

ICPSR 34315

The Irish Longitudinal Study on Ageing (TILDA), 2009-2011

Rose Anne Kenny
Trinity College Dublin (Ireland)

Release Notes

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Terms of Use

The terms of use for this study can be found at:
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34315/terms>

Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

NOTICE

WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

THE IRISH LONGITUDINAL STUDY ON AGEING

TILDA Wave 1 Anonymised Data Release Notes

Version 1 – February 2012

1 Frequently Asked Questions

1.1 What is TILDA?

The Irish Longitudinal Study on Ageing is a nationally representative study of the population of Ireland aged 50 and above. TILDA aims to understand how the health, social and financial circumstances of the older Irish population and how these factors interact.

The first wave of data collection was conducted between October 2009 and July 2011. In total, 8175 individuals aged 50 and over participated in the study. 329 interviews were also conducted with younger spouses or partners of participants, leading to a total sample size of 8504. The second wave of TILDA interviews will be undertaken in 2012.

The design of TILDA is described in full elsewhere (see documentation section below) but in brief, each participant underwent a home interview administered by a trained interviewer, was asked to complete and return a questionnaire including more sensitive questions, and was invited to undertake a health assessment, either at a dedicated centre or in their own home if travel was impracticable.

1.2 Who supports TILDA?

TILDA is based at Trinity College Dublin, and involves many scientific collaborators within Ireland and internationally. TILDA is funded by the Department for Health and Children, Irish Life and The Atlantic Philanthropies.

Ethical approval was granted by the Trinity College Research Ethics Committee.

1.3 What does the first public release dataset include?

The first public release includes data from all 8504 TILDA participants. The data included in the first release includes that from the home interview, the self-completion questionnaire, and certain other variables derived from these data.

The data are fully described in the documentation (see documentation section below). A brief description follows:

Identifiers and audit data (id-nml). A unique identifier for each participant is provided, along with identifiers for the household and geographic cluster to which they belong, along with some basic demographic information and variables relevant to complex sample analysis (see section 3 below).

Computer Aided Personal Interview (CAPI) data (cm003-x14d). All of the information collected as part of the initial home interview are included, subject to the restrictions described below. See the questionnaires and codebook for more information on these.

Self-completion questionnaire (SCQ) data (SCQsocact1-SCQageprc32). Each participant was left a questionnaire to be completed and returned to the TILDA offices. In total, 7191 (85%) participants returned the SCQ. Data from the SCQ is included in the first public release, the

name of each variable containing SCQ data is prefixed 'SCQ'. See the questionnaires and codebook for more information on these.

Derived variables (DISimpairments-BEHcage). Several commonly used scales have been constructed based on the CAPI and SCQ data and these are included along with other derived variables. See the derived variable codebook for more information on these.

1.4 What is not included?

Data that is highly sensitive or that is potentially identifiable is not included in the public release dataset. In addition, some variables are recoded in order to avoid possible identification of individuals (see section 4 below).

At present, data on specific medication use, data from the health assessment centre and the health assessment conducted in the participant's home (including data from blood samples) are not included. This data is currently being prepared for release and access to this data is restricted to those contributing to the process of cleaning and preparation for public release being conducted on site at Trinity College Dublin. Contact the TILDA team for information on preliminary access to these data and how to assist with the cleaning process.

1.5 What other documentation is available?

Design Report. The Design Report published in 2009 describes the design of TILDA and the motivation for the selection of each of the assessments undertaken in the study.

First Findings Report. A first findings report based on a preliminary version of the TILDA dataset was published in May 2011 and is available on the TILDA website.

Codebook. A codebook including meta-data for all variables is available. The **derived variable codebook** includes notes on each of the derived variables and how each was created.

Questionnaires. The TILDA wave 1 CAPI questionnaire and SCQ are available from the TILDA website.

Cohort Profile. See Kearney et al (2011) *Int. J. Epidemiol.* 40 (4): 877-884 for a description of the cohort and the assessment.

Details of the anonymisation process: a list of the variables that were removed from the first public release, those that were modified and how such modification was conducted.

1.6 Who do I contact to report an error or for more information?

This data is supplied without any guarantee of its accuracy. It is possible that errors and inconsistencies exist with the dataset. Future releases of the dataset will address any issues that we are made aware of.

Contact the TILDA team (tilda@tcd.ie) to report errors in the data or for more information.

2 Target population and Design of TILDA

The sampling framework and design of each component of TILDA is described in detail in the TILDA Design Report¹, and is briefly summarised below. The TILDA target population includes all members of the population of Ireland who are 50 years old or over and who live in the community (that is they do not live in a long term care institution). While only around 1% those between the ages of 65 and 74 currently live in long term care, this figure rises to around 6% of those aged between 75 and 84 and 21% of those aged 85 and over.

To generate the TILDA sample, all postal addresses in Ireland were assigned to one of 3155 geographic clusters, and a sample of 640 of these clusters was selected, stratified by socio-economic group and geography to maintain a population representative sample. Clusters were selected with a probability proportional to the number of individuals aged 50 and over in each cluster. Forty households were selected from each cluster (it was estimated that 25600 addresses in total would be required to achieve the required sample size of 8000). Each of the selected addresses was visited by an interviewer, who attempted to ascertain the eligibility of the address, to contact a household member and determine whether any individuals aged 50 or over lived at that address. All individuals aged 50 or over in each selected household and their partners (even if aged less than 50 themselves) were invited to be included in the study.

2.1 Fieldwork report and response rate

A total of 8175 interviews were conducted with respondents aged 50 and over belonging to 6279 households. An additional 329 interviews were conducted with younger partners of eligible individuals. The first interview took place on 18 October 2009, with a steady accrual until the last interview was conducted on 22 February 2011.

The response rate is the proportion of selected households including an eligible participant from which an interview was successfully obtained. Interviewers were sent to all of the initially allocated 25600 addresses. Of these, 22321 were occupied residential addresses. At 11819 addresses contact was made and it was determined that no person aged 50 or over was at that address. In 9818 it was determined that there was a person aged 50 or over. At 684 addresses either no contact was made or contact was made but it was impossible to determine whether there was anybody over 50 living at that address. Based on those households in which eligibility was determined, it is estimated that $9818 / (9818 + 11819) \times 684 = 310.4$ of those households were eligible.

The estimated number of selected eligible households is therefore $9818 + 310.4 = 10128.4$. Successful interviews were obtained in 6279 households, leading to a response rate of 62.0%.

2.2 Components of the study

¹ Available from www.tilda.ie

The initial respondent (first eligible household member interviewed) in each household provided details of all of household members. Any household members eligible for the study were subsequently invited to take part in the study.

Each individual agreeing to participate in the study underwent a structured 'CAPI' interview in their own home with a trained interviewer, which included questions on many domains of health, wellbeing, family and financial circumstances

Where two participants were married or were living together as if married, a **'financial respondent'** and a **'family respondent'** were identified, providing the detailed responses on family and financial circumstances respectively. The financial and family respondents were not necessarily different individuals.

Each participant was invited to undergo a health assessment, either a full health assessment at a centre in Dublin or Cork or a partial assessment in their own home where travel to either centre was not practicable.

Each participant was also left a 'self-completion questionnaire' including potentially sensitive questions for them to fill in and return to TILDA by mail. This included a range of questions on quality of relationships, quality of life, perceptions of ageing, emotional wellbeing and health behaviours. There was also a single blank page for each respondent to make any further comment they chose.

The detailed design of each component of the study, including the rationale for the design and the comparability with other European and international studies is described in the TILDA design report.

3 Applying Complex Survey Analysis Methods

3.1 Weights

For inferences based on the TILDA dataset to be applicable to the Irish population, weights must be applied to correct for selection bias. These weights are supplied with the dataset; a description of how they are calculated follows:

The weight for each participant is equal to the number of individuals in the population represented in the study by that participant. Those individuals who come from groups less likely to participate in the study therefore have a higher weight. A 'weighted' estimate based on the sample is an unbiased estimate of the respective parameter in the population.

There are two sets of weights supplied in the dataset. The first is the 'CAPI' weight, to be applied when the whole TILDA sample is included in an analysis. This is equal to the number of individuals in the population (based on the numbers in the 2011 census) represented by each individual among the 8175 individuals that underwent the home 'CAPI' interview. The second weight is the self-completion questionnaire (SCQ) weight, to be applied when only those who returned the SCQ are included. This takes into account the CAPI weight as

described above, but also the fact that some subgroups of the sample were more likely to return their SCQ than others.

CAPI Weight

The CAPI weight is calculated by comparing the distributions of age, sex, education, marital status and geographic location in the sample to those derived from census data. The external sources of information used to calculate the weights were:

- The distribution of age (50-64/65-74/75+), sex, marital status (single/widowed/married/divorced or separated) and geographic location (Dublin/other city or town/other rural area) from the 2006 census.
- The distribution of age (50-64/65-74/75+), sex and educational attainment (primary/secondary/tertiary+) from the 2010 Quarterly National Household Survey
- The distribution of age (50-64/65-74/75+) and sex from the preliminary results of the 2011 census.

The weighting algorithm rakes over the first two sources using twenty iterations, leading to a weight that yields the correct distribution of marital status, educational attainment and geographic location within each age/sex group. Finally the weight is adjusted to match the numbers in each age/sex group to those obtained from the 2011 census of the Irish population.

Therefore the numbers of people and the distribution of age and sex match those from the 2011 census, and the distributions of marital status, educational attainment and geographic location within those groups are assumed to be as they were in 2006/2010. As more information is released from the 2011 census, these weights will be updated.

SCQ Weight

The SCQ weight inflates each CAPI weight by the probability that each individual returned the SCQ. In total, 7191 (85%) of participants returned the SCQ. Multiple logistic regression was used to generate a predicted probability of returning the SCQ for each individual, based on age, educational attainment, marital status, whether the participant agreed to health assessment, and the participants prospective memory test results, immediate recall score, disability, and depression. Each of these factors independently predicts returning the questionnaire. This probability ranged from 0.36 to 0.95.

The SCQ weight is created by dividing the CAPI weight by this predicted probability, and incorporates a minor rescaling adjustment to ensure that the total size of the weighted population remains the same.

Variables

The variables '**capiweight**' and '**scqweight**' include the weights calculated as described above. Applying these weights to analyses yields estimates that are applicable to the Irish population in 2011. Weights are only supplied for the 8175 participants aged 50 or over.

Those aged less than 50 should not be directly included in analyses aimed at yielding estimates applicable to the general population.

The code used to generate the weights is available on request.

3.2 Clusters

The TILDA sample was recruited from households selected in geographic clusters, and when a household was selected every eligible member of that household was invited to participate. Failing to take into account the correlation between participants introduced by this sampling design will lead to biased estimates of the precision of estimates.

The two variables supplied in the dataset that indicate the cluster and household to which participants belong are '**cluster**' and '**household**'.

3.3 Stratification

The selection of geographic clusters was stratified, so that equal numbers of clusters were selected from each of three socio-economic groups. The socio-economic status of a cluster was defined by the proportion of individuals in that cluster. The variable '**stratum**' indicates to which of the three strata the cluster from which each participant was recruited belonged.

4 Anonymisation Process

The TILDA public release dataset was anonymised in collaboration with the Central Statistics Office. The criteria for anonymization included the following rules:

- No sensitive data was to be released
- Any potentially identifiable data, that is data based on which an individual could be identified, on its own or in combination with other publically available data sources was to be either top-coded, grouped or dropped completely
- It would not be possible to eliminate self-identification. The process to do this would be to eliminate a large amount of the data, and as such would greatly diminish the quality of the data released.

Top-coding: This was performed on variables where extreme results occurred at the end of the scale. As such, all respondents who answered over a given amount or value were grouped together, so as to eliminate identification.

Grouping: This was performed on variables where very specific answers occurred at different stages in the variable. As such, the decision was taken to band answers together, so as to eliminate identification.

In both the top-coding and grouping phases of anonymization, the criteria for banding and grouping answers together came from the criteria that there should be no fewer than 20 respondents under a given category for each variable. Less than 20 would be deemed to be too identifiable, and as such, the variable was either top-coded or recoded.

A spreadsheet detailing the variables that were deleted, top-coded or grouped prior to this release are available.