78.5	Number responding	9311	10392
English-speaking					
Other	58.6	67.8			

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 were subject to 100 per cent verification (i.e., the data was entered in twice and any discrepancies corrected). 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 (ASCO) and industry questions are coded to four-digit Australian and New Zealand Standard Industry Classification (ANZSIC).

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 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 expect to run a training course around the same time as the HILDA Survey Research Conference, which is scheduled for 19-20 July, 2007. 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. A modest registration fee will apply.

We plan to conduct user training sessions at least every two years.

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 data files along with a growing bibliography of research papers that use the HILDA Survey data.

- Contact the HILDA team at the Melbourne Institute – if your query is about getting access to the data or you have lost your password, contact the HILDA Administrative Assistant on 03 8344 2108 or email hilda-inquiries@unimelb.edu.au. Alternatively, if you have questions about the data files or variables, contact Nicole Watson on 03 8344 2088 or email n.watson@unimelb.edu.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

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.

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.

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. 2/02, 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.

APPENDIX 1a: SUMMARY OF HILDA SURVEY CONTENT, WAVES 1 – 5

The following table provides a guide to topics covered in the HILDA Survey across the first 5 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 5.0 of the HILDA CD-Rom.

HOUSEHOLD FORM	*Wave*				
	1	*2*	*3*	*4*	*5*
Sex[a]	X	X	X	X	X
Date of birth[a]	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
Marital status of household members	X				
Employment status of household members	X		X	X	X
Household relationships	X	X	X	X	X
Entrants – reasons for, and date of, joining household		X	X	X	X
Movers – reasons for, and date of, leaving household		X	X	X	X

a Pre-printed in Waves 2, 3 and 4.

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

HOUSEHOLD QUESTIONNAIRE (c'td)	*Wave*				
	1	*2*	*3*	*4*	*5*
Dwelling type[d]	X	X	X	X	X
Condition of dwelling (interviewer assessed)[d]	X	X	X	X	X
Housing Wealth					
Owner IDs and share owned		X			
First home buyer	X	X			
Year home purchased		X			
Purchase price of home		X			
Current value of home	X	X	X	X	X
Value of housing debt	X	X	X	X	X
Value of housing loans repayments	X	X	X	X	X
Year expect housing loan to be paid off	X	X	X	X	X
Value of initial housing loans		X			
Household Assets *[special module]*					
Other properties		X			
Equity investments		X			
Trust funds		X			
Childrens bank accounts		X			
Other cash-type investments		X			
Business wealth		X			
Value of vehicles		X			
Life insurance		X			
Collectibles		X			
Other					
Number of motor vehicles	X				
Weekly expenditure on groceries / food	X		X	X	X
Weekly expenditure on meals out	X		X	X	X
Adequacy of household income	X				
Total household income (bands)				X	X

b 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.
c One item added in Wave 3.
d Collected as part of the HF in Wave 1.

CONTINUING PERSON QUESTIONNAIRE	*Wave*				
	1	*2*	*3*	*4*	*5*
Country of birth & language					
Country of birth	X			X	
Year of arrival	X			X	
English as first language	X				

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*				
	1	*2*	*3*	*4*	*5*
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 14	X				
Why not living with parents	X				
Parents ever separated / divorce	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
Year left school	X	X	X	X	X
Type of school last attended	X	X	X	X	X
Qualifications studying for		X	X	X	X
Qualifications completed	X	X	X	X	X
Date completed qualification		X	X	X	X
Country in which completed qualification	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				
Employment status – ABS definition (9 questions)	X[e]	X	X	X	X

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

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*				
	1	*2*	*3*	*4*	*5*
Persons not in paid employment					
Job search activity	X	X	X	X	X
Looking for work – When began looking for work	X	X	X	X	X
Looking for work – Hours spent in job search in last wk		X	X	X	X
Looking for work – Intensive Assistance	X	X	X	X	X
Looking for work – Availability to start work	X	X	X	X	X
Looking for work – Difficulties finding a job	X	X	X	X	X
Looking for work – Number of job offers	X	X	X	X	X
Not looking for work – Main activity	X	X	X	X	X
Not looking for work –Preference to work	X	X	X	X	X
Not looking for work –Reasons for not looking	X	X	X	X	X
Not looking for work – Availability to start work	X	X	X	X	X
Reservation wage	X	X	X	X	X
Desired hours of work	X	X	X	X	X
Expected probability of finding a job	X	X	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
Whether retired	X	X	X	X	X
Year / age retired	X	X	X	X	X
Age plan to retire	X	X	X	X	X
Labour market calendar	X	X	X	X	X
Mutual Obligation activity	X	X	X	X	X
Income					
Current wage and salary income	X	X	X	X	X
Current income from government benefits	X	X	X	X	X
Financial year income by source	X	X	X	X	X
Credit card use and payment strategy	X	X	X	X	X
Wealth *[special module]*					
Bank accounts		X			
Credit card debt		X			
Other debts		X			
Superannuation		X			
Family Formation					
Number of children	X	X	X	X	X
Non-resident children characteristics	X	X	X	X	X
Financial support for non-resident children	X	X	X	X	X

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*				
	1	2	3	4	*5*
Amount of contact with youngest non-resident child	X	X	X	X	X
Employment status of other parent			X	X	X
Resident children characteristics	X	X	X	X	X
Financial support from other parent	X	X	X	X	X
Amount of contact other parent has with youngest child	X	X	X	X	X
Employment status of other parent			X	X	X
Desire to have another child	X	X	X	X	X
Likelihood of having another child	X	X	X	X	X
Number of additional children intend to have	X	X	X	X	X
Year intend to have next child		X			X
Fertility *[special module]*					
Partner currently pregnant					X
Self currently pregnant					X
Time stopped/started work pre/post birth of baby					X
Birth Control					X
Partnering / Relationships					
Changes in marital status		X	X	X	X
Current marital status	X	X	X	X	X
Current living circumstances		X	X	X	X
Single persons – Likelihood of marriage	X	X	X	X	X
De facto relationships – Year relationship started	X	X	X	X	X
Number of other de facto relationships		X	X	X	X
Retirement *[special module]*			X		
Health / Disability					
Whether has disability / health condition	X	X	X	X	X
Type of disability			X	X	X
Whether disability commenced in last year	X	X		X	X
Year of onset of disability			X		
Impact of disability on work (2 questions)	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	
Employment difficulties				X	
Education difficulties				X	
DVA Treatment Entitlements Card				X	
Private health insurance (8 questions)				X	

CONTINUING PERSON QUESTIONNAIRE (c'td)	*Wave*				
	1	2	3	4	5
Hospital visits in past 12 months (7 questions)				X	
Caring for others					X
Whether respondent is a carer in hh					X
Whether respondent is a carer outside hh					X
Youth module				X	
Other					
Life satisfaction (9 items)	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
Movers – Date moved to current address		X	X	X	X
Movers – Date left previous address			X	X	X
Movers – Reasons for moving		X	X	X	X
Intentions/Plans for next 3 years					
Move house					X
Where move					X
Stop/start working					X
Change Employment					X

e In Wave 1 a shorter series of questions was used.
f Additional sub-question included to better enable weekend workers to be identified.

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

SELF-COMPLETION QUESTIONNAIRE (c'td)	*Wave*				
	1	*2*	*3*	*4*	*5*
Employment					
Job characteristics (12 items)	X	X	X	X	X
Family friendliness of the workplace (3 items)	X				
Access to family friendly benefits (7 items)	X	X	X	X	X
Parenting					
Parenting stress (4 items)	X	X	X	X	X
Fairness of childcare	X	X	X	X	X
Work family gains and strains (13 items)	X	X	X	X	X
Other					
Attitudes about work and gender roles (14 items)	X				X
Attitudes to marriage/ children (10 items)					X
Benefits of employment (14 items)					X
Personality (36 items)					X
Sex		X	X	X	X
Age group		X	X	X	X

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 HILDA Survey instruments. 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.	FaCSIA.
W5: D16	Factors influencing decision to start looking for work	FaCSIA.
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 (W2: J23) was significantly modified. 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)

Rotation pattern	*First wave: qstn #*	*Data item / topic*	*Notes on origin / source*
Every wave	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.
Every wave	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.
From Wave 2 on	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.
From Wave 2 on	W2: B3	Number of cigarettes smoked per week	Developed with advice from staff at the AIHW.
Every wave	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.
Every wave	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.
Every wave	W1: B6	Frequency of feeling pressed for time	ABS 1999 Survey of Living Standards pilot (q. L1).
Every wave	W1: B7	Frequency of spare time	ABS 1999 Survey of Living Standards pilot (q. L3).
Waves 1 to 4, and then every 2nd wave	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.
Wave 6	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).
Waves 1 and 2 only	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.

SELF COMPLETION QUESTIONNAIRE (c'td)			
Rotation pattern	***First wave: qstn #***	***Data item / topic***	***Notes on origin / source***
Every wave	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.
Every wave	W1: B12	Perception of whether doing fair share of the housework	Negotiating the Life Course Study.
Every wave	W1: B14	Frequency of social interaction	Based on a question asked in the Tasmanian HCS.
Every wave	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).
Every wave	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.
From Wave 2 on	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.
Waves 3 and 4 and then every 3rd wave	W3: B10	Self-efficacy	Pearlin and Schooler (1978).
Every 3rd wave starting Wave 3	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.
Every 3rd wave starting wave 4	W4: B18	Religious denomination	Pre-coded categories selected on the basis of the most frequent responses to the 2001 Census.
Wave 5	W5: B10	Satisfaction with division of household tasks	Adapted from two questions asked in the GGS (wave 1, q. 202 and q. 402).
Wave 5	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.
Wave 5	W5: B17-B18	Use of domestic help	B17 came from GGS (wave 1, q. 404).

SELF COMPLETION QUESTIONNAIRE (c'td)			
Rotation pattern	***First wave: qstn #***	***Data item / topic***	***Notes on origin / source***
Wave 5	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.
From Wave 6 on	W6: B6-B7	Height and weight	Generic questions, but format influenced by like questions included in the 1994 AYS.
Wave 6	W6: B21	Community participation	Helen Berry, National Centre for Epidemiology and Population Health, ANU.
Every wave	W1: C1	Financial well-being (self-assessed prosperity)	Tested as part of IsssA 2000 (q. 5, p. 84).
Every wave	W1: C2	Stressful financial events	Based closely on ABS 1999 Survey of Living Standards pilot (q. H6).
Every wave	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.
Every wave	W1: C3b	Source of money in an emergency	Categories based on those used in Canadian Survey of Financial Security (q. L14).
Waves 1 to 4, and then every second wave	W1: C4	Family's savings habits	1998 SCF (X3015-3020).
Waves 1 to 4, and then every second wave	W1: C5	Savings time horizon	1998 SCF (X3008).
Waves 1 to 4, and then every second wave	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.
Waves 1 and 2 only	W1: C7	Attitudes to borrowing	Based closely on 1998 SCF (X402-406).
From Wave 2 on	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.
From Wave 5 on	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.

SELF COMPLETION QUESTIONNAIRE (c'td)			
Rotation pattern	***First wave: qstn #***	***Data item / topic***	***Notes on origin / source***
Wave 1 and Wave 5	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).
Waves 5 and 6	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).
Wave 5	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.
Wave 5	W5: D4	Perceived benefits of paid employment	Based on a proposal from FaCS.
Every wave	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).
Wave 1 only	E2	Family friendliness of workplace.	Inspired by work of Marshall and Barnett (1993).
Every wave	F2	Parenting stress	PSID Child Development Supplement 1997, Primary Caregiver of Target Child – Household Questionnaire (q. A29).
Every wave	F3	Perception of whether doing fair share of the child care	Negotiating the Life Course Study.
Every wave	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)
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 2a: CHANGES TO THE DATA FROM RELEASE 5.0

Changes to wave 5 files between releases 5.0 (Jan2007) and 5.1 (March 2007)

Cross wave

- Subset of history variables.

  - calendar calculations were incorrectly overlapping by one time period (0.03 year) which affected time since leaving school (_*ehtse*, 103 cases)/ time since leaving school in paid work (_*ehtjb*, 68 cases)/ time since leaving school unemployed and looking for work (_*ehtuj*, 17 cases)/ time since leaving school not working and not looking for work (_*ehto*, 19 cases).

  - time in current marital status variables were incorrectly set to -4 when there was an intervening period of defacto which ended and the respondent reverted to their prior marital status. Affected a handful of cases for time in current marriage (_*mrcdur*)/time widowed (_*mrwdur*)/time currently separated or divorced (_*mrsdur*). These are now calculated each wave after bringing forward the marriages grid, rather than attempting to update the derived variables from previous waves.

Corrections in calculations

- *efmagelh* 'History: Age left home' Check for presence of parents in current wave incorrect. After correction, 135 cases set to an age; 56 cases with an age reset to still at home.
- *efmnsib*. 'History: How many siblings' 1 case corrected number of siblings from 6 to 7
- *efmfocc2* and *efmmocc2*. 'History: 2-digit mothers and father occupation' (4 digit version not affected). 165 *efmfocc2* and 129 *efmmocc2* missing cases set to a 2-digit occupation code. Data not brought forward correctly from earlier waves when no wave 4 information.
- *ertage* 'History: Age retired/intends to retire' 55 missing cases reset to age. Data not brought forward correctly from earlier waves when no wave 4 information.
- *eehtse* 'History: Time since FT education - years' 97 corrections
- *eehtjb* 'History: Time in paid work - years' 80 corrections
- *eehtuj* 'History: Time unemployed and looking for work - years' 11 corrections
- *eehto* 'History: Time not working and not looking for work - years' 31 corrections Data not brought forward correctly from earlier waves when no wave 4 information.
- *emrplvt* 'History: Years living together before present or most recent marriage' 39 corrections
- *emr1lvt* 'History: Years living together before first marriage'. 6 corrections
- *emrcdur* 'History: Current marriage duration - years' 3 corrections
- *emrsdur* 'History: Current separated or divorced from date of separation - years' 3 corrections. Data not brought forward correctly from earlier waves when no wave 4 information.

- A number of other history variables had 1 or 2 cases change (but no more) when the data was brought forward correctly (*eanatsi, efmlwop, efmpdiv, efmhsib, efmnsib, efmeldst, efmfcob, efmmcob, efmfemp, efmfocc2, efmfuemp, efmmemp, efmmocc2, eedagels, eedhists, eedtypes, eedcly, emrcdur, emrsdur, ehsyrcad, efmfoccs, efmmoccs, efmfi82, efmmi82*).
- ehgagef, ehgagf1 to ehgagf14. Flag for imputed age or dob contained no data.
- dobf 'Date of birth flag' (in in-confidence master file)
- Corrected calculation of DV's for HQ14 & HQ16 non-employment related childcare as calculations had only been made if child received any employment-related and any nonemployment related childcare. (*ecnsu_bs, ecnsu_gu, ecnsu_ge, ecnsu_au, ecnsu_ae, ecnsu_fo, ecnsu_ft, ecnsu_ps, ecnsu_fd, ecnsu_pd, ecnsu_fc, ecnsu_o1, ecnsu_o2, ecnsu_na, ecnsu_np, ecnsh_bs, ecnsh_gu, ecnsh_ge, ecnsh_au, ecnsh_ae, ecnsh_fo, ecnsh_ft, ecnsh_ps, ecnsh_fd, ecnsh_pd, ecnsh_fc, ecnsh_o1, ecnsh_o2, ecnpu_bs, ecnpu_gu, ecnpu_ge, ecnpu_au, ecnpu_ae, ecnpu_fo, ecnpu_ft, ecnpu_ps, ecnpu_fd, ecnpu_pd, ecnpu_kp, ecnpu_o1, ecnpu_o2, ecnpu_na, ecnpu_np, ecnph_bs, ecnph_gu, ecnph_ge, ecnph_au, ecnph_ae, ecnph_fo, ecnph_ft, ecnph_ps, ecnph_fd, ecnph_pd, ecnph_kp, ecnph_o1, ecnph_o2*). (The majority of corrections were from -1 not asked to 0 not used)
- Total non-employment related childcare costs for school aged children and not yet at school children
- 2 additional households calculated as late SCQs data entered. SCQ derived variables now calculated for these 2 cases (SF36 scales; personality scales, Braziers SF6D; Time use: combining hours and minutes variables, household expenditure.)
- Supplementary data for 8460103 collected at wave 6 entered for new entrant variables at wave 5

Deleted variables

- Uses data in HQ14 and HQ16 for up to 6 children (general release files)
- Hours data in HQ14 and HQ16 for up to 6 children (general release files)

Documentation corrections

- HQ 14 & HQ16 Non-employment childcare; Costs and derived variables relating to used and hours renamed from *ensc* to *ecnsc*; and *enpc* to *ecnpc* to conform to prior waves
- 2 digit occupation and industry (asco isic and isco) were labelled main job. Labels changed to current main job
- *ahhrepgp* 'DV: Replicate group' now labeled (master file).

<u>Changes made to wave 4 between releases 5.0 (Jan 2007) and 5.1 (March 2007)</u>

Corrections in calculations

- *dfmfocc2* 'History: 2-digit father occupation' (4 digit version is not affected as it is not a DV). 21 missing cases set to an occupation
- *dfmmocc2* 'History: 2-digit mothers occupation' 13 missing cases set to an occupation

- *drtage* History: Age retired/intends to retire. 98 missing cases reset to age.
- *dmrplvt* 'History: Years living together before present or most recent marriage'. 102 missing cases set to a value.
- *dmr1lvt* 'History: Years living together before first marriage'. 15 missing cases set to a value.
- *dmr2lvt* 'History: Years living together before second marriage'. 7 missing cases set to a value.
- *dmrsdur* 'History: Current separated or divorced from date of separation - years'.1 case set to 0.
- *dhgagef*, *dhgagf1* to *dhgagf14*. Flag for imputed age or dob contained no data

New variables

- Total non-employment related childcare costs for school aged children and not yet at school children.

Documentation corrections

- *_jbmwpsz* Corrected question number from C29 to C31 in variable label
- 2 digit occupation and industry (asco isic and isco) were labelled main job. Labels changed to current main job
- Relabelled *dcnsu_np* and *dcnpu_np* from 'not applicable' to 'none'

Changes made to wave 3 between releases 5.0 (Jan 2007) and 5.1 (March 2007)

- **Corrections in calculations**

- *chgagef*, *chgagf1* to *chgagf14*. Flag for imputed age or dob contained no data

New variables

- Total non-employment related childcare costs for school aged children and not yet at school children

Documentation corrections

- *cjbmwpsz* Corrected qn no from C29 to C31 in variable label
- 2 digit occupation and industry (asco isic and isco) were labelled main job. Labels changed to current main job

Changes made to wave 2 between releases 5.0 (Jan 2007) and 5.1 (March 2007)

New variables

- Total non-employment related childcare costs for school aged children and not yet at school children

Documentation corrections

- 2 digit occupation and industry (asco isic and isco) were labelled main job. Labels changed to current main job

Changes made to wave 1 between releases 5.0 (Jan 2007) and 5.1 (March 2007)

Documentation corrections

- 2 digit occupation and industry (asco isic and isco) were labelled main job. Labels changed to current main job

APPENDIX 2b: CHANGES TO THE DATA FROM RELEASE 4.1

Changes to wave 4 files between releases 4.1 and 5.0

Cleaning

- Cleaned 25 relationships in 15 households.

Corrections in calculations

- FTB-A if receiving child support. Program corrected so reduction only applied if child support exceeds maintenance-free threshold
- Mother and father ids (HF person number and cross wave identifiers: dhhmid dhhfid dhhmxid dhhfxid) are now calculated for all parent-child relationships in household. Previously only supplied for individuals classified as < 15, dependent or non-dependent children (dhhrih=8,9,10)
- dedhigh Highest level of education: 79 corrections for achieved year 10 and have cert II to cert II; 7 corrections for currently in year 12 or 11 and have cert II to cert II.
- dehto 'History: Time not working and not looking for work - years' - incorrectly had values of years unemployed
- Current wages and salaries: dwsce corrected by $15 for 20 cases where net annual income was given for main job and (40228 < net income <= 46318). Grossing up parameter was not updated from previous financial year, though it had been updated correctly for the ‘financial year’ wages and salaries and ‘other jobs’ wages and salaries.
- dedagels. History: Age left school. If respondent incorrectly recalled that they were in school at previous interview and skipped date left school, set to last recorded age at school.

New variables

- dhgagef Imputed imputed ages for missing cases (and imputed dates of birth in the unconfidentialised (In-Confidence) files). Only imputed variables are supplied along with a flag to identify whether a case has been imputed.
- dghsf6d Brazier’s SF-6D Health state classification.
- dhhsos ABS section of state
- djhlhruw 'History: Hours a week usually worked in last job (not currently in paid work)'
- djhljcnt 'History: Contract of employment, last job (not currently in paid work)'

- djhlji82 'History: ISCO-88 2-digit: Last job (not currently in paid work)'
- djhljii2 'History: ISIC 3.1 Last job (not currently in paid work) International industry classification 2-digit'
- djhljoc2 'History: Occupation last job (not currently in paid work) 2-digit ASCO'
- djhljocs 'History: ANU4 occupational status scale: Last job (not currently in paid work)'
- djhljrea 'History: Main reason stopped working (not currently in paid work)'
- djhljtha 'History: Employment length - answered weeks or years, last job (not currently in paid work)'
- djhljtwk 'History: Weeks worked for last employer (not currently in paid work)'
- djhljtyp 'History: Employment status, last job (not currently in paid work)'
- djhljtyr 'History: Years worked for last employer (not currently in paid work)'
- djhtsjha 'History: How long since last worked for pay answered weeks or years (not currently in paid work)'
- djhwku 'History: How long since last worked for pay - less than 1 year - weeks (not currently in paid work)'
- djhyru 'History: How long since last worked for pay - years (not currently in paid work)'
- doiint, doidiv, doiroy, doidvry, doirntp, doirntn. Dividends, royalties, interest and rent. Nil amounts are included. These variables are calculated as inputs into the imputation process and are now provided in the files.
- doiflss, doiflssa, doiflsw, doiflswa. Lump sum workers compensation and lump sum superannuation. At F32 (any other source of financial year income), lump sum payments are now provided in their own variables, and counted in windfall income. They are no longer counted in regular financial year income.
- dtcr, dtcnr. Number of own resident and non-resident children. General release file. Due to an inconsistency in the allocation of children into G1b & G1d; and into G1c & G1e (all waves), the components are no longer supplied on the general release file. Instead sums are provided in these new DV's. Note that the number of children in the resident children grid (at G15) can exceed dtcr, as the number of grandchildren, stepchildren and foster children without a responsible parent in the household (G13) are also listed in the resident children grid.

Deleted variables

- dtchere, dtcnpd, dtcpdwl, dtcptnpd (see note for new variables dtcr and dtcnr)

Documentation corrections

- Added value labels for HQ Q5a to Q5l Difficulties with childcare dccdifgq to dccdifsn -3 'Dont know/not applicable'.
- Added 9999 labels for other industry codes to match djbmind dpjoind dpjotind .
- Added value labels dujljind 9999 'Not enough information'. To match later waves
- w4 confidentialised marked up person questionnaires non-resident child grid G3 showed dncmth1-dncmth13 & dncyr1-dncyr13; but these variables are not supplied on the confidentialised file

Change made to master file

- hhsm Sample membership has been updated

Changes made to wave 3 between releases 4.1 and 5.0

Cleaning

- Cleaned 6 relationships in 5 households.

Corrections in calculations

- Reduction in FTB-A if receiving child support. Code corrected so reduction only applied if child support exceeds maintenance-free threshold
- Mother and Father ids (HF person number and xwave, chhmid chhfid chhmxid chhfxid) are now calculated for all parent-child relationships in household. Previously only supplied for individuals classified as aged < 15, and dependent or non-dependent children (chhrih=8,9,10)
- cedhigh Highest level of education: 74 corrections for achieved year 10 and have cert II to cert II; 7 corrections for currently in year 12 or 11 and have cert II to cert II.
- cehtse 'History: Time since FT education - years' was zero for all 833 new entrants. Now has 547 valids, 285 not asked and 1 missing
- cedagels. History: Age left school. If respondent incorrectly recalled that they were in school at previous interview and skipped date left school, set to last recorded age at school.

New variables

- chgagef Imputed imputed ages for missing cases (and imputed dates of birth in the unconfidentialised (In-Confidence) files). Only imputed variables are supplied along with a flag to identify whether a case has been imputed.
- cghsf6d Brazier's SF-6D Health state classification.
- chhsos ABS section of state
- cjhlhruw 'History: Hours a week usually worked in last job (not currently in paid work)'
- cjhljcnt 'History: Contract of employment, last job (not currently in paid work)'
- cjhlji82 'History: ISCO-88 2-digit: Last job (not currently in paid work)'
- cjhljii2 'History: ISIC 3.1 Last job (not currently in paid work) International industry classification 2-digit'
- cjhljoc2 'History: Occupation last job (not currently in paid work) 2-digit ASCO'
- cjhljocs 'History: ANU4 occupational status scale: Last job (not currently in paid work)'
- cjhljrea 'History: Main reason stopped working (not currently in paid work)'
- cjhljtha 'History: Employment length - answered weeks or years, last job (not currently in paid work)'
- cjhljtwk 'History: Weeks worked for last employer (not currently in paid work)'
- cjhljtyp 'History: Employment status, last job (not currently in paid work)'
- cjhljtyr 'History: Years worked for last employer (not currently in paid work)'
- cjhtsjha 'History: How long since last worked for pay answered weeks or years (not currently in paid work)'
- cjhwku 'History: How long since last worked for pay - less than 1 year - weeks (not currently in paid work)'
- cjhyru 'History: How long since last worked for pay - years (not currently in paid work)'
- coiint, coidiv, coiroy, coidvry, coirntp, coirntn. Dividends/royalties, interest and rent. Nil amounts are included. These variables are calculated as inputs into the imputation process and are now provided in the files and as components for FY investment income on the responding person financial year income model diagram.

- coiflss, coiflssa, coiflsw, coiflswa. Lump sum workers comp and lump sum superannuation. At F32 (any other source of financial year income), lump sum payments are now provided in their own variables, and counted in windfall income. They are no longer counted in regular financial year income

- ctcr, ctcnr. Number of own resident and non-resident children. General release file. Due to an inconsistency in the allocation of children into G1b & G1d; and into G1c & G1e (all waves), the components are no longer supplied on the general release file. Instead sums are provided in these new DV's. Note that the number of children in the resident children grid (at G15) can exceed ctcr, as the number of grandchildren, stepchildren and foster children without a responsible parent in the household (G13) are also listed in the resident children grid.

Deleted variables

- ctchere, ctcnpd, ctcpdwl, ctcptnpd (see note for new variables dtcr and dtcnr)

Documentation corrections

- Added value labels for HQ Q5a to Q5m Difficulties with childcare variables bccdifcq to bccdifch -3 'Dont know/not applicable'.

Changes made to wave 2 between releases 4.1 and 5.0

- ***Cleaning***

- Cleaned 6 relationships in 3 households.

Corrections in calculations

- Reduction in FTB-A if receiving child support. Program corrected so reduction only applied if child support exceeds maintenance-free threshold

- FTB-A Corrected wave 2 parameters for reduction in FTB-A if receiving child support: Maintenance free area parameters from (1007.4+((totkids-1)*335.8)) to (1062.15+((totkids-1)*354.05)).

- FTB-A Approximation for the large family supplement, corrected $208.05 per child to 219.00

- Mother and Father ids (HF person number and xwave, bhhmid bhhfid bhhmxid bhhfxid) are now calculated for all parent-child relationships in hh. Previously only supplied for individuals classified as < 15, dependent or non-dependent children (bhhrih=8,9,10)

- bedhigh Highest level of education: 69 corrections for achieved year 10 and have cert II to cert II; 10 corrections for currently in year 12 or 11 and have cert II to cert II.

- bedagels. History: Age left school. If respondent incorrectly recalled that they were in school at previous interview and skipped date left school, set to last recorded age at school.

New variables

- bhgagef Imputed imputed ages for missing cases (and imputed dates of birth in the unconfidentialised (In-Confidence) files). Only imputed variables are supplied along with a flag to identify whether a case has been imputed.
- bghsf6d Brazier's SF-6D Health state classification.
- bhhsos ABS section of state
- bjhlhruw 'History: Hours a week usually worked in last job (not currently in paid work)'
- bjhljcnt 'History: Contract of employment, last job (not currently in paid work)'
- bjhlji82 'History: ISCO-88 2-digit: Last job (not currently in paid work)'
- bjhljii2 'History: ISIC 3.1 Last job (not currently in paid work) International industry classification 2-digit'
- bjhljoc2 'History: Occupation last job (not currently in paid work) 2-digit ASCO'
- bjhljocs 'History: ANU4 occupational status scale: Last job (not currently in paid work)'
- bjhljrea 'History: Main reason stopped working (not currently in paid work)'
- bjhljtha 'History: Employment length - answered weeks or years, last job (not currently in paid work)'
- bjhljtwk 'History: Weeks worked for last employer (not currently in paid work)'
- bjhljtyp 'History: Employment status, last job (not currently in paid work)'
- bjhljtyr 'History: Years worked for last employer (not currently in paid work)'
- bjhtsjha 'History: How long since last worked for pay answered weeks or years (not currently in paid work)'
- bjhwku 'History: How long since last worked for pay - less than 1 year - weeks (not currently in paid work)'
- bjhyru 'History: How long since last worked for pay - years (not currently in paid work)'
- boiint, boidiv, boiroy, boidvry, boirntp, boirntn. Dividends/royalties, interest and rent. Nil amounts are included. These variables are calculated as inputs into the imputation process and are now provided in the files and as components for FY

investment income on the responding person financial year income model diagram.

- boiflss, boiflssa, boiflsw, boiflswa. Lump sum workers comp and lump sum superannuation At F32 (any other source of financial year income), lump sum payments are now provided in their own variables, and counted in windfall income. They are no longer counted in regular financial year income

- btcr btcnr Number of own resident and non-resident children. General release file. Due to an inconsistency in the allocation of children into G1b & G1d; and into G1c & G1e (all waves), the components are no longer supplied on the general release file. Instead sums are provided in these new DV's. Note that the number of children in the resident children grid (at G15) can exceed btcr, as grandchildren, stepchildren and foster children without a responsible parent in the household (G13) are also listed in the resident children grid.

Deleted variables

- Combined file: dropped processing flag called 'last'

- btchere, btcnpd, btcpdwl, btcptnpd (see note for new variables dtcr and dtcnr)

Documentation corrections

- HQ added value labels for types of HQ Q4a to Q4m childcare used bccdifgq to bccdifch -3 'Dont know/not applicable'

Changes made to wave 1 between releases 4.1 and 5.0

Cleaning

- aujljrsn D20 Main reason stopped working, swapped backcoded responses 15 (to 'too much travel time') and 16 (to 'migrated to a new country') to match backcoding at subsequent waves

- Cleaned 5 relationships in 2 households.

- Education - asced coding. For conformity with later waves, backcoded C7a qualifications obtained (mult response, 9 changes) and C12a qualifications studying for (5 changes). These cleans had been made in the construction of the ASCED derived variables, but not in the originating variables. One change of asced 400 -> asced 421 for qualifications obtained. Changes of 310 -> 321 & unknown -> 521 for continuing qualifications. These changes do not alter highest qualification obtained aedhigh.

Corrections in calculations

- Reduction in FTB-A if receiving child support. Program corrected so reduction only applied if child support exceeds maintenance-free threshold

- Mother and Father ids (HF person number and xwave: ahhmid ahhfid ahhmxid ahhfxid) are now calculated for all parent-child relationships in hh. Previously only supplied for individuals classified as < 15, dependent or non-dependent children (ahhrih=8,9,10)

- aedhigh Highest level of education: 68 corrections for achieved year 10 and have cert II to cert II;

New variables

- Imputed DOB and imputed ages for missing cases introduced w1-w5. Separate date of birth and age variables are not supplied, but imputed age or date of birth flags supplied.

- Added Brazier's SF-6D Health state classification aghsf6d.

- ahhsos ABS section of state

Deleted variables

- ajbmocc was renamed from ajbmoccs (which was inconsistent with the variable naming in other waves).

Documentation corrections

- HQ added value labels for types of HQ Q4a to Q4m childcare used accdifgq to accdifch -3 'Dont know/not applicable'

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