

# Key Longitudinal Variable Datasets

## Data Documentation

Compiled by Nicholas Egan with input from the Data Management Group (April 2023)

### Overview

The Key Longitudinal Variable (KLV) datasets are intended to save data users time by reducing duplication of work and programming errors. Until now, it was a substantial undertaking to generate longitudinal survey variables because of the different variable names at each survey wave for the original three cohorts.

There is one longitudinal KLV dataset per cohort; as of April 2023, the following survey waves are included: 1989-95 Waves 1-6, 1973-78 Waves 1-9, 1946-51 Waves 1-9 and 1921-26 Waves 1-6. The included variables are presented on the next page of this document. The raw survey variables used to derive each longitudinal variable can be identified by viewing the source code.

### Collapsed variables

We have included commonly used collapsed versions of some variables; these are denoted by the variable suffix `_#gp` where `#` refers to the number of levels in the collapsed variable. All variables with their original levels are retained in the survey datasets, in the event data users wish to use an alternative categorisation of a variable. The categories included in the collapsed versions of variables are provided in the table below.

### Missing values

Missing values were left unchanged. If the raw survey variable contained a missing value for a given survey, then the derived longitudinal variable is also set to missing for that survey.

### Values applied across surveys

Country of birth was only asked at one survey wave for each cohort. As this variable is time-invariant, it has been applied across all survey waves. Note that in the 1989-95 cohort, country of birth was asked at Wave 2, and so any participants who did not respond to Wave 2 will have missing data for country of birth.

Highest qualification was only asked at Wave 1 in the 1921-26 cohort and Wave 1 and Wave 6 in the 1946-51 cohort. As highest qualification is less likely to change in these older women, highest qualification is 'filled-in' for the later survey waves in the 1921-26 based on their response at Wave 1. Similarly, for the 1946-51 cohort, their Wave 1 response is used to fill-in Waves 1 to 5, and their Wave 6 response is used to fill-in Waves 7 to 9.

### Available formats

The KLV datasets are available in SAS, Stata and SPSS formats.

### Summary of variables and categories

The KLV datasets include the following variables presented in the table below. Cells shaded in yellow denote commonly used collapsed variables. SF-36 scores are not included in the 1989-95 KLV dataset.

Variable	Label	Categories
idalias/idproj	ID number	NA
survey	Survey wave	NA
age	Participant age at time of survey return	NA
ariapgp	Area of residence (original levels)	1='Major cities' 2='Inner regional' 3='Outer regional' 4='Remote' 5='Very remote' 6='Overseas'
marital	Marital status (original levels)	1='Married' 2='Defacto' 3='Widowed' 4='Separated' 5='Divorced' 6='Never married'
smokst	Smoking status (original levels)	1='Never-smoker' 2='Ex-smoker' 3='Smoker, less than 10 per day' 4='Smoker, 10-19 per day' 5='Smoker, 20 or more per day' 6='Smoker, unknown cigarettes per day'
alcnhmrc	Alcohol consumption (original levels)	1='Low risk drinker' 2='Non-drinker' 3='Rarely drinks' 4='Risky drinker' 5='High risk drinker'
mh	SF-36 Mental Health score	NA
pf	SF-36 Physical Functioning score	NA
srh	Self-rated health (original levels)	1='Excellent' 2='Very good' 3='Good' 4='Fair' 5='Poor'
hq	Highest qualification (original levels)	Varies by cohort see Data Dictionary.
moi	Ability to manage on available income (original levels)	1='Easy' 2='Not too bad' 3='Difficult sometimes' 4='Difficult always' 5='Impossible'
mnstrs	Mean stress score	NA
bmigroup	Body mass index (WHO categories for people of non-Asian ethnicity)	1='Underweight (<18.5 kg/m <sup>2</sup> )' 2='Acceptable weight (>=18.5 to <25 kg/m <sup>2</sup> )' 3='Overweight (>=25 to <30 kg/m <sup>2</sup> )' 4='Obese (>30 kg/m <sup>2</sup> )'

Variable	Label	Categories
hcc	Health care card	0='No' 1='Yes'
cob	Country of birth (original levels)	Varies by cohort see Data Dictionary.
ariapgp_3gp	Area of residence (3 levels)	1='Major cities/overseas' 2='Inner regional' 3='Outer regional/remote/very remote'
income_3gp	Ability to manage on available income (3 levels)	1='Not too bad/easy' 2='Difficult sometimes' 3='Difficult always/impossible'
marital_2gp	Marital status (2 levels)	0='Partnered' 1='Non-partnered'
alcnhmrc_3gp	Alcohol consumption (3 levels)	1='Low risk drinker' 2='Non-drinker/rarely drinks' 3='Risky/high risk drinker'
smokst_3gp	Smoking status (3 levels)	1='Non-smoker' 2='Ex-smoker' 3='Current smoker'
srh_3gp	Self-rated health (3 levels)	1='Excellent' 2='Very good/good' 3='Fair/poor'
hq_3gp	Highest qualification (3 levels)	1='Year 12 or below' 2='Certificate/diploma' 3='University'
hq_4gp	Highest qualification (4 levels)	1='Less than Year 12' 2='Year 12 or equivalent' 3='Certificate/diploma' 4='University'
cob_3gp	Country of birth (3 levels)	1='Australian born' 2='Other English-speaking country' 3='Other'