Household, Income and Labour Dynamics in Australia (HILDA) Survey

Release 21.0

Published December 2022

Extracting the zips

General Release

To extract the General Release data zips, you can use the Windows 7/8/10 explorer, the default MacOS archive utility or most shareware and open source extraction programs.

Restricted Release

The Restricted Release data zips are encrypted / password protected. The Australian Data Archive will send you an email containing the password. The data zips need to be extracted before use. You will need access to a program such as 7-Zip, WinZip or MacOS UnArchiver which supports AES encrypted zip files[1].

You cannot extract HILDA's password protected zips using Windows 7/8/10 explorer, which gives the error message "Windows cannot complete the extraction - The destination file could not be created" or "ERROR 0x80004005". You cannot use most shareware extraction programs or the default MacOS archive utility.

Windows users of 7-Zip can 'right-click' on the data zip file → Open with → '7-Zip' to extract the files to your destination. You will be prompted for the password. If you get the 7-Zip error message "File skipped, unknown compression method", upgrade 7-Zip to version 9.20 or later (for 32 or 64-bit Windows) at http://www.7-zip.org/download.html.

Post-release data issues

Problems found with the data post the release are documented online at http://melbourneinstitute.unimelb.edu.au/hilda/for-data-users/data-and-documentation-issues.

Features of wave 21

- *Questions introduced for the first time in wave 21*
    - CPQ:K76a,b/NPQ:AA15a,b and CPQ:K77/NPQ:AA16: Mother's location of birth
    - C36a to C36e: Presenteeism
    - K5a, K5b, K7, K9a and K9b: COVID-19 specific content
    - CPQ:K78/NPQ:K76, CPQ:K79/NPQ:K77 and CPQ:K80/NPQ:K78: Service in the Australian Defence Force
    - SCQ B25a and B25b: Pharmaceutical drug use
    - SCQ B26: E-cigarettes
- *Wave 21 HQ repeats these special topic modules and questions*
    - HQ: Children's health and health care utilisation

[1] Apple users can purchase Keka, "Winzip for the MacOS" or use the shareware Unarchiver (http://unarchiver.c3.cx/).

- *Wave 21 PQ repeats these special topic modules and questions*
    - PQ Section K: Physical and mental health, diet and quantity of sleep
- *Wave 21 SCQ repeats these special topic questions*
    - SCQ B6: Binge drinking
    - SCQ B12: Waist circumference
    - SCQ B13-B16: Dieting and Perceptions of weight
    - SCQ B17: Food frequency
    - SCQ B18: Kessler 10
    - SCQ B28: Personality
    - SCQ B30 and SCQ B31: Sleep quality
    - SCQ E4: Work-family gains/strains

Additions and corrections for Release 21

Revised / new derived variables

- ASGS 2021 geography. The ABS created a new geography for the 2021 census and this is supplied from Release 21 of HILDA. Statistical Areas Levels 1 to 4 and Local Government Area are added to the Restricted Release files. In both the General Release and Restricted Release files, Greater Capital City Statistical Area (GCCSA) and Section of State (SOS) are available. Remoteness Area (RA) will be available in the next release.

  The variable names for the ASGS 2021 are as follows:

Variable Name	Label
_hhs3lga	DV: ASGS 2021 Local Government Area (LGA)
_hhs3sa1	DV: ASGS 2021 Statistical Area Level 1 (SA1) 11-digit
_hhs3sa2	DV: ASGS 2021 Statistical Area Level 2 (SA2) 9-digit
_hhs3sa3	DV: ASGS 2021 Statistical Area Level 3 (SA3) 5-digit
_hhs3sa4	DV: ASGS 2021 Statistical Area Level 4 (SA4) 3-digit
_hhs3gcc	DV: ASGS 2021 Greater Capital City Statistical Area (GCCSA)
_hhs3sos	DV: ASGS 2021 Section of State (SOS)

- ASGS 2016 geography. The ASGS 2016 Statistical Areas Levels 1 is now supplied for all waves. _ hhsa116 (DV: ASGS 2016 Statistical Area Level 1 (SA1) 7-digit) was only available from wave 16 in previous releases.

- Mother's location at birth. Responses were used to generate two new derived variables, _fmlbbd (DV: Where your mother was living when you were born – broad) and _fmlbdt (DV: Where you mother was living when you were born – detail). _fmlbbd values are the ASGS 2021 Greater Capital City Statistical Area (GCCSA) where the respondent's mother was living when they were born. _fmlbdt is available only in the Restricted Release. Where the respondent's mother was living in one of the seven capital cities listed in the questionnaire a 2-digit GCCSA value is provided at _fmlbdt, and for those with a mother living outside these capital cities who provided an 'Other' or a rural locality a 3-digit ASGS 2021 Statistical Area Level 4 (SA4) value is provided.

- Revision to calculation of FTB A in all waves to factor in child support income.

- Included _bnfcvp (COVID-19 Payments) in the calculation of _bnfoni (DV: Financial year Australian Government. Non-income support other than family payments).

Corrections / backcoding

- Added 15 adults and five children to wave 20, and three adults and one child to wave 19. These were missed on the Household Form for these waves.
- Resolved the inconsistencies in the age imputation flag for all waves.
- Corrected the derive variable _slhrwk (DV: Hours of sleep per week). When naps are not taken (_slenapa = 2 or _slunapa = 2), the value for number of hours of sleep from naps (_slenaph or _slunaph) is -1 (not asked) since it is skipped. As a result, a value of 1 is incorrectly taken off the total weekly sleep amount. As a fix, if _slenapa = 2 or _slunapa = 2, then _slenaph or _slunaph is set to 0 instead of -1. In addition, _slhrwk is set to -4 or -3 if the response in _slenapa or _slunapa is Refused/Don't know. The waves affected are waves 13, 17 and 21.
- Corrected the value label of the derived variable _helv10 (DV: How likely that you will live to 75 or at least 10 more years) for value = 1 from "Very unlikely" to Very likely" The waves affected are waves 9, 13, 17 and 21.

- Included Australia (1101) in aedcly (DV: Country of last school year). Those who responded No to C5a (Did you do your last year of school overseas?) in wave 1 were coded -1 (Not asked) in aedcly. The response was recoded to Australia (1101) as the history variable in the subsequent waves includes Australia. There were 10,892 cases affected in wave 1.

HILDA Survey User Manual

The *HILDA Survey User Manual* can be downloaded from http://melbourneinstitute.unimelb.edu.au/hilda/for-data-users/user-manuals. Any updates to the HILDA Survey User Manual will be provided on this page of the HILDA website. The manual describes:

- missing data conventions;
- introduction to the derived variables;
- how to match the wave data files together and create longitudinal files;
- the income model;
- income, wealth and expenditure imputation;
- industry and occupation variables in both Australian and International coding schemes;
- overview of the data documentation;
- summary of the data quality issues;
- how to use the weights;
- summary of the survey design and data collection procedures; and
- answers many other frequently asked questions.

To quickly find variable names, the manual can be used in conjunction with the subject level coding framework ("Subject Level Coding Framework 210c.pdf"), which can be searched on question number, keyword or variable name (excluding the first character wave identifier). It shows which waves a variable is available in, the source questionnaire / question ('DV' for derived, and 'History' for history variables), the values the variable takes and notes on the construction of variables.

Online data dictionary

A variable database is available on the HILDA website (https://www.online.fbe.unimelb.edu.au/HILDAodd/Default.aspx).

You can search the online data dictionary by keyword, subject area, or variable name.

File structures

Zip files are produced for:

- Documentation
- Frequencies
- SAS data files and programs
- SPSS data files
- STATA data files

The files have a '*wrt*' (wave, release, type) naming convention of 'a210c', 'b210c', 'c210c', etc.:

- a210c refers to wave 1, Release 21.0, General Release
- b210c refers to wave 2, Release 21.0, General Release

  . . .
- u210c refers to wave 21, Release 21.0, General Release[2].

Datasets

- Household = Information collected or aggregated for the household as a whole. Data from the Household Form and Household Questionnaire and derived variables relating to the household.
- Eperson = Enumerated person in current wave (member of a fully or partly responding household, irrespective of whether an individual interview was completed or not, or if they are aged less than 15). Data from the Household Form is transformed to the person level. Derived and imputed variables in this file relate to everyone in-scope in responding households (i.e. where at least one adult provided an interview), including interviewed adults, non-interviewed adults and non-interviewed children (_hgni).
- Rperson = Responding person (individual interview completed). Data from the New Person Questionnaire, the Continuing Person Questionnaire and the Self Completion Questionnaire and associated derived and history variables.
- Combined = All data from Eperson and Rperson and Household datasets combined into an enumerated person level dataset. Household data is repeated for each person from the same household. Sorted by cross wave person identifier (xwaveid) for longitudinal matching (wide file).
- Master = Enumerated in any wave. A longitudinal sample description file which contains identifiers, household status and interview status at each wave. Also includes distance moved, sex, year of death, age at death, strings showing the pattern of waves enumerated and waves interviewed, the year entered, year left, number of times a person interview has been completed, year of first interview, year of last interview.
- Longitudinal Weights = Enumerated in any wave. Contains weights for the balanced panel from any wave to any other wave and for any pair of waves.
- CNEF_long = A long HILDA dataset for the Cross-National Equivalence File, a project started by Cornell University which creates comparable longitudinal social and economic population datasets for other countries, such as the USA, Germany, the UK, Switzerland, Russia, Korea and Canada.

[2] The release number shows the cumulative releases of HILDA data and does not imply there has been a prior release of wave 20. For historical reasons, General Release files have type 'c' and Restricted Release files type 'u'.

With the exception of the longitudinal sample description ("Master") file and the CNEF_long file, the datasets are supplied in individual waves. There are example programs in the Program Library on the HILDA website for creating wide and long multi-wave datasets (http://melbourneinstitute.unimelb.edu.au/hilda/for-data-users/program-library).

HILDA quick reference guides[3]

Wave	1	2	3	4	5	6	7	8	9	10
_	a	b	c	d	e	f	g	h	i	j
Current Year[4]	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Financial Year[5]	2000-2001	2001-2002	2002-2003	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010

Wave	11	12	13	14	15	16	17	18	19	20
_	k	l	m	n	o	p	q	r	s	t
Current Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Financial Year	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	2019-2020

Wave	21
_	u
Current Year	2021
Financial Year	2020-2021

Alternate

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u

[3] You may find it helpful to attach one of these to your monitor for quick reference.
[4] Current year. Year in which the bulk (98%) of the PQ interviews were conducted. E.g., current hours worked, current wages and salary, wage rates, life satisfaction.
[5] Financial (fiscal) year. The period from 1st July to 30th June. The detailed income model is constructed for the financial year preceding the date of interview.

General release file sizes

Wave	Filename[6]	Cases	Variables
W1	Combined a210c	19914	3402
W2	Combined b210c	18295	4490
W3	Combined c210c	17690	4481
W4	Combined d210c	17209	4279
W5	Combined e210c	17467	4714
W6	Combined f210c	17453	4979
W7	Combined g210c	17280	4896
W8	Combined h210c	17144	5019
W9	Combined i210c	17632	5112
W10	Combined j210c	17855	5384
W11	Combined k210c	23415	5579
W12	Combined l210c	23182	5436
W13	Combined m210c	23299	5361
W14	Combined n210c	23114	5618
W15	Combined o210c	23305	5686
W16	Combined p210c	23507	5338
W17	Combined q210c	23442	5600
W18	Combined r210c	23267	5991
W19	Combined s210c	23260	6161
W20	Combined t210c	22951	5768
W21	Combined u210c	22434	5891
W1	Eperson a210c	19914	232
W2	Eperson b210c	18295	299
W3	Eperson c210c	17690	282
W4	Eperson d210c	17209	282
W5	Eperson e210c	17467	282
W6	Eperson f210c	17453	298
W7	Eperson g210c	17280	282
W8	Eperson h210c	17144	282
W9	Eperson i210c	17632	292
W10	Eperson j210c	17855	326
W11	Eperson k210c	23415	314
W12	Eperson l210c	23182	316
W13	Eperson m210c	23299	316
W14	Eperson n210c	23114	333
W15	Eperson o210c	23305	317
W16	Eperson p210c	23507	316
W17	Eperson q210c	23442	323
W18	Eperson r210c	23267	359
W19	Eperson s210c	23260	343
W20	Eperson t210c	22951	343
W21	Eperson u210c	22434	343
W1	Household a210c	7682	928
W2	Household b210c	7245	1301
W3	Household c210c	7096	1051
W4	Household d210c	6987	1082
W5	Household e210c	7125	1146
W6	Household f210c	7139	1403
W7	Household g210c	7063	1160
W8	Household h210c	7066	1160

[6] Stata and SAS filenames have underscores instead of spaces.

Wave	Filename[6]	Cases	Variables
W9	Household i210c	7234	1425
W10	Household j210c	7317	1459
W11	Household k210c	9543	1252
W12	Household l210c	9537	1319
W13	Household m210c	9555	1440
W14	Household n210c	9538	1543
W15	Household o210c	9631	1203
W16	Household p210c	9750	1329
W17	Household q210c	9741	1460
W18	Household r210c	9638	1746
W19	Household s210c	9664	1409
W20	Household t210c	9555	1595
W21	Household u210c	9358	1654
W21	Longitudinal weights u210c	45570	1288
W21	Master u210c	45570	162
W1	Rperson a210c	13969	2411
W2	Rperson b210c	13041	3072
W3	Rperson c210c	12728	3315
W4	Rperson d210c	12408	3082
W5	Rperson e210c	12759	3453
W6	Rperson f210c	12905	3461
W7	Rperson g210c	12789	3621
W8	Rperson h210c	12785	3744
W9	Rperson i210c	13301	3569
W10	Rperson j210c	13526	3808
W11	Rperson k210c	17612	4207
W12	Rperson l210c	17475	3999
W13	Rperson m210c	17500	3803
W14	Rperson n210c	17511	3957
W15	Rperson o210c	17605	4365
W16	Rperson p210c	17693	3891
W17	Rperson q210c	17570	4021
W18	Rperson r210c	17434	4106
W19	Rperson s210c	17462	4613
W20	Rperson t210c	17070	4034
W21	Rperson u210c	16549	4098

Restricted Release (by special request to DSS via Dataverse)[7]

Wave	Filename[8]	Cases	Variables
W1	Combined a210u	19914	4300
W2	Combined b210u	18295	5231
W3	Combined c210u	17690	5225
W4	Combined d210u	17209	5092
W5	Combined e210u	17467	5904
W6	Combined f210u	17453	6151
W7	Combined g210u	17280	6055
W8	Combined h210u	17144	6197
W9	Combined i210u	17632	6213
W10	Combined j210u	17855	6497
W11	Combined k210u	23415	6700
W12	Combined l210u	23182	6539
W13	Combined m210u	23299	6464
W14	Combined n210u	23114	6728
W15	Combined o210u	23305	6809
W16	Combined p210u	23507	6441
W17	Combined q210u	23442	6705
W18	Combined r210u	23267	7101
W19	Combined s210u	23260	7286
W20	Combined t210u	22951	6875
W21	Combined u210u	22434	6994
W1	Eperson a210u	19914	270
W2	Eperson b210u	18295	338
W3	Eperson c210u	17690	323
W4	Eperson d210u	17209	323
W5	Eperson e210u	17467	323
W6	Eperson f210u	17453	339
W7	Eperson g210u	17280	323
W8	Eperson h210u	17144	323
W9	Eperson i210u	17632	328
W10	Eperson j210u	17855	362
W11	Eperson k210u	23415	350
W12	Eperson l210u	23182	352
W13	Eperson m210u	23299	352
W14	Eperson n210u	23114	369
W15	Eperson o210u	23305	353
W16	Eperson p210u	23507	352
W17	Eperson q210u	23442	359
W18	Eperson r210u	23267	395
W19	Eperson s210u	23260	379
W20	Eperson t210u	22951	379
W21	Eperson u210u	22434	379
W1	Household a210u	12497	1754
W2	Household b210u	8394	1949
W3	Household c210u	8393	1701
W4	Household d210u	8280	1800
W5	Household e210u	8263	2222
W6	Household f210u	8114	2481
W7	Household g210u	7915	2236
W8	Household h210u	7940	2236

[7] For use within Australia only. Household files include limited data on non-responding households.
[8] Stata and SAS filenames have underscores instead of spaces.

Wave	Filename[8]	Cases	Variables
W9	Household i210u	8136	2423
W10	Household j210u	8244	2461
W11	Household k210u	11580	2250
W12	Household l210u	10641	2317
W13	Household m210u	10898	2438
W14	Household n210u	10873	2545
W15	Household o210u	10935	2201
W16	Household p210u	10910	2327
W17	Household q210u	10869	2458
W18	Household r210u	10645	2748
W19	Household s210u	10689	2407
W20	Household t210u	10712	2593
W21	Household u210u	10607	2652
W21	Longitudinal weights u210u	45570	1288
W21	Master u210u	45570	211
W1	Rperson a210u	13969	2511
W2	Rperson b210u	13041	3187
W3	Rperson c210u	12728	3429
W4	Rperson d210u	12408	3197
W5	Rperson e210u	12759	3587
W6	Rperson f210u	12905	3575
W7	Rperson g210u	12789	3724
W8	Rperson h210u	12785	3866
W9	Rperson i210u	13301	3697
W10	Rperson j210u	13526	3944
W11	Rperson k210u	17612	4355
W12	Rperson l210u	17475	4129
W13	Rperson m210u	17500	3933
W14	Rperson n210u	17511	4090
W15	Rperson o210u	17605	4515
W16	Rperson p210u	17693	4021
W17	Rperson q210u	17570	4153
W18	Rperson r210u	17434	4239
W19	Rperson s210u	17462	4765
W20	Rperson t210u	17070	4168
W21	Rperson u210u	16549	4228

Special Topic Modules

Wave 1 - History (family background/marital history/employment history).
Wave 2 - Household wealth.
Wave 3 - Retirement intentions and the impact of retirement.
Wave 4 - Youth issues; Health insurance; Disability.
Wave 5 - Fertility; Non co-residential relationships; Personality; Household expenditure; Intentions and plans.
Wave 6 - Household wealth.
Wave 7 - Retirement intentions and the impact of retirement; Diet; Smoking history.
Wave 8 - Fertility; Non co-residential relationships; Intentions and plans.
Wave 9 - Health (including Diet/Health Insurance/Disability); Personality.
Wave 10 - Household wealth.
Wave 11 - Fertility; Retirement intentions and the impact of retirement; Intentions and plans.
Wave 12 - Non co-residential relationships; Education, skills and abilities; Literacy and Numeracy; Cognitive Ability Tasks.
Wave 13 - Health (including Diet/Health Insurance/Disability); Personality.
Wave 14 - Household wealth; Material deprivation.
Wave 15 - Retirement; Fertility; Non co-residential relationships.
Wave 16 - Education, skills and abilities; Literacy and Numeracy; Cognitive Ability Tasks.
Wave 17 - Health (including Diet/Health Insurance/Disability); Personality.
Wave 18 - Household wealth; Material deprivation.
Wave 19 - Retirement; Fertility; Non co-residential relationships.
Wave 20 - Coronavirus; Education, skills and abilities; Literacy and Numeracy.
Wave 21 - Health (including Diet/Health Insurance/Disability); Personality.

For the special topic modules through wave 21, see the interview modules rubric http://melbourneinstitute.unimelb.edu.au/hilda/for-data-users/questionnaires-and-fieldwork-materials

Zip files

The **Documentation** zip contains:

- Subject Level Coding framework (pdf) for all variables; and a (summarized version) cross wave variable index, master file coding framework and longitudinal weights file coding framework.
- Marked-up questionnaires and showcards (pdfs) showing the questionnaires and associated variable names (excludes derived and history variables).

The **Frequencies** zip contains:

- Frequencies (Stata tabulate) are provided for each wave. String variables (ids, dates and timestamps) are usually excluded from the frequencies. Numeric variables are unweighted. Some variables may appear multiple times in a single frequency listing (for multiple populations: Eperson, Rperson or Household).
- Topcoding thresholds for financial questions.

The two **SAS** zips contain:

- SAS version 9 or later household-level and person-level data (as 32-bit datasets. These can be read by 64-bit SAS). To reduce SAS file sizes, the files have been compressed by SAS:

```
libname x "C:\SAS" compress=yes;
```

- SAS formats for each file and SAS program to create the formats (create formats *wrt*.sas) and template SAS programs to associate the formats with the SAS data files (read sas files *wrt*.sas).

Note: If you are using 64-bit SAS on 64-bit windows you may encounter this type of error: "ERROR: File WAVE1.EPERSON_AXXXC cannot be updated because its encoding does not match the session encoding or the file is in a format native to another host, such as WINDOWS_32". This error occurs because SAS will not allow a single directory to be designated as both 32- and 64-bit (and you are trying to read 32-bit files and write 64-bit files). One solution is to use SAS to copy the 32-bit HILDA files from a temporary directory into a final directory.

```
libname inlib cvp 'c:\temp';
libname outlib 'c:\hilda' outencoding='wlatin1';
proc copy noclone in=inlib out=outlib;
run;
```

The supplied "Read SAS files *wrt*.sas" programs can be run on the copied files.

The two **SPSS** zips contain:

SPSS version 12 or later system files for the household-level and person-level data.
The SPSS files have HILDA's global missing values (see list below) set to missing, which will cause cases with these values to be excluded from most procedures. To turn off this setting (for example if you need to include those who are coded as -3 Don't know and -4 No answer), use the SPSS command: "missing values *varnames* ()."

The two **STATA** zips contain:

Stata version 11 to 16 / .dta-format. If you are using Stata 11, 12 or 13 "update all" before opening the HILDA data files.

If you have Stata 10, install (ssc install use13) and use the USE13 command to open the files. If this gives the error "unexpected text found: s" install and use the USE12 command. If you have Stata 9, use the SaveTo9 tool http://www.radyakin.org/transfer/saveto9/cs/index.htm.

Note that the Combined files and Rperson files will require 'Stata/SE' or 'Stata/MP' as the number of variables exceeds the limit of 2047 variables for 'Intercooled Stata' (Stata/IC). The student version 'Small Stata' cannot open any of the HILDA datasets. If you are restricted to Stata/IC you can work within the 2047 variable limit by pre-specifying a list of variables names in the "use varnames using filename" command:

```
use xwaveid ktifefn ktifefp using "Rperson_k210c.dta"
```

Stata is case-sensitive, all HILDA variables names are lowercase.

In Stata, variable labels are truncated at 80 characters and HILDA variable labels often exceed this limit. The full variable labels can be seen in the coding frameworks or Stata notes field, view with the "notes varname" command. Value labels are not truncated.

Stata/SE/MP "set" commands may be required before using Combined or Rperson files:

```
set maxvar 6000
[set memory 200m for Stata 10/11]
```

It is not recommended to create a wide longitudinal file using all the Combined files with all variables, as this produces a file containing over 100,000 variables which consumes around 8G of memory. Newer versions of Stata (tested on Stata/MP 15) will allow you to do this, but older versions of Stata/SE/MP will run out of memory (on older computers) or exceed the limit of 32,767 variables. You should instead select the subset of variables you will be analysing, save

each file, then “merge 1:1 xwaveid using filename”; or you can merge files using a loop and the keepusing option for merge:

```
use "Master_u210u.dta"
forvalues i=1/21 {
   local w = word(c(alpha), `i')
   merge 1:1 xwaveid using "Combined_`w'210u.dta", ///
        gen(m`i') keepusing(`w'edhigh1)
}
```

If your computer has sufficient memory, you can create a long longitudinal file using the program on the HILDA website. On older computers it is usually not possible to create a long longitudinal file with ALL the combined file variables, as Stata/SE/MP runs out of memory. You could select a subset of analysis variables, strip the wave letter off the variable names, calculate a wave number and combine the resulting files.

Should you wish to create **R** files, you can do so by importing the HILDA Stata (*.dta) files:

```
library(foreign)
Combined_u210c <- read.dta(file="Combined_u210c.dta")
head(Combined_u210c)
```

Q Professional

The HILDA SPSS (*.sav) files can be read using File → Import New Data File (New Project). Note that Q is sensitive to compression and virus checking. In particular, check that you have unzipped the files correctly as described in the section “Extracting the zips”, above.

PanelWhiz

PanelWhiz metadata is no longer supplied in the standard HILDA *Combined* files. PanelWhiz is still available for the HILDA General and Restricted Release Stata files but first requires running the PanelWhiz setup programs. PanelWhiz updates for each release will be available through the usual PanelWhiz channels.

PanelWhiz was developed by Canadian economist Prof. John P. DeNew while working at the German Institute for Economic Research (DIW), Berlin (John is now a Professor at the University of Melbourne). Markus Hahn, while an economics student at the Ruhr-Universität Bochum contributed to many components of the package. PanelWhiz is charityware, and suggests the user make a direct donation to UNICEF. You can download PanelWhiz from http://www.panelwhiz.eu/.

Cross National Equivalence File (CNEF)

The CNEF_long (long-format) file is included in the SAS, STATA and SPSS zips. This file contains demographic, income and some health measures equivalised to match the structure of the German and American CNEF panel datasets. The HILDA CNEF file is supplied to encourage the exploration of this small but powerful dataset for cross-national comparisons of such things as poverty rates and income inequality. It is also possible to pair the various surveys by a common selection metric using the CNEF to select cases from the originating surveys. By application CNEF data can be obtained for other countries, such as the USA, Germany, the UK, Switzerland, Korea, Russia and Canada. The two HILDA-Cross-National-Equivalent-File-codeframes in the Documentation zip describe the algorithms used to convert HILDA variables into CNEF variables and the means or frequencies for each variable. http://cnef.ehe.osu.edu/

Global missing values used in HILDA

-1 Not asked (question skipped due to preceding answer)
-2 Not applicable
-3 Don't know
-4 Refused or not answered
-5 Multiple boxes marked (Self Completion Questionnaire only)
-6 Given value implausible: after intensive checking was overwritten
-7 Unable to determine value
-8 No matching Self Completion Questionnaire
-9 Non-responding household this wave
-10 Non-responding adult or child from a household with at least one responding person

Inclusion of these values without declaring them as missing may lead to spurious results as you will be including non-response as a component of your analysis.

Previous users please note

Users of the previous releases of HILDA are advised against mixing data from prior and current releases. Within a release the data in each wave is updated for cross-wave consistency, particularly with respect to age, sex, the income model, the imputation of income, wealth and expenditure and the methods used to construct the derived and history variables.

Data Use

Access to this data is conditional on you agreeing to the terms and conditions set out in the Confidentiality Deed Poll and having been approved to use the data by the Department of Social Services. Agreement to the terms of the Confidentiality Deed Poll is required as part of the process to submit the online application form for this data. See the *DSS Longitudinal Studies Data Access and Use Guidelines* for more details.

Contact details

If you have any problems extracting the data or documentation zips, have a query about a data anomaly or have difficulty in finding a variable, please submit your query via Dataverse. Click on the email icon button part way down the HILDA Dataverse page on the right or the 'Support' option at the top right of the page.