

# The Irish Longitudinal Study on Ageing (TILDA) Pseudonymised Microdata File

## Frequently Asked Questions

If you have any queries on the The Irish Longitudinal Study on Ageing (TILDA) Pseudonymised Microdata File (PMF) please read through these frequently asked questions (FAQs) which cover the most common queries received. If your query is not answered here, please contact [issda@ucd.ie](mailto:issda@ucd.ie).

### **DATA ACCESS OUTSIDE OF EUROPE**

*Q. Can I access the TILDA data from the United States or other countries outside of Europe?*

A. Due to the differing data protection legislation across jurisdictions, the TILDA PMF datasets are only available to those who reside within the European Economic Area (EEA). Requests to access the PMF datasets from outside of the EEA will be forwarded to the TILDA team for review on a case-by-case basis.

### **VARIABLES PRESENT IN EACH WAVE**

*Q. Is there a way to check what data is collected in each different wave of TILDA?*

A. As well as the documentation included with each release of data on ISSDA detailing what is contained in each dataset, TILDA also has several resources on its website to allow you to see what data has been collected in each wave of the study. One such resource is a searchable database of questions that appear in the TILDA CAPI and SCQ across each wave, which can be found here: <https://tilda.tcd.ie/data/questionnaire/QuestionnaireHarmonisation/index.php>.

This database is searchable by variable names as well as keywords and highlights which waves each variable is included in or omitted from.

You can also find more information at <https://tilda.tcd.ie/data/questionnaire/> that include resources such as an Interactive Dataset Map. This page also includes metadata catalogues that are currently being generated for all waves of TILDA.

## **AVAILABLE VARIABLE DOCUMENTATION**

*Q. I am unsure of what some of the variables mean. Where can I find more information?*

A. Along with the individual PMF dataset releases, we include documentation to assist with understanding the contents of the datasets. The TILDA release guide contains introductory information about the study as well as information about any changes made to the PMF datasets. We have also included a derived variables codebook which outlines additional variables generated for the datasets with notes on how these variables were generated. We also include copies of the CAPI questionnaire and the Self Completions Questionnaires for each wave of the study which detail the question name, question wording, response options and routing specifications for each question.

## **ACCESSING ADDITIONAL VARIABLES**

*Q. I would like to use a variable that isn't in the PMF dataset. Is it possible to send any additional variables for my research?*

A. We are unable to facilitate requests to send additional variables that are not currently present in the PMF. All requests for additional variables reviewed for the next release of the PMF. To access variables not available in the PMF, an application can be made to access the RMF files via the TILDA hotdesk system).

## **RESEARCHER MICRODATA FILES (RMF) – TILDA HOTDESKS**

*Q. Is it possible to access the Researcher Microdata Files (RMF) for TILDA?*

A. TILDA have RMF versions of the datasets available to access through the on-site hotdesk system based in the TILDA offices in Trinity College Dublin. All proposals to access the RMF are reviewed individually, and researchers will be given access to only the data required and requested for their research. If a researcher wishes to use the RMF, they can go to <https://tilda.tcd.ie/data/accessing-data/hotdesk/index.php> where they can find further information on the application process as well as the TILDA Hot desk application form.

## **MEDICAL CONDITIONS**

*Q. I am looking at medical conditions from Wave 1 but can't find them in Wave 2. Is there a way to access these?*

A. From Wave 2 onwards, incidence conditions are release using ICD-10 classification codes. This is due to the low incidence rates of changes between waves for individual conditions. To see the individual conditions for the later waves of TILDA, an application to request access to the researcher microdata files (RMF) can be made.

## **CAGE QUESTIONNAIRE SCALE**

*Q. What is the scale used as BEHcage: CAGE alcohol scale?*

A. The BEHcage scale is the CAGE Questionnaire scale which is asked in the SCQ part of TILDA. It is used for screening alcoholism. The questions are –

- § Have you ever felt you needed to cut down on your drinking?
- § Have people annoyed you by criticizing your drinking?
- § Have you ever felt guilty about drinking?
- § Have you ever felt you needed a drink first thing in the morning (eye-opener) to steady your nerves or to get rid of a hangover?

Those who answer yes to any question are given a +1 to their score. Those who have a missing value (don't know / refused) for any of the questions are excluded from the final scoring of the scale. There are also a number of missing values for respondents who did not fill out the SCQ booklet. The four original variables used to create the scored variable are SCQcage1, SCQcage2, SCQcage3 and SCQcage4. The study found here - <http://www.sciencedirect.com/science/article/pii/S0140673682915793> outlines how the scoring works for identification of excessive drinking and alcoholism.

## **CHILDREN IN HOUSEHOLDS**

*Q1. I cannot figure out the children living in the household because variables cm010\_\* are missing.*

*2. I cannot get the age of children who live outside the household since the variables cs031\_\* are missing. I have no idea how I can get such information.*

A. These variables were removed from the public dataset as there was a potential to identify participants by using them in conjunction with other variables. It was decided that although we understand these omitted variables reduce the quality of the dataset in terms of the research that can be carried out, the risk to data protection must take priority.

## **SELF-RATED SENSES**

*Q. Self-rated vision, hearing, taste and smell have been used twice in the questionnaire as well as the derived variables. I am confused which to use? What is the difference between them?*

A. Self-rated vision

PH102 asks respondents to rate their general vision. PH103 asks to rate their vision at a distance and PH104 asks to rate vision close up.

DISvision is identical to PH102 and was only derived to group a number of the disability related questions within the questionnaire (those with DIS as the prefix).

The response options are slightly different in PH102 and DISvision than in the questionnaire. 1 – 4 stay the same as excellent, very good, good and fair while 5 and 6 have been grouped into one option of 5 for those who rate their vision as poor or legally blind. This is for anonymity purposes with the publically available data.

PH102 or DISvision can be used for general investigation into respondent's vision while PH103 and PH104 can be used if you want to investigate anything specific to close or distance vision problems.

#### Self-rated hearing

PH108 = DIShearing (General self-rated hearing from excellent to poor)

DIShearing and PH108 are identical variables derived with the DIS prefix to group them with other disability related variables. Either can be used for general investigation into respondents' level of hearing.

PH109 looks at difficulty following conversation with one person while PH110 looks at difficulty following conversation with four people. These would be used for more specific hearing related research.

#### Self-rated taste/smell

DISsmell is derived from PH112 to group the DISsmell variable with other disability related variables

DIStaste is derived from PH113 to group the DIStaste variable with other disability related variables

For the DISsmell and DIStaste variables, the only difference from their original variables are that the don't know responses from PH112 and PH113 have been coded as general missing responses (.).

### **VERBAL TASKS**

*Q. Word list learning is presented from code ph117-ph120, where people are asked to recall a list of words which are presented to them. The values 98 and 99 indicate 'don't know' and 'refusal to answer' but there are quite a lot of responses coded as -1. Please advise what this code means as I would suppose that if people did not remember any of the words on the list they would have scored zero. Is this correct?*

A. With regards to the ph117-ph120, the reason for the '-1' score relates to the variable called ph116. This variable states whether the respondent had the list of words read out to them by computer, or whether it was read out to them by the interviewer. If the words were read out by the computer, then the respondents scores were recorded in ph117 and ph118. If the words were read out by the interviewer, then the respondents score were recorded in ph119 and

ph120. There should be no respondent who has answered all four variables. The score of '-1' simply denotes that the respondent did not record anything for that variable. I hope that clears it up.

*Q. In the verbal fluency task data is presented that indicates how many animals were named in one minute. I wonder if data is available which gives a breakdown of the actual animals recalled, in what order and what timeframe within the one minute.*

A. Unfortunately, in the verbal fluency data, there is no breakdown of actual animals recalled, in what order and the timeframe. The interviewers were not told to do this, so we do not have any data available. It would have added time to the interview, which we could not afford.

## **WIDOWHOOD**

*Q. For the individuals who report that they are widowed, I am looking for year of widowhood to construct a variable that reflects the duration of widowhood at baseline. I can't seem to find anything that relates to year the individual lost their spouse/partner. Please advise.*

A. While we do collect the year the person became widowed, but we do not include it in the PMF as it was deemed too identifiable for inclusion.

## **RETIREMENT**

*Q. Am I correct in saying that the questions asking respondents for their reason for retirement (we604/605) were only asked in wave 1?*

A. In the public archive datasets, the reasons for retirement questions were fully included in Wave 1. These were coded as variables ranging from we605\_01 to we605\_13. These variables were not included in the Wave 2 public archive dataset.

In the Wave 3, Wave 4 and Wave 5 datasets, the following reason for retirement variables were included: we605\_01, we605\_04, we605\_06, we605\_07, we605\_95, we605\_98, we605\_99

From wave 3 onward, we updated our variable naming protocol so that variables with a response of "other", "don't know" and "refused" were given the suffixes of 95, 98 & 99 respectively. Due to this, you may want to recode variables we605\_11, we605\_12 & we605\_13 from the wave 1 archive to match this if you wanted to do direct comparisons across the different datasets. With each of the datasets, there is a variable anonymisation actions file that can be referred to check the status of each variable in the public dataset. If the variable has been dropped for example, it will be noted here.

## **BEREAVEMENT**

*Q. I wanted to add a control variable for if the respondent has recently experienced the death of a loved one. I was thinking of using if martial status (cs006) and number of living children*

*(cn002a) changes between waves but it looks like these variables were only asked in wave 1 as well?*

A. With regards to marital status, we have included a derived variable called “married” in each of the archive datasets. This variable, however, may not be a suitable proxy indicator for the death of a loved one, as you wouldn’t be able to distinguish between a participant who got divorced or a participant who remarried in the period between each wave of data collection.

The variable regarding the number of living children is only included in the first wave of the public archive dataset and not in any of the others.

We do however include variables “SOCpalive, SOCpaliveW2, SOCpaliveW3, SOCpaliveW4 & SOCpaliveW5” to track the living status of each participants parents across the wave so this may be a more suitable proxy indicator for the loss of a loved one.

While TILDA does gather more information regarding the death of loved ones, this isn’t included in the public archive datasets currently, as the data is considered identifiable and not appropriate for the public dataset. This will be noted in the variable anonymisations actions document.

## **PERSONALITY**

*Q. I see that wave 2 asked questions on personality, I can’t seem to find these in the user guide. Are there any derived variables in the dataset that would give me a measure of the big-five personality traits?*

A. While TILDA does record the Revised NEO Personality Inventory, this isn’t currently available in the public archives.