

ELSA
50+ Health & Life

NatCen
Social Research

User Guide to the Main Interview Datasets

Waves 1 to 9

Authors: NatCen Social Research

Date: May 2020 (version 1)

Contents

1 - Overview of the survey	3
1.1 Background and aims.....	3
1.2 Ethical clearance.....	4
1.3 Contact details	4
2 - Sample	5
2.1 Sample design.....	5
2.2 Refreshment samples.....	5
2.3 Sample status and eligibility	6
2.3.1 Household splits and eligibility.....	8
2.4 Proxies 8	
2.5 Institution interviews	9
3 - Survey content	10
3.1 Development and piloting	10
3.2 Details of survey content – overview	10
3.2.1 End of Life interviews	11
3.2.2 End of Life interviews	11
3.2.3 Follow-up nurse visit.....	11
3.3 Questionnaire content – Waves 1 to 9	11
4 - Fieldwork and response	14
4.1 Fieldwork – overview	14
4.2 Fieldwork – tracing movers.....	14
4.3 Fieldwork periods	15
4.4 Number of interviews.....	16
4.5 Self-completions in ELSA	16
4.5.1 Main self-completion (all Waves).....	16
4.5.2 Health and Work self-completion at Wave 3	16
4.5.3 Sexual activity and relationships self-completion at Waves 6 and 817	
4.5.4 Two versions of the Wave 7 self-completion.....	17
5 - Weighting, clustering and stratification	18
5.1 Clustering and stratification	18
5.2 Weighting for Waves 1 and 2	18
5.3 Weighting for Waves 3–9.....	19
5.4 Longitudinal weights Waves 3–9.....	19
5.5 Cross-sectional weights Waves 3-9	20
5.6 Self-completion weights	21
5.7 Nurse data weights	21
6 - Dataset information	22
6.1 Datafiles available at the UKDS	22
6.1.1 Main interview datasets	22

6.1.2 Other ELSA datasets at the UKDS	22
6.1.3 Geographic variables	23
6.1.4 Access to non-archived data.....	23
6.2 Order and content of the main interview dataset	23
6.2.1 Dropped variables	24
6.3 Household level modules	24
6.4 Serial numbering	25
6.5 Missing values	25
6.6 Multi-coded and dichotomous variables – Wave 3 onwards	26
6.7 Outcome code variables	27
6.8 Questionnaire errors	28
6.8.1 Wave 8	28
6.9 Coding and editing	29
Appendices	31
➤ Notes about particular variables in the Health module.....	31
➤ Notes about particular variables in the Risk module (Wave 5 only)	36

1 - Overview of the survey

This User Guide provides detailed information about the first eight waves of the ELSA data, all of which are available to download from the UK Data Service [here](#).

1.1 Background and aims

The English Longitudinal Study of Ageing (ELSA) began in 2002. It is a large scale longitudinal panel study of people aged 50 and over and their partners, living in private households in England. The original sample was drawn from households that had previously responded to the Health Survey for England (HSE) between 1998 and 2001. The sample has been refreshed at several waves (waves 3, 4, 6, 7 and 9) so not all respondents have participated since 2002 (see Section 2 for more details).

The same group of respondents have been interviewed at two-yearly interviews, known as 'waves', to measure changes in their health, economic and social circumstances. ELSA can complete the picture of what it means to grow older in the 21st century, and help us understand what accounts for the variety of patterns that are seen.

Although new topics can be introduced at different waves, every module has been reviewed to ensure that it will provide data that can measure change over time. This is achieved by repeating some measures exactly, by asking directly about change and by adopting questions to allow people to update or amend past responses. The information collected provides data about:

- Household and individual demographics
- Health – physical and psychosocial
- Social care (from Wave 6)
- Work and pensions
- Income and assets
- Housing
- Cognitive function
- Social participation
- Effort and Reward (voluntary work and caring)
- Expectations
- Walking speed
- Weight

In addition, certain waves contain one-off modules and questions as outlined in section 3.

For a tabular overview of the topics included on ELSA until Wave 9 please see [here](#) or the Appendix A of the ELSA Wave 9 Questionnaire & Data Documentation available on the [UKDS](#).

ELSA is the result of collaboration between University College London (UCL), the Institute for Fiscal Studies (IFS), the University of Manchester and NatCen Social Research. Other academic collaborators based at the Universities of Cambridge, Exeter and the University of East Anglia provided expert advice on specific modules. Funding for the first eight waves of ELSA has been provided by the US National Institute on Aging, and a consortium of British Government departments including:

- Department of Health;
- Department for Transport;

- Department for Work and Pensions;
- Communities and Local Government (formerly Office of the Deputy Prime Minister);
- HM Treasury;
- Department of Environment, Food and Rural Affairs;
- HMRC (formerly Inland Revenue and HM Customs and Excise);
- Office for National Statistics.

Many of the measures adopted in ELSA are comparable with measures used in the US Health Retirement Study (HRS) and the Survey of Health, Ageing and Retirement in Europe (SHARE).

1.2 Ethical clearance

Ethical approval for all the ELSA waves was granted from NHS Research Ethics Committees under the National Research and Ethics Service (NRES). For further information see here:

<http://www.nres.nhs.uk/>

1.3 Contact details

Any queries related to this study or the data should be sent to: elsadata@natcen.ac.uk

2 - Sample

2.1 Sample design

The ELSA sample has been designed to represent people aged 50 and over, living in private households in England. The sample is based on respondents who participated in the Health Survey for England (HSE). The original sample was selected from three years of HSE: 1998, 1999 and 2001. These years were chosen because they were recent and could provide a sufficiently large sample size. ELSA used the core samples for these years, all of which were nationally representative. The HSE 1999 sample design also included a boost sample that represented ethnic minorities. Because of funding constraints, it was not possible to follow-up the boost sample and it was discarded. Together these three HSE years contained 23,132 responding households.

Households were removed from the HSE sampling frame for ELSA Wave 1 if it was known that there was no adult of 50 years or older in the household who had agreed to be re-contacted at some time in the future. Individuals in the remaining households provided the basis for the ELSA Wave 1 sample (11,578 households containing 18,813 eligible individuals). The Wave 1 Technical Report, available from the UK Data Service, provides more details.

2.2 Refreshment samples

ELSA is a study of people aged 50 and over. As the study progresses, the sample members get older leaving the youngest ages unrepresented unless new sample members are recruited to fill the gap. Those aged 50-51 at any given wave will be 52-53 by the next wave, therefore a refreshment sample of people in their early 50s will be needed if the sample is to fully represent those aged 50+.

The ELSA sample has been refreshed five times now, at waves 3, 4, 6, 7 and 9. When a refreshment sample is drawn, households from HSE are selected for ELSA if at least one HSE interview was conducted with an age eligible respondent who agreed to be re-contacted.

Table 2.1 summarises the HSE years of sample origin, the eligible age ranges and the number of people interviewed in each cohort at the initial wave.¹

The W1 original sample was drawn from HSE respondents aged 50+ on the 1st March 2002. As ELSA is a study of those aged 50+, it needs to be periodically refreshed with additional sample aged from 50. Subsequent refreshment samples were drawn from other HSE years with differing age criteria to correct the age profile of the resulting cohort. Each refreshment sample added then becomes part of the cohort issued again at subsequent waves. The table below summarises the original and refreshment samples drawn from HSE for the ELSA sample.

¹ A small number of respondents who were previously classified in the original sample taken from HSE 2001 as a 'younger partner' became eligible under the Wave 3 Refreshment sample rules to become an ELSA Core Member. Please see the Wave 3 Technical Report for further details.

Table 2.1: Cohorts and source HSE years

Cohort	Invited to join ELSA	Origin HSE years	Age on 1st March of Y1
1	2002/3	1998, 1999, 2001	50+
3	2006/7	2001, 2002, 2003, 2004	50-52
4	2008/9	2006	50-74
6	2012/3	2009, 2010, 2011	50-55
7	2014/5	2011, 2012	50-51
9	2018/9	2013, 2014, 2015	50-53

All the age eligible people in the household that responded to HSE and also conduct an ELSA interview in the initial wave when invited to joint ELSA become what is known as ELSA 'Core Members', discussed further in the next section.

For further information about the refreshment samples in waves 3, 4, 6, 7 and 9, please refer to the Technical Reports for the relevant waves, available at:

<http://www.elsa-project.ac.uk/publications/case/technical>.

2.3 Sample status and eligibility

Across all waves, different types of individuals have taken part in an ELSA interview. The variable **FINSTAT**² indicates the detailed status of the individual by status and cohort. It indicates when the person joined ELSA and their status when they did so (e.g. Cohort 1 Core Member = C1CM; Cohort 1 Younger Partner = C1YP; Cohort 1 Older Partner = C1OP; Cohort 1 New Partner = C1NP).

In addition, a summary variable of sample member types is also available in each wave's file (**SAMPSTA** in Waves 2 and 3 and **SAMPTYP** in Waves 4 to 9³).

Diagram 2.1 below provides a visual summary of the sample type assignment rules on ELSA.

At the heart of eligibility to take part in ELSA were so-called ELSA '**Core Members**'. As the diagram illustration, ELSA core members have each met three criteria:

1. Fitted the age eligibility criteria of a given ELSA cohort
2. Participated in the sample-origin HSE survey
3. Participated in the first wave of ELSA when invited to join the study.

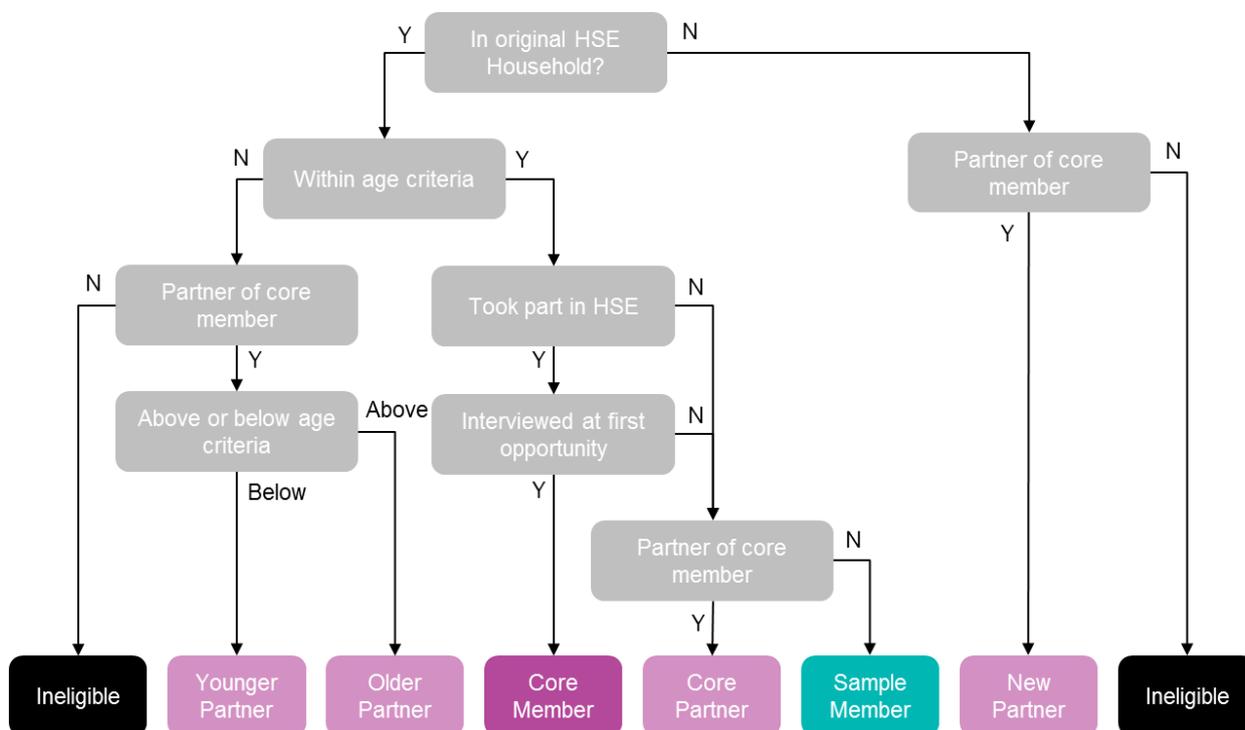
Core Members have remained eligible for an ELSA interview (personal or proxy) over the waves, as long as they have not died or moved outside of Great Britain. Core Members remained eligible if they had moved to Wales or Scotland. Core Members have remained

² The finstat variable name has a wave identifier in some files e.g. Wave 7 'finstatw7'.

³ In some waves, new partners were grouped into the summary sample type categories of sampsta/samptyp, while other times they were incorporated into them. From Wave 8 the latter approach is being taken consistently.

eligible if they have moved to an institution from their original residential address (within Great Britain).

Diagram 2.1: ELSA sample type assignment rules



In addition to Core Members, all **cohabiting partners** of Core Members (who were not Core Members themselves) have also always been eligible to take part. These ELSA ‘Partners’ have further been categorised into four different types to illustrate their relative age range and duration of co-habitation with the Core Member:

- Partners already present at the time of the HSE interview were categorised as either ‘**Core Partners**’ (age eligible but missed the baseline HSE and/or initial wave ELSA interview), ‘**Young Partners**’ (younger than the eligible age range at initial wave) or ‘**Old Partners**’ (older than the eligible age range at initial wave)
- Partners of any age who joined the household after the initial HSE interview are called ‘**New Partners**’, with the finstat variable indicating the wave at which they started cohabitation with the Core Member (e.g. C3NP5 finstat value represents a Cohort 3 New Partner joining at Wave 5).

Finally, people who were age eligible to become Core Members, but were not productive nor cohabiting with a participating Core Member at the initial wave of contact, became ineligible to take part in ELSA and not contacted again for an interview after the initial wave. They are called ELSA ‘**Sample Members**’. (As a non-participating, ineligible group, Sample Members do not feature in the archived productive ELSA data.)

2.3.1 Household splits and eligibility

If a household consisting of two Core Members splits, i.e. the cohabitation ends, both Core Members remain eligible to take part in their new households as long as they still are still living in Great Britain.

If a Core Member separates from their Partner, the Core Member remains eligible to take part in their new household as long as they still are still living in Great Britain.

If a non-Core-Member ELSA Partner separates from their Core Member partner, whether through divorce, separation or widowhood, they remain eligible to take part until they have been successfully interviewed for **one** subsequent wave of the study; in order to understand their circumstances after this event had occurred. Such 'lone' ELSA Partners are followed up two consecutive waves after separation from the Core Member. If after that time they have still not participated then they are removed from the eligible sample. The only exception to this following rule is if the partner has moved into an institution. In this case a subsequent interview is **not** pursued.

2.4 Proxies

A proxy interview was pursued if an eligible respondent was physically or cognitively impaired; in hospital; or temporarily in care for the whole of the fieldwork period. Those who refused to take part in person but agreed that someone else could do the interview on their behalf could also have a proxy interview.

Interviewers were asked to identify a proxy informant as any responsible adult (aged 16 years or over) who knew enough about the respondent's circumstances to be able to provide information about them. Where possible, a close family member such as a partner or child fulfilled this role.

The CAPI program guided the interviewer through the proxy interview automatically. However, only a subset of questions was asked during a proxy interview. The proxy interview contained the following modules:

HD*	Household Demographics
ID	Individual Demographics
HE	Health
CA	Social Care
SP	Social Participation
WP	Work and Pensions
IA*	Income and Assets
HO*	Housing
FQ	Final Questions

Some specific questions only for proxies were included in the Health, Social Care and Social Participation and Final questions modules.

The asterisked modules were not asked for all respondents. The modules on household demographics and housing were done as part of the proxy interview only when no-one else in



the household was eligible for interview. In cases where no-one else in the same financial unit was eligible for interview, the proxy interview included the Income and Assets module. If two proxy interviews were needed for a couple, the Income and Assets module would only appear in one of the interviews (asking about both their finances). For couples comprising of one person who was interviewed in person and another who needed a proxy interview, the former would automatically be asked the Income and Assets module on behalf of the couple. The question about whether they keep finances together or separate would not be asked.

2.5 Institution interviews

For existing sample members, an ‘institution interview’ was sought with core members who had moved from a private household at the first ELSA interview into a residential care home or similar establishment. Institution interviews could be carried out in person or by proxy. Routing has been developed within the interview specifically for people who have moved into institutions. Institution interviews were not pursued for partners of core members, or for members of the refreshment sample in their first wave of ELSA.

The content of these interviews is broadly the same as the main interview. People in institutions were asked different questions in the Housing and consumption module and were not asked some questions in other modules. There are also certain questions which are *only* asked of/about institutional respondents such as **MIREL** to **MIFIN** (within Individual Demographics).

People in institutions who were interviewed by proxy were also not asked questions that non-institution proxy informants were not asked (see Section 2.4).

The household identifiers for those who have moved out of a household and into an institution have been kept the same as for the household they have left, rather than being assigned a new one. This is only visible in the data where, for example, one respondent remains in the original household but another has moved into an institution. This decision was taken as some modules ask about the household as a whole and it would be confusing, especially in the Household Demographics module, if a respondent was being referred to who was now in a different household. Additionally, these modules may be asked of the respondent in the institution or the respondent still remaining in the previous household, depending on the circumstances.

3 - Survey content

3.1 Development and piloting

For each wave, at least one pilot of the computer-assisted personal interview (CAPI) instrument, the self-completion questionnaire and associated documents were conducted around a year to six months before the beginning of the mainstage fieldwork. All pilot respondents were drawn from a separate sample of households that participated in the ELSA Wave 1 pilots or were selected to join the refreshment pilot sample at subsequent waves. The pilots tested the fieldwork procedure and interview content and in some waves, new questions underwent cognitive testing.

For further information on specific waves' survey development please see the relevant technical report.

3.2 Details of survey content – overview

Overall, the intention of ELSA was to collect data about the same topics in each wave. However, at some waves there were some additions to the content of the interview to respond to new areas of enquiry. Some questions at certain waves were omitted, as it was decided that they did not need to be asked at every wave. Furthermore, several elements of the questionnaire were amended to take account of responses given at the previous waves. At each wave all interviews comprised of a personal face-to-face CAPI interview and a self-completion questionnaire.

The ELSA program allowed flexibility in administering the interview. Respondents could be interviewed individually, or in houses with more than one eligible respondent, interviewed at the same time (in a single session) using concurrent interviewing techniques. In a concurrent session the same block of questions was asked alternately of each person. Concurrent interviews tended to be quicker than two separate individual interview sessions, and were generally more convenient for respondents.

ELSA uses a technique called 'feeding forward' data. It is a technique that feeds responses individuals made at earlier interviews to either aid recall and/or improve consistency of responses across interviews. ELSA uses proactive dependent interviewing: i.e. earlier responses are provided to the individual before they respond to a question (rather than reactive dependent interviewing when the earlier response is used after the individual has responded as a form of validation). Dependent interviewing was used to check some information collected during previous waves (or HSE), to determine changes in status, and to control routing within the questionnaire.

There are various modules each covering a different area of interest. The content and major routing of each module at waves 1–9 is summarised in section 3.3 below. Although interviews tended to follow the same module order, some flexibility was given to the interviewer. For example, the timed walk could be administered at any time after the Health module, and it was possible for interviewers to skip the Income and Assets or Housing modules if it was more convenient to do them at another time.

Five of the modules (Cognitive Function, Expectations, Effort and Reward, Psychosocial Health and Final Questions) form the 'private modules' block. Wherever possible, these modules were



administered with no other household members present. If two respondents were being interviewed concurrently, whilst the first respondent was being asked the private block, the second responding individual was asked to fill in the self-completion in a separate room. The two respondents then switched places.

3.2.1 End of Life interviews

On Waves 2, 3, 4 and 6, separate 'End of Life' interviews were also conducted among a sample of family members or carers of ELSA respondents who had recently passed away, asking about the circumstances around the respondent's final stages of life. Separate data and documentation about the EOL interviews is available at the UKDS.

3.2.2 End of Life interviews

After Wave 3, a subset of the ELSA sample was invited to complete an additional 'Life History' interview, collecting information on a range of areas spanning their whole life. Separate data and documentation about the EOL interviews is available at the UKDS.

3.2.3 Follow-up nurse visit

In waves 2, 4, 6, 8 and 9, eligible respondents were offered a follow-up visit by a qualified nurse. The ELSA nurse visit have collected various physical examination and performance data and collected biological samples for analysis.

In waves 2, 4 and 6, all ELSA Core Members who completed a personal CAPI interview (not proxies) were eligible for a nurse visit. Some ineligible Partners were also offered a nurse visit on the nurse's discretion.

In Wave 8, only a sub-sample of Core Members who completed a person CAPI interview were eligible for a nurse visit: people who had consistently taken part in a nurse visit when available prioritised for inclusion in the eligible sample, to maximise the collection of longitudinal nurse data. A small number of ineligible Core Members or Partners were also offered a nurse visit on the nurse's discretion.

The nurse visit was offered in Wave 9 to those participants who:

- completed the main interview;
- were not offered a nurse visit during Wave 8;
- took part in nurse visits in earlier waves whenever they were eligible.

Separate data and documentation about the nurse visits interviews is available at the UKDS – please see the ELSA Nurse Data User Guide for more details.

3.3 Questionnaire content – Waves 1 to 9

The table below gives an overview of the modules and topics covered in waves 1 to 9. Please see the Appendix of *ELSA Wave 9 Questionnaire & Data Documentation*, available on the UKDS, for a more detailed tabular summary of the topics covered at each wave.

Table 3.1: Questionnaire modules

Module	Description – changes from previous waves
Household Demographics (HD)	Answered by one person on behalf of the household with answers copied onto the data for other household members. Collects basic demographic information on all in the household. Uses feed forward data from previous wave.
Individual Demographics (ID)	Answered by all. Includes details on: legal marital status; living children (including adopted, fostered or stepchildren), grandchildren & great-grandchildren; number of siblings; childhood circumstances. Also includes information on parents.
Health (HE)	Answered by all. Includes: self-reported general health; long-standing illness or disability; eyesight and hearing; specific diagnoses and symptoms; pain; difficulties with activities and instruments of daily living; health behaviours. Aged 60+ were asked about falls and fractures. From Wave 6 this module includes a sub module on receipt of and payment for social care.
Social Participation (SP)	Answered by all. Questions on use of public transport; use of services e.g. transport provided by hospitals/day centres and meals on wheels.
Work & Pensions (WP)	Answered by all. Includes: current work activities; current or past pensions. If retired and receiving a pension details collected on the pension, including amounts.
Income & Assets (IA)	If a couple's finances are kept together one of them answers. If separate, both answer. Questions include: wages; state & private pensions; annuity income; state benefits; financial & non-financial assets held, income from these assets; regular & one-off transfers from non-household members; life insurance.
Housing (HO)	Answered by one person on behalf of the household. Includes: current housing situation; housing-related expenses; ownership of durables, cars and pets; expenditure on food, clothes, gifts and leisure activities.
Cognitive Function (CF)	Answered by all, start of 'private modules' block. Measured: cognitive function including memory & speed.
Expectations (EX)	Answered privately by all. Expectations about certainty of future events and financial decision-making including moving home in future
Effort and Reward (ER)	**MODULE INTRODUCED AT WAVE 2 ** Answered privately by all. Assesses respondents' motivations for voluntary work, caring for others, relationship between effort and reward and provision of care and use of respite care services.
Psychosocial Health (PS)	Asks how respondent views their life across a variety of dimensions.
Final Questions and consents (FQ)	Answered privately by all. Included: demographic information & stable address contact. Consent for nurse visit follow-up on Waves 2, 4, 6, 8 and 9. Verbal reminder of data linkage consent if already given; otherwise asked for consent to obtain health and economic data from administrative sources.
Measurement Module (MM)	Walking speed completed by those aged 60+ when safe to do so. Respondents walked distance of 8 feet (2.4m) twice, timed using a stopwatch. At Wave 8: Weight measurement taken for those aged 51+. Where measurement was not possible or refused the respondent was asked to estimate their weight.
Self-completion	The main (core) self completion includes: respondents' quality of life, social

Module	Description – changes from previous waves
<p>(“SC”) questionnaire</p>	<p>participation, control at work, life satisfaction, social networks and alcohol, fruit and vegetable consumption and wellbeing.</p> <p>Some waves included additional self-completion questionnaires, e.g. work and health questionnaires with vignettes in Wave 3 and sexual activity in Waves 6 and 8.</p> <p>For more detail on questions included at each wave, please see the cross-wave topic summary in Appendix A of <i>ELSA Wave 9 Questionnaire & Data Documentation</i>. For more detail on the administration of self-completions please see section 4.5 below.</p>

4 - Fieldwork and response

4.1 Fieldwork – overview

The fieldwork process for each wave of ELSA has been very similar. Eligible individuals satisfying a number of criteria (see section 2.3 above) were sent an advance letter inviting them to take part, along with a survey leaflet. Interviewers then either telephoned (for existing sample members aged under 80 who had taken part in the preceding wave) or visited the households to explain the study, and interview willing individuals straight away or to make appointments to call back at a convenient time. The main survey comprised of a personal face-to-face interview and a paper self-completion questionnaire. Waves 2, 4, 6, 8 and 9 also included a separate follow-up Nurse visit; the data from both these elements of ELSA is available from the UK Data Service.

At waves 1–4, before starting work all new interviewers underwent a two-day personal briefing by a researcher. From Wave 5 the two-day briefing for new interviewers was reduced slightly by half a day. At waves 1-5, those interviewers who had taken part in a previous wave of ELSA underwent a one day refresher briefing. At Wave 6, this was reduced to a half day briefing. Experienced ELSA interviewers had a self-briefing at home which included a training DVD and a follow up teleconference with the research team. In waves 7, 8 and 9, face to face briefings were undertaken, with a full day briefing for new interviewers and half-day refresher briefing for experienced interviewers. The briefings covered all fieldwork procedures including training on how to administer the assessments (walking speed, weight measurement and cognitive function), the documents needed for the study and provided an introduction to all questions within the CAPI interview. Interviewers were provided with written study guidelines to reinforce the briefing.

Addresses within nearby postcode sectors were clustered and issued to interviewers. At waves 1-5, all interviewers were instructed to report to the police station local to where they were working before starting to carry out their visits. This was to advise the police station that they would be working in the area. Interviewers were asked to show a copy of the ELSA advance letter, leave their name and NatCen's contact details. From Wave 6 this process was no longer required.

For further information on the specific fieldwork procedures at each wave please see the technical reports available at: <http://www.elsa-project.ac.uk/publications/case/technical>.

4.2 Fieldwork – tracing movers

To minimise the sample attrition caused by respondents moving address, procedures were in place to track respondents who move between waves, or since their original HSE interview, to ensure that the more mobile sections of the ELSA sample are not lost.

If the whole household had moved since the last interview, or a core member who had consented to be re-contacted in future waves had moved away, interviewers were asked to try the following to trace movers:

- Attempt telephone contact with the respondent
- Attempt to find a follow-up address

- Approach the present occupants, neighbours or friends to obtain the new address
- From Wave 3, approach the person(s) living at the “stable address” provided previously by the respondent – Since Wave 2 respondents had been asked to give the name and contact details of someone who could be contacted if they moved
- Consider phone books, electoral register, local shops, letting agency, estate agent, post office

A ‘tracing letter’ was offered if interviewers identified a member of the public who was aware of the core member’s new address but was reluctant to reveal it to the interviewer. This letter, which was forwarded with a pre-paid envelope by the member of the public who had been identified, asked the core member to contact the office with their new address.

At Wave 3 the DWP assisted with the tracing of core members using their state pension databases. Each respondent’s name, date of birth and address were provided to DWP and they matched this to their databases in order to identify the most up-to-date contact details. If a new address was found, an advance letter was sent to the respondent.

At Waves 3-6 tracing was conducted through the National Health Service Central Register (NHSCR). For those respondents who we could not trace and who agreed to have their data linked to the NHSCR we were able to obtain health authority information, which enabled us to write to the health authority and ask them to pass on an advance letter to the respondent via their GP.

4.3 Fieldwork periods

The ELSA fieldwork takes place over the course of approximately one year, with the sample issued to field in a number of batches to balance the field interviewer workload. The table below summarises the fieldwork periods of Waves 1-9.

On the longitudinal Waves 2-9, wherever possible, interviewers were assigned to the same households they had interviewed at the previous wave.

Table 4.1: Fieldwork periods by wave

	Fieldwork period
Wave 1	March 2002 – March 2003
Wave 2	June 2004 – July 2005
Wave 3	May 2006 – August 2007
Wave 4	May 2008 – July 2009
Wave 5	June 2010 – July 2011
Wave 6	May 2012 – June 2013
Wave 7	June 2014 – May 2015
Wave 8	May 2016 – June 2017
Wave 9	June 2018 – July 2019

4.4 Number of interviews

Detailed analysis of response rates (field contact and response rates, cohort response rates) at each ELSA wave can be accessed in the [technical reports](#) (Waves 7 and 8 *forthcoming) and [methods chapters](#) within each wave's main report (Wave 9 forthcoming October 2020).

The table below summarises the number of productive interviews archived at each wave, and the number of these which were Core Member and Partner interviews.

Table 4.2: Number of interviews by waves

	Total archived interviews	Core Member interviews	Partner interviews
Wave 1	12,099	11,391	708
Wave 2	9,432	8,780	652
Wave 3	9,771	8,810	961
Wave 4	11,050	9,886	1,164
Wave 5	10,274	9,090	1,184
Wave 6	10,601	9,169	1,432
Wave 7	9,666	8,249	1,417
Wave 8	8,445	7,223	1,222
Wave 9	8,736	7,289	1,406

4.5 Self-completions in ELSA

4.5.1 Main self-completion (all Waves)

At each wave of ELSA interviewers have asked respondents to complete a core self-completion questionnaire covering questions such as wellbeing, relationships, alcohol consumption. Everyone with an in-person interview (not proxy) is invited to complete this questionnaire. This may be completed while the other person in a couple completes sections in the 'private' block of the questionnaire, or may be left for completion afterwards. For existing sample members the interviewer may leave it for respondents to complete in advance. A summary of whether or not a respondent completed a core self-completion for each Wave of the study can be found by looking at the variable OutScw1, OutScW2 etc in the Index file.

4.5.2 Health and Work self-completion at Wave 3

At Wave 3 a sub-sample of respondents were invited to complete a second self-completion in addition to the main self-completion. Thus there were three groups of people (from among those eligible for a self-completion). The numbers in each group are about equal (a third of the sample in each):

- Main self-completion only
- Main self-completion and 'work' self-completion
- Main self-completion and 'health' self-completion

Both the work and health questionnaires contained:

- a section of 7 questions about the respondent's health over the previous 30 days (the same in both questionnaires)
- a section of anchoring vignettes which were different in the two questionnaires (9 vignettes in the work questionnaire and 18 vignettes in the health questionnaire). See the Wave 3 report methodology chapter for more information on anchoring vignettes. The vignettes in the work questionnaire related to work related health and those in the health questionnaire were about health unrelated to work.

Respondents were assigned in advance in the sample file to each of the three groups (core only, core and work, core and health). There was only one version of each questionnaire 'work' and 'health'. This means that the names and details used in each vignette were not randomized – each respondent got the same name in the vignette (this is because it was administered by paper self-completion).

The ELSA Index file contains variables which show which questionnaire each respondent was invited to complete and which they actually completed:

- Outscww3 – work self-completion
- Outschw3 – health self-completion

4.5.3 Sexual activity and relationships self-completion at Waves 6 and 8

At Waves 6 and 8 all respondents who had an interview in person were invited to complete a second self-completion in addition to the core self-completion. This asked people about their attitudes towards sexual activity and their recent and life time experiences. At each wave, there were two versions of the questionnaire (one for men and one for women) and each person was asked to complete one according to which sex they were. These questionnaires were completed during the interview or afterwards and respondents were given an envelope to seal the questionnaire in before handing it back to the interviewer.

4.5.4 Two versions of the Wave 7 self-completion

Part-way through Wave 7 fieldwork, it was noticed that the answer scale of some questions on the self-completion questionnaire was set out as six point 1-6 scale, as opposed to the intended seven point 0-6 scale. A new version of the self-completion was issued part-way through fieldwork with the corrected scale. Therefore, two versions of each of the affected variables are provided in the dataset, with the variable name suffix indicating the self-completion version completed (_v1: Orange questionnaire, scale 1-6; _v2: Green questionnaire, scale 0-6). Additionally, the variable SCVER in the Wave 7 dataset provides information on which version of the self-completion questionnaire was completed by each respondent.

5 - Weighting, clustering and stratification

5.1 Clustering and stratification

With complex sample designs such as ELSA, the effects of clustering and stratification should be taken into account when conducting analyses.

Analysts should use the following cluster and stratification variables, according to wave.

Table 5.1: Cluster and stratification variable by wave

	Cluster variable	Stratification variable
Wave 1	ahsecls2	astratif
Wave 2	hseclst	astratif
Wave 3 onwards ⁴	idahhwn (household serial number: 'idahhw' plus Wave number e.g. for Wave 5, this is idahhw5) (*)	gor

Where possible we recommend that analysis be conducted on weighted data since this will help to minimise bias from differential non-response amongst key sub groups. This is particularly important if conducting Wave 8 full sample analyses due to the relatively smaller size (in comparison to the proportions in the population) of the youngest age groups 52-54 in the ELSA sample. More details will be provided in the Wave 8 Technical Report (forthcoming).

5.2 Weighting for Waves 1 and 2

The variable in the dataset to be used for weighting at Wave 1 is **W1wgt** and the variable for Wave 2 is **W2wgt**. Weights were calculated for core sample members only (including proxy and partial interviews). All other non-sample individuals who were interviewed (Core Partners, New, Younger and Older Partners) have a weight of zero.

When running weighted analyses, researchers should remember to exclude these respondents (with zero weights) from the un-weighted base, if quoted. The data for Partners can be used as supplementary information for Core Members⁵.

For further information on Wave 1 and 2 weighting, please see the technical reports available at: <http://www.elsa-project.ac.uk/publications/case/technical>.

⁴ From Wave 3, attrition means that geographical clustering is negligible; however, clustering within household should be taken into account.

⁵ If non-core sample members are to be analysed they should be analysed unweighted.

5.3 Weighting for Waves 3–9

The weighting for Waves 3–9 can be divided for analysis cross-sectionally or longitudinally:

- Cross-sectional analysis uses data collected only at a particular wave.
- Longitudinal analysis involves data collected from more than one wave for the purposes of analysing change.

For Waves 3–9 cross-sectional and longitudinal weights support these two different objectives.

As with Waves 1 and 2, only ELSA Core Members living in private households in England (at the time of interview) have been given a weight. Cross-sectional interviewer weights are “system missing” for ELSA Partners, and zero for all Core Members not eligible to receive a weight (institutional respondents or those living outside of England). To receive a longitudinal weight, respondents must have taken part in all nine waves of ELSA; all other respondents are ‘system missing’.

When running weighted analyses, researchers should remember to exclude these respondents (with zero or ‘system missing’ weights) from the un-weighted base, if quoted. The data for Partners can be used as supplementary information for Core Members.

5.4 Longitudinal weights Waves 3–9

These weights are defined for the subset of cases who have taken part in all waves, up to and including the wave in question. From Wave 8 an additional longitudinal weight has been added, defined for the subset of cases who have taken part in all waves since Wave 4, up to and including the wave in question.

At each wave, the fully responding core members are re-weighted to take account of respondents at the previous wave that were lost through refusal at the current wave or through some other form of sample attrition. Core members from the original sample who returned to the study having missed a wave therefore have no longitudinal weight.

Table 5.2: Longitudinal weight variables by wave

	LONGITUDINAL WEIGHTS
Wave 3	<p>Variable name: W3LWGT</p> <p>Calculated for the set of 7,168 core members who have responded to all three waves of ELSA, and remain living in private households.</p>
Wave 4	<p>Variable name: W4LWGT</p> <p>Calculated for the set of 5,971 core members who have responded to all four waves of ELSA, and remain living in private households.</p>
Wave 5	<p>Variable name: W5LWGT</p> <p>Calculated for the set of 5,262 core members who have responded to all five waves of ELSA, and remain living in private households.</p>

Wave 6	<p>Variable name: W6LWGT</p> <p>Calculated for the set of 4,711 core members who have responded to all six waves of ELSA, and remain living in private households.</p>
Wave 7	<p>Variable name: W7LWGT</p> <p>Calculated for the set of 4,062 core members who have responded to all seven waves of ELSA, and remain living in private households.</p>
Wave 8	<p>Variable name: W8W1LWGT</p> <p>Calculated for the set of 3,470 core members who have responded to all eight waves of ELSA, and remain living in private households.</p> <p>Variable name: W8W4LWGT</p> <p>Calculated for the set of 5,623 core members who have responded to all five waves since Wave 4, and remain living in private households.</p>
Wave 9	<p>Variable name: W9W1LWGT</p> <p>Calculated for the set of 2,959 core members who have responded to all nine waves of ELSA, and remain living in private households.</p> <p>Variable name: W9W4LWGT</p> <p>Calculated for the set of 4,848 core members who have responded to all six waves since Wave 4, and remain living in private households.</p>

5.5 Cross-sectional weights Waves 3-9

The cross-sectional weights for Wave 3-9 are computed for the following groups:

- Longitudinal sample members (i.e. those given a longitudinal weight)
- Cohort 3 core members (new entrants at Wave 3)
- Cohort 4 core members (new entrants at Wave 4)
- Cohort 6 core members (new entrants at Wave 6)
- Cohort 7 core members (new entrants at Wave 7)
- Cohort 9 core members (new entrants at Wave 9)
- Wave non-responders (those core members from a previous cohort who return to the study having missed a previous wave)

The cross-sectional weights were calculated separately for each cohort. In each case the weighting aimed to adjust for differences in the propensity to respond amongst key sub-groups. The final step in the calculation of the cross-sectional weight involved computing a scaling factor to ensure that the original sample (Cohort 1) and refreshment samples (Cohorts 3, 4, 6, 7 and 9) were represented in the same proportions, with respect to age, in which they

appear in the population (based on the mid-year household population estimates for the year in which fieldwork started provided by the Office for National Statistics). For example, for Wave 8 weights this would be the 2016 midyear household population estimates.

Table 5.3: Cross-sectional weight variables by wave

	MAIN INTERVIEW CROSS-SECTIONAL WEIGHT
Wave 3	Variable name: W3XWGT
Wave 4	Variable name: W4XWGT
Wave 5	Variable name: W5XWGT
Wave 6	Variable name: W6XWGT
Wave 7	Variable name: W7XWGT
Wave 8	Variable name: W8XWGT
Wave 9	Variable name: W9XWGT

5.6 Self-completion weights

Self-completion weights were created at Waves 4–8 and 9 for the Core Members whose self-completion questionnaire was received with the majority of questions answered. For all others, the weights are set to ‘system missing’. These weights (**W4SCWT**, **W5SCWT**, **W6SCWT**, **W7SCWT**, **W8SCWT** and **W9SCWT** respectively) built upon the Wave 4--9 cross-sectional weights and aimed to adjust for differential non-response to the self completion questionnaire amongst respondents.

At Waves 6 and 8, a separate weight was also created for the sexual health self completion questionnaire, for the Core Members whose sexual self-completion questionnaire was received with the majority of questions answered. For all others, the weights are set to ‘system missing’. The variable names for this weight are **W6SSCWT** and **W8SSCWT**.

5.7 Nurse data weights

Weights are also available for analysing the nurse data collected at Waves 2, 4 6. Please see the User Guide for ELSA Nurse Data on the UK Data Service for more information about analysing data collected in the nurse visit.

6 - Dataset information

This User Guide refers to the main interview datasets that have been archived for ELSA Waves 1-9.

6.1 Datafiles available at the UKDS

6.1.1 Main interview datasets

For Waves 1–8, the UK Data Service contains a single ELSA data file containing the majority of variables relating to the ELSA interviewer visit.

For Wave 9, the data related to the main interview is archived under a number of separate files, reflecting the relative sensitivity or disclosure risk of the variables:

- A dataset available for holders of the UKDS End User Licence (EUL) contains the majority of the interview data, with sensitive or disclosive variables removed
- A Primary Special Licence (SL) dataset, available for users under the UKDS Special Licence conditions, contains all the variables in the EUL dataset, as well as additional more sensitive/disclosive variables
- Further separate SL datasets containing:
 - State pension age variables (year and age in years reached state pension age)
 - Geographic variables (LA, IMD quintiles, population density quintiles, urban/rural)
- A Primary Special Access (SA) dataset, available using the UKDS Secure Access gateway, contains all the variables in the Primary SL dataset and additional highly disclosive variables.
- Further separate SA datasets containing:
 - State pension age variables (exact date and age reached state pension age)
 - Geographic variables (lower level Census geographic variables)

A detailed outline of variables available within each respective Wave 9 dataset is available in the Wave 9 Questionnaire & Data Documentation available via the [UKDS website](#).

Data for ELSA respondents collected during HSE (referred to as “Wave 0” data) and ELSA Waves 1-9 can be used for longitudinal analysis. All this data is available from the [UK Data Service](#).

6.1.2 Other ELSA datasets at the UKDS

In addition to the main ELSA datasets for Waves 0–9, there are a number of additional ELSA datafiles available (under EUL) via the UKDS:

- Nurse visit datasets (Waves 2, 4, 6, 8 and 9* forthcoming) – see ELSA Nurse Visit User Guide for more details.
- Life History interview dataset (Wave 3)
- ‘End of Life’ interview datasets for Waves 2, 3, 4 and 6
- Pension grid: data and documentation of data originating in the main interview pensions loop is available in a separate ‘pensions grid’ file for all waves (by IFS)
- IFS (non-financial) Derived Variables and documentation produced for all waves by IFS

- Financial Derived Variables and documentation produced for all waves (by IFS)
- Pension wealth derived variables for Waves 1–5 (by IFS)
- Mortgage dataset (Wave 2) and Mortgage grid dataset (Wave 3)
- Ryff self-completion dataset (Wave 2)
- ELSA Index File (Waves 0–5) including key status variables for each wave such as fieldwork issue status and outcome codes, mortality status. It has not been possible to update the Index File since Wave 5 as no further updates to mortality data have been received from NHS Digital; negotiations are ongoing.
- Harmonised ELSA dataset and documentation (by RAND US)

A summary of data files produced by NatCen and IFS available across the waves is also available at: <http://www.elsa-project.ac.uk/availableData>

6.1.3 Geographic variables

Government Office Region (GOR) is included in the main interview dataset for all waves.

Various other more disclosive geographical variables are available under secure arrangements. For Waves 8 and 9* (*forthcoming) geographic variables are available in the Special Licence and Secure Access datasets as detailed above. Please see the Wave 9 Questionnaire & Data Documentation for details. For Waves 1-7, please contact elsadata@natcen.ac.uk if you would like to request access to any geographic variables.

6.1.4 Access to non-archived data

Any further variables not included in the archived datasets may be requested via the NatCen Data Release Panel. Please contact the ELSA Data Manager at elsadata@natcen.ac.uk for more details.

6.2 Order and content of the main interview dataset

The main interview datasets (EUL, SL and SA versions) are individual level files, including all respondents with a productive personal or proxy interview at a given wave. This includes both ELSA Core Member and Partners. The main group of respondents for analysis is the Core Members. Data on Partners can be used as characteristics of the Core Members (i.e. to provide supplementary information), but these partners should not be analysed as individuals in their own right. The ineligible partners are unrepresentative, and any analysis using them would need to be unweighted.

The datasets contain variables in the following order:

- Key variables not in the questionnaire (e.g. serial number, interview month and year, outcome codes)
- Variables in the questionnaire (in the order they appear in the CAPI interview) unless marked for not to be archived (see next section). A small number of additional computed variables that are associated with particular questionnaire variables are located alongside these variables in the data. These are annotated in the questionnaire.

- Variables in relevant self-completion questionnaires
- Other variables not in the questionnaire including derived variables, variables fed forward from previous waves, and survey weights. Derived variables are denoted with “(D)” at the beginning of the variable label⁶.

6.2.1 Dropped variables

All variables in the questionnaire documentation with an @ symbol next to their name have been removed from the archived datasets, in order to reduce the potential to identify individuals and for other reasons (specified below):

1. Uncoded open text responses
2. Variables containing a personal identifier (e.g. name/address)
3. Other variables considered to be disclosive, such as:
 - Detailed ethnicity
 - Specific country of birth
 - Council tax payments (Different councils charge different amounts and therefore the amount may reveal the area the respondent lives in)
 - Water and sewerage charges (These vary in different areas and therefore the amount may reveal the area the respondent lives in)
4. Timing variables, which give the time at specific points in the interview (used for administration purposes)
5. Administrative variables (e.g. link variables at start and end of sections) and variables that only contain missing values – excluded because they are not useful. Such variables have only been kept if they are integral to the structure of the data.

All efforts have been made to release as much of the ELSA data as possible to researchers while safeguarding the confidentiality of ELSA participants.

6.3 Household level modules

There are three ELSA modules that can be or are always answered by one person on behalf of the whole households. The dataset includes a variable for each of these modules that identifies the person who answered the module:

- Household Demographics = **HHRESP**
- Income and Assets = **IAPID**
- Housing = **HOPID**

Where modules were answered by one person on behalf of others, information was copied directly to the other member(s) of the household or financial unit.

⁶ Derived variables were not computed for Wave 7, but will be added following a retrospective review and re-archiving of the main datasets for Waves 1-7.

6.4 Serial numbering

Constant Individual Serial Number

All the ELSA data files deposited in the archive contain a unique individual analytical serial number (**IDAUNIQ**) to enable users to link the different files. Each respondent has a unique value for **IDAUNIQ**, which will remain constant across all datasets at all waves.

Wave-specific household serial number

The five digit household serial number (for example, for Wave 9 **IDAHHW9**) was randomly generated for each Wave, and does not relate to the serial number used during interviewing, or to the household serial numbers of previous waves. The datasets for each wave of ELSA (including "Wave 0", i.e. HSE) contains a different set of household serial numbers (**IDAHHW0**, **IDAHHW1** etc.). It is necessary to have a different household serial number for each wave as respondents can change households between waves. The ELSA Index File, available from the Data Archive, enables data users to link the household serial numbers in order to compare data for each respondent and household at different waves.

Person number

Each person within the household was given a number, starting from 01, at the time of the HSE interview (**PERID**). The numbering was continued for new people that entered the household after the HSE interview.

This number is used for several variables. For example, it is used to identify the respondent's partner within the household (see variable **CPID**), or to identify the person in the household that responded to particular questions such as the Income and Assets module (see variable **IAPID**).

6.5 Missing values

The following missing value codes are used in the ELSA interviewer datasets:

Table 6.1: Missing value codes in interviewer datasets

Code	Description	Relevant modules/sections	Waves
-9	Refusal	All	All
-8	Don't know	All	All
-4	Missing by error	All	Wave 8 only
-3	No valid answer	Self-completion & Sexual self-completion variables	Wave 8 only
-2	Self-completion instrument not completed	Self-completion & Sexual self-completion variables	All
-1	Item not applicable	All	All

As summarised in the table, all codes except -2 and -3 are relevant for the CAPI questionnaire variables.

Most questions in the CAPI questionnaire permit 'refusal' and 'don't know' answers. For the small number of questions where a response of 'don't know' or 'refusal' was not permitted in the CAPI instrument, this is indicated in the questionnaire.

For various reasons, some respondents did not complete the entire interview. For these 'partial' interviews, the questions towards the end of the questionnaire that were not asked are coded as '-1' (not applicable).

6.6 Multi-coded and dichotomous variables – Wave 3 onwards

A change in the way that ELSA data is archived occurred from Wave 3 onwards regarding multi-coded variables. A variable is defined as multi-coded if more than one response could be given e.g. favourite activities are reading *and* going to the cinema. For Wave 3, and beyond, such variables are archived as dichotomous variables, with each answer option being coded as 'mentioned' or 'not mentioned'. Please note that although the way the data is presented is different, no change was made to the way the questions were asked during the interview.

For example, data from the series of questions **EXRELE01 – EXRELE10** ("Whether having too little money stops them from doing things") was outputted as multi-coded variables at previous waves.

Therefore, there was a variable for the respondent's first response to the question, second response and so on. Each variable had the following answer options:

- 1 Buy your first choices of food items
- 2 Have family and friends round for a drink or meal
- 3 Have an outfit to wear for social or family occasions
- 4 Keep your home in a reasonable state of decoration
- 5 Much better off
- 6 Pay for fares or other transport costs to get to and from places you want to go
- 7 Buy presents for friends or family once a year
- 8 Take the sorts of holidays you want
- 9 Treat yourself from time to time
- 96 None of these.

As of Wave 3, the data was outputted as dichotomous variables so there is a variable for each response category for that question which shows whether or not the respondent gave that response.

The variables in the data are:

EXRELEFO: Things interviewee is not able to do because of having too little money: Buy your first choices of food items.

EXRELEME: Things interviewee is not able to do because of having too little money: Have family and friends round for a drink or meal.

EXRELEOU: Things interviewee is not able to do because of having too little money: Have an outfit to wear for social or family occasions.

...

EXREL96: Things interviewee is not able to do because of having too little money: None of these.

These questions now have much simpler answer options of:

0 Not mentioned (*i.e. respondent did not give this answer*)

1 Mentioned (*i.e. respondent did give this answer*).

The new dichotomous variables have been given different names in order to distinguish them from their multi-coded equivalents at previous waves. The variable names now have a common 'stem' for each question (**EXRELE** in the example above), and then two or three letters to indicate the answer option. Where answer options are the same between questions, the same two or three letters have been used each time for consistency.

6.7 Outcome code variables

Individual and Household interview outcome codes

Individual and household level interview outcomes are given for each wave. The individual outcome is a two-digit code. The first digit indicates whether the interview was full (1) or partial (2). The second digit indicates whether the interview was in person (1) or by proxy (3). The household outcome is a three-digit code. The first digit indicates whether all eligible residents were interviewed (1) or some interviewed (2).

Table 6.2: Individual and Household interview outcome codes by wave

	Individual outcome	Household outcome
Wave 1	Indoc	Rthhout
Wave 2	W2INDOUT	W2HHOUT
Wave 3	W3INDOUT	W3HHOUT
Wave 4	OUTINDW4	OUTHHW4
Wave 5	W5INDOUT	W5HHOUT
Wave 6	W6INDOUT	W6HHOUT
Wave 7	W7INDOUT	W7HHOUT
Wave 8	W8INDOUT	W8HHOUT
Wave 9	W9INDOUT	W9HHOUT

Self completion outcome codes

An outcome code variable for the core self-completion was included at Wave 3 (**OUTSCW3**), and this was reintroduced as **W8SCOUT** in Wave 8. The self-completion outcome codes of Waves 1–5 are available in the ELSA Index File.

For the Wave 3 additional self-completion outcome codes for the health and work vignette self-completions (**OUTSHW3** and **OUTSWW3** respectively), the value of '-1' it means they were not part of the sub-sample selected to receive that questionnaire.

An outcome code for the sexual self-completion was included at Wave 8 as **W8SSCOUT**

Nurse outcomes

The Wave 8 main interviewer ELSA dataset also includes the outcome of the Wave 8 follow-up nurse visit (**W8NUROUT**) and the blood sample outcome (**W8BSOUT**).

6.8 Questionnaire errors

6.8.1 Wave 8

Routing errors

The following questions have cases with missing data by error identified in post-fieldwork data checking. The affected cases have the non-response value of -4 'Missing by error' for the relevant variables.

Table 6.3: Routing errors

Variable(s)	Number of cases affected	Additional notes
respwts	6	The weight measurement is in a separate parallel block. These interviews were marked as ready without this parallel block being completed.
FqMQua	9	
exhvr, exhva	165, 169 respectively	These respondents were not routed to these questions despite the household fulfilling the routing condition HoTenu = 1, 2 or 3.
fqnurse	5	Eligible for a nurse visit but end of questionnaire question on nurse follow-up visit not completed.
scnta to scntf, scmena to scmend	440	The Wave 8 paper self-completion questionnaire had an error in the routing specification: respondents answering 'No' to question 23 were instructed to skip to question 29, instead of the intended question 27 (please see the Self-Completion Questionnaire, included in the ELSA Wave 8 Questionnaire & Data Documentation). 440 respondents followed the routing instruction and missed out questions 27 and 28 (dataset variables scnta to scntf and scmena to scmend)

Variable(s)	Number of cases affected	Additional notes
		respectively). 536 respondents did not follow the instruction and still completed questions 27 and 28. No edits have been made to the data in these cases – all responses have been retained.

Feed-forward loading errors

For 168 respondents, the individual level feed-forward data from previous waves did not load for their interview. These individuals are marked as 'Yes' in the variable 'fferror'. For this reason, they were routed to a number of questions as if they were new respondents, rather than asked a dependent interview variant of a question, checking existing data. For example, these respondents would have been routed to HEOPT to ask about eye conditions, instead of HEOPC asking to confirm previously diagnosed eye conditions even if they had previously reported one already). If analysing information for the full sample, therefore please look at both the confirmation and newly-asked variants of questions to account for these cases.

6.9 Coding and editing

Additional coding and editing tasks were performed after the interviews were conducted.

Coding

Any newly-reported jobs were coded to the Standard Industry and Occupation code frames, SIC and SOC.

The coding of other responses was mostly dealt with by the CAPI questionnaire through the use of fully closed questions. There were also a number of open questions where an answer was recorded verbatim. These open text answers are not archived for disclosure reasons, but the responses to these 'open' questions have been coded in the office.

The majority of the open variables are follow-up questions to questions with code frames that included an 'other' option (e.g. **DIKLIV**). In these 'other-specify' questions, interviewers could use this option if the respondent's answer did not fit any of the codes or if they were not confident of coding into the prescribed codes. In these cases, the interviewer recorded the full 'other' answer at a follow up question (e.g. 'other' responses relating to **DIKLIV** were recorded in **DIKLO** – not archived).

If these other-specify questions were 'single coded', i.e. when only one option could be chosen (e.g. **DIKLIV**) then the text answers were coded and incorporated into the original coded variable (i.e. **DIKLIV**).

For all previously multi-coded variables (and from wave three dichotomous variables) that were coded, there are two sets of variables. The first are the original variables that contain the answers recorded by the interviewer (e.g. **SPTRAB1** to **SPTRAB95**). The second set of variables contains the original coding plus the codes assigned to "other answers" (e.g. named **SPTRM01** to **SPTRM95**). Note that the letter 'm' added to these variable names after the 'stem'



(**SPTR**) means that they contain ‘merged’ original and coded answers. This naming convention is followed consistently so the final merged variables can be identified by name. It is recommended that the merged variables should be used instead of the original variables as they present a more comprehensive picture of all respondents’ answers.

Editing

As with the coding, most of the editing was carried out by the interviewers in the field. However, there were some additional checks that related to inconsistencies in the data that were carried out after the interview.

Appendices

➤ Notes about particular variables in the Health module

Variable name (in CAPI questionnaire)	Note
All HEOPT, HEDIAA and HEDIAB questions	<p>The idea of the diagnosed disease questions is to track diagnoses across time, but not to attempt to derive an exact history. The first time the respondent is interviewed for ELSA they are asked if they have ever been diagnosed with any of the diseases or conditions specified. For several of these conditions they are asked their age of diagnosis. In subsequent interviews they are asked to confirm that they had been diagnosed and, for most conditions, to say whether or not it is continuing. They are then asked if any additional conditions have been diagnosed since the previous interview and, if so (for several conditions), when.</p> <p>First-time ELSA interviewees are not asked whether they still have the condition but in some circumstances (e.g. heart attack, stroke, and psychiatric conditions) there are subsidiary questions that ask if they have experienced the disease/condition during the previous two years.</p> <p>A report of a diagnosis at these questions does not necessarily signify incidence since last interview for two reasons:</p> <ul style="list-style-type: none"> • Respondents interviewed by proxy do not have information fed forward so may report an old diagnosis here; • Some respondents report diseases for the first time although they were diagnosed a long time ago – analysts are advised to look at dates of diagnosis where available. <p>Users wishing to ascertain incidence or prevalence are advised to link data together across the waves and to note whether the interview was by proxy or not. Apart from the note above regarding incidence please also note;</p> <ul style="list-style-type: none"> • Proxies are asked if the respondent has ever had any of the listed conditions, even if they have previously reported diseases. They are not asked the questions about confirming a previous report - this is to preserve confidentiality. It has been found that proxies do not report all conditions that a respondent has previously reported in person; • Occasionally a respondent re-reports an old disease or condition that had been 'dropped' from the feed

Variable name (in CAPI questionnaire)	Note
	<p>forward, because the respondent had told the interviewer at an earlier wave that they no longer had the condition.</p> <p>For first time ELSA interviewees and for proxies, Code 96 indicates that the respondent (or proxy) believes they have never been diagnosed with any of the conditions in the set.</p> <p>For subsequent interviews, Code 96 indicates that no additional diagnoses have been made since the last interview.</p>
All HEOPF, HEDIAW and HEDIBW questions	These are the variables that tell the interviewer if a respondent has previously reported a condition. They are not used if the interview is by proxy.
HEDIAW	The feedback takes the answers as originally given to the interviewer, so ignores any recoding of “other” CVD conditions, e.g. if a respondent had said that they had a leaky heart valve at the previous interview, this would be recoded for the archived data set for that wave to Code 5 (variable HEDIMHM). However, for the current interview it would be fed forward as an “other” condition and the wording used by the respondent reported back to them for confirmation.
HEDIAS	Respondents who have previously reported the following conditions, and have not since denied every having them, are not asked if they still have the condition: heart attack, stroke, diabetes or high blood sugar. In Wave 2 this question was asked of those with diabetes or high blood sugar.
HEDIDS	Those who have previously reported the following conditions, and not since denied every having them, are not asked if they still have the condition: psychiatric condition, Alzheimer’s disease.
HEACD	Users will need to use the feed forward variable DHEACD to identify all those who have ever reported diabetes. In Wave 4 the variable applies to those who newly reported a diagnosis of diabetes or high blood sugar and to those who had reported diabetes or high blood sugar previously but had Code 2 at HEACD for their previous interview or who had not answered this question. In Wave 2 this question simply applied to those who had reported diagnosis of diabetes or high blood sugar either at Wave 1 or Wave 2.
HEHIBPB1, HEMDA1	These questions only apply to respondents who had had a stroke but had not already been asked about high blood pressure in the set of questions that specifically follow a diagnosis of high blood pressure (variables HEHIBPB and HEMDA).

Variable name (in CAPI questionnaire)	Note
HEHRTMD	Eligibility for this question has changed – previously it was those respondents with angina or myocardial infarction, but at Wave 4 it was extended to those with diabetes or stroke.
HEDIDS	The archived name of this variable has changed from HEDBLU at Wave 3 to HEDBSLU at Wave 4. At Waves 5 and 6 it is called HEDBLU again.
HEIQA – HEIQQ	Eligibility for this question has changed as a result of a change to the rules for proxy interviewing. The questions are intended to refer to those who are not capable of responding for themselves. Previously, proxy interviewers were only done in these circumstances but in Wave 4 the circumstances under which proxy interviews were allowed were extended. So, whereas previously this question applied to all proxies, in Wave 4 a condition was added to limit the questions to cases where the ELSA sample member was ill or away.
HEPAIN – HEDIZ	The pain questions were revamped for Wave 4 to enable a better fit to the ACOVE indicators (please refer to the “Wave 2 Quality of Care Indicators” document, available from the Data Archive, for more information). Although some questions have the same wording and eligibility analysts are reminded that the question order may have changed.
HEAND1	This question is not part of the standard Rose Angina question sequence, but was included because of problems with pinpointing whether those who report that they never/cannot walk do in fact experience chest pain that might indicate angina. The reference is to walking at an ordinary pace on level ground if the answer to HEANC is Code 4, otherwise to walking uphill. This variable was not included in Wave 6.
HERPE1	This question is not part of the standard MRC respiratory question sequence, but was included because of problems with pinpointing whether those who report that they never/cannot walk do in fact experience being short of breath. The reference is to walking with people of own age on level ground if HERPE is Code 3, otherwise to "uphill or hurry". This variable was not included in Wave 6.
HEKNEA	At Wave 3, this question referred to knee pain among those reporting osteoarthritis and knee pain in range 1-5 (HEKNE) and not hip pain in range 6-10 (HEHIP) whereas from Wave 4 it refers more simply and extensively to those who had knee pain whether severity in any of range 1-10 and reported osteoarthritis at question HEART (variable HEARTOA on dataset).
HEHIPA	This question applies to respondents who rated their hip pain

Variable name (in CAPI questionnaire)	Note
	between 6 and 10 (with 10 being severe/excruciating) at HEHIP and also reported osteoarthritis at HEART.
HEPMED	This question applies if the respondent reports osteoarthritis and either had knee pain for more than six months with a severity between 6 and 10 (same scale as above), or the equivalent for hip pain.
HEROSMD	In Wave 3 this question applied to those reporting chest pain at HEANK or chest pain at sites 1-4 at HEANI. From Wave 4, those applicable were only those reporting chest pains at HEANK.
HEADLB	Two new activities were added in Wave 4: appreciating risk of danger and communication. This means that the order of the difficulties on the questionnaire has changed but variable names on the dataset have been assigned so that the same name in successive waves refers to the same difficulty, e.g. HEADLMO refers to difficulty with managing money at Waves 4, 5 and 6 even though the position in the list of difficulties has changed.
HEHP SERIES and CA series	These questions were included from Wave 1 to Wave 5. Details of particular issues to note are provided below but only apply up to Wave 5. From Wave 6 questions about care were covered in a new module. All the variables from this module start CA. They follow on from the HEMOBWA and HEADLDR series. Those who reported needing help with any tasks were asked about whether they received help with a subset of the tasks and if so details were collected on who helped, how much help was received and how it was paid for.
HEHPSOT HEHPSCH HEHPSEX (UP TO w5) CAHPSOT CAHPSCH CAHPSEX (w6)	These questions about services used are equivalent between W5 and W6, but note that at W5 they referred to current use and took the form HEHPSOT etc and at W6 they referred to the last three months and took the form CAHPSOT etc.
HeClub (up to W5) CaClub (W6)	These are also equivalent but HeClub they have a different reference period. HeClub (up to Wave 5) refers to ever while CaClub (W6) refers to the last month.
HEHPB	Up to Wave 5 HEHPB applies to those who reported help at HEHPA, had difficulty with an activity not covered by the previous questions and also had not yet reported any specific source of help. These questions are to be used in conjunction with all the other help questions to find whether a particular source of help was used at all by a respondent. They do not necessarily cover all the sources of help for

Variable name (in CAPI questionnaire)	Note
	these residual activities.
HEHPB1 – HEHPM11, HEHPSM5 – HEHPSM9	For these questions, the categories of who has given the help have changed since Wave 3. So, for example, at Wave 3 code 3 on the interview list represented “Whether receives help moving around house from son ”, whereas at Wave 4 it represented “Whether receives help moving around house from daughter ”. However, the variable names on the dataset have been designed such that the same name should apply to the same grouping at each wave, e.g. HEHPBSO refers to help from son at both waves regardless of the code on the interview list. These only apply up to Wave 5
HEHPSM	<p>The coding frame for these questions was revised for Wave 4. Analysts are advised to check variable labels carefully if making any comparisons to Wave 3.</p> <p>Also note that Code 3 (variable HEHSMEX) may not include all responses of “swimming” or “dancing” mentioned in HEHPSO. The former has occasionally been coded to hydrotherapy (variable HEHSMHY) and both may also be included in ‘other’ (variable HEHSM85 on the dataset). These only apply up to Wave 5</p>
HEHPLA, HEHPPD	These are new questions which apply to those respondents who reported a particular source of help for at least one activity with which they have difficulty. These only apply up to Wave 5
HEMEB, HEHYY, HEOVY, HEPEE, HEHRT	The questions on menstruation are intended to update previous information collected during the life history interview.

➤ Notes about particular variables in the Risk module (Wave 5 only)

Variable name (in CAPI questionnaire)	Note
ritpcons	This is a categorical variable that indicates whether the time preference choices are consistent on tasks 1-6, on task 7-12, both, neither or both + consistent across the two sets
ridis1c	This is a categorical variable that assigns respondents to a patience category on the basis of tasks 1 to 6. the categories have upper and lower bounds for the weekly discount rate. The top and bottom categories are unbounded ("less than"... "more than") and there is a final category for those whose choices are inconsistent (and hence can't be classified).
ridis2c	As above but for tasks 7 to 12.