

## English Longitudinal Study of Ageing (ELSA) User Guide for the Wave 1 Core Dataset Version 3

This document accompanies the first version of data from Wave 1 of the English Longitudinal Study of Ageing (ELSA). The guide provides an outline of the sampling and methodology, content of the ELSA interview, and a description of the deposited dataset (including key variables).

The following Wave 1 files are deposited at the UK Data Archive ([www.esds.ac.uk](http://www.esds.ac.uk)), and should be used in conjunction with this User Guide.

**Table 1 Files deposited at the UK Data Archive**

Title	Format
ELSA Wave 1: Core dataset	Various statistical packages
ELSA Wave 1: Financial derived variables dataset	Various statistical packages
ELSA Wave 1: Pension wealth derived variables dataset	Various statistical packages
ELSA Wave 1 : Core Data User Guides and Documentation	Word (available as part of PDF documentation)
ELSA Wave 1: Variable relationships for financial derived variables	Excel
ELSA Wave 1: Financial derived variables name lookup	Text (details variable name truncation that may occur with SPSS data)
ELSA Wave 1: Pension wealth derived variables name lookup	Text (details variable name truncation that may occur with SPSS data)
<i>Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing.</i> Technical Report (National Centre for Social Research). To be deposited at a later date.	Word
<i>Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing.</i> Report on Wave 1 data.	Link to website: <a href="http://www.ifs.org.uk/elsa">www.ifs.org.uk/elsa</a>

### **BACKGROUND AND AIMS**

The English Longitudinal Study of Ageing (ELSA) is a study of people aged 50 and over and their younger partners, living in private households in England. The sample was drawn from households that had previously responded to the Health Survey for England (HSE) in 1998, 1999 or 2001.

Every two years we hope to interview the same group of people to measure change in their health, economic and social circumstances. ELSA can complete the picture

of what it means to grow older in the new century, and help us understand what accounts for the variety of patterns that are seen.

This User Guide relates to data deposited for the first wave of ELSA, which was carried out between March 2002 and March 2003. The data can be analysed cross-sectionally, or longitudinally in conjunction with HSE Wave 0 data and ELSA Wave 2 data, all of which is available from the Data Archive.

ELSA is the result of collaboration between University College London, the Institute of Fiscal Studies, and the National Centre for Social Research (NatCen). Other academic collaborators based at the Universities of Cambridge, Nottingham and Oxford provided expert advice on specific modules.

Funding for the first two waves of ELSA has been provided by the US National Institute on Aging, and a consortium of British Government departments, specifically: Department for Education and Skills; Department of Environment, Food and Rural Affairs; Department of Work and Pensions; HM Treasury; Inland Revenue; Office of the Deputy Prime Minister and 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 and Retirement in Europe (SHARE).

### **ETHICAL CLEARANCE**

Ethical approval for ELSA Wave 1 was granted from the Multicentre Research and Ethics Committee (Reference number MREC/01/2/91) approved on 7 February 2002.

### **DEVELOPMENT AND PILOTING**

Two pilots of the computer-assisted personal interview (CAPI) instrument, the self-completion questionnaire and associated documents were conducted in August and November 2001. All pilot respondents were drawn from households that participated in HSE 2000. The pilots tested the fieldwork procedure and interview content. Some new questions underwent cognitive testing, and some innovative measures were introduced. For example the use of unfolding brackets was used to militate against non-response problems on financial variables (see financial derived variables and imputation procedures).

### **SAMPLE DESIGN**

The ELSA sample has been designed to represent people aged 50 and over, living in private households in England. Three years of the Health Survey for England (HSE) were selected as the sampling frame: 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 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 sample (11,578 households containing 18,813 eligible individuals). The Technical Report provides more details.

### **Sample Types**

Within households there were three different types of individual who were eligible to be invited to take part in the study: core sample members, younger partners and new

partners. The variable **Finstat** determines the sample type of the respondent. Each type of respondent is described further:

- **Core sample members (C1CM)** are individuals who were living within the household at the time of the HSE interview and were born on or before 29<sup>th</sup> February 1952. This date was chosen to ensure that all sample members were aged 50 or over at the beginning of March 2002, i.e. in time for the start of ELSA fieldwork. In order for the individual to be eligible, the interviewer had to ascertain that the individual was living in a private residential address in England at the time of the ELSA interview. Eligible core members who responded to the ELSA survey form the baseline sample for analysis.
- **Younger partners (C1YP)** are the cohabiting spouses or partners of core sample members, who were living within the household at the time of the HSE interview and were born *after* 29<sup>th</sup> February 1952. In order for the individual to be invited to take part, the interviewer had to ascertain that he or she was still living with an eligible core sample member. Younger partners are not part of the core sample and should not be included in all analyses. Their information has been collected to make it possible to carry out an analysis of a representative sample of couples where at least one spouse is 50 or older.
- **New partners (C1NP1)** are the cohabiting spouses or partners of core sample members at the time of the first ELSA interview, who had joined the household since the HSE interview. Like young partners, they are not part of the sample and are should not be included in all analyses. Their information has been collected to make the data as complete as possible.

In the Wave 1 deposited dataset there are 11,392 core members, 636 younger partners, and 72 new partners.

### **ACHIEVED SAMPLE**

The ELSA Wave 1 fieldwork produced 12,100 productive interviews. As explained above, 11,392 of these were with core sample members. Of all core sample member interviews, 204 were partial responses and 158 were proxy responses (see variables **Indoc** and **Rthhout**). In addition 636 productive interviews were conducted with younger partners, and 72 with new partners.

Of the 11,578 households that were issued for ELSA, the interviewer made contact with the household at 94% of them (the household contact rate). Nearly two thirds of non-contacts had moved since the HSE interview and could not be traced. Separate analyses show that around a tenth of issued households had moved house by the Wave 1 fieldwork period. Two thirds of the movers were eventually traced.

A responding household is defined as one where at least one eligible person was interviewed. Using this criterion, a household response rate of 70% was achieved. The majority of non-responding households refused to participate (22% of the eligible sample of households).

A small percentage of respondents within the responding households were ineligible (3%). Of the remaining sample of individuals within responding households, a response rate of 96% was achieved. Non-response within households was almost always because of refusal to take part.

The overall response rate for individuals (calculated using the total number of eligible individuals within all issued households) was 66%.

More information about the response rates is provided in the Wave 1 Technical Report.

### **WEIGHTING**

The variable in the dataset to be used for weighting is **W1wgt**. Weights were calculated for core sample members only (including proxy and partial interviews), as this is the sample of interest. All other non-sample individuals that were interviewed (i.e. new and younger partners) have a weight of zero. When running weighted analyses, researchers should remember to exclude non-sample members in the unweighted base, if quoted. The data for partners can be used as characteristics of the core members (i.e. supplementary information).<sup>1</sup>

Where possible we recommend that analysis be conducted on weighted data since this should help to reduce bias from non-random non-response. The aim of weighting is to take account of any bias from non-response in order to make the respondent sample more representative of the population. The equal probability design of the HSE samples, and the fact that the ELSA sample included all eligible adults from the HSE, eliminates any need for weights to account for selection probabilities. However, non-response at HSE, refusals to be re-interviewed post-HSE and non-response at ELSA Wave 1 all have the potential to make the ELSA respondent sample unrepresentative of the population. In addition, the original complex sample design of the HSE samples has to be considered. The HSE sample design is described in the ELSA Technical Report.

A thorough analysis of non-response was conducted for ELSA to examine different stages of drop out, and the extent of drop out at each stage.

Two stages were identified where there had been a significant level of drop out that was likely to result in bias and, as a result, required weighting. The two stages were:

- Drop out of households where none of the age-eligible individuals *consented to be re-approached for a further interview*,
- Drop out of households where no one agreed to be interviewed (that is, Wave 1 household level non-response).

The predicted probability of response for the responding households was inverted to provide the initial non-response weight.

A further round of weighting was needed to adjust the initial household non-response weight to ensure that the weighted responding sample of individuals matched the population of interest. This adjustment helps to account for any bias caused by households non-responding to HSE. The calibration method proposed by Lemaitre and Dufour (1987)<sup>2</sup> was appropriate to use in this situation (for further details see Deville and Sarndall (1992)<sup>3</sup>). The rationale behind calibration weighting is that it attaches an estimated probability of response to each household that 'explains' the discrepancy between the survey age-sex distribution and the population age-sex distribution. Each individual in the same household has the same value of weight.

---

<sup>1</sup> If non-core sample members are to be analysed they should be analysed unweighted.

<sup>2</sup> Lemaitre, G.E. and Dufour, J. (1987). An integrated method for weighting persons and families. *Survey Methodology*, 13, 199-207.

<sup>3</sup> Deville J-C and Sarndal C-E (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

This weight can be used for individual-level and household-level estimation. This is illustrated in the example below. A key advantage of this approach is that, because the household and individual weights are identical, in the absence of substantial within-household non-response, estimates about individuals derived from household-level data should match estimates derived from the individual level data.

Example:		<i>Individual weight</i>	<i>Household weight</i>
	HH 1 Person 1	w1wgt =1.01	HH 1 Weight=1.01
	HH 1 Person 2	w1wgt =1.01	
	HH 2 Person 1	w1wgt =0.98	HH 2 Weight=0.98
	HH 3 Person 1	w1wgt =1.05	HH 3 Weight=1.05
	HH 3 Person 2	w1wgt =1.05	
	HH 3 Person 3	w1wgt =1.05	

### Standard Errors

Standard errors for survey estimates should account for the complex sample design and the weights. There are two aspects of the sample design that impact on standard errors: clustering and stratification. Clusters can be identified with variable **Ahsecls2**. The stratification can be replicated using variable **Astratif**. Strictly speaking, the age-sex post-stratification should also be accounted for, but it is difficult to achieve this in most software packages. The post-stratification is likely to reduce standard errors of estimates slightly so, by ignoring this, the standard errors calculated may be slightly conservative.

The ELSA Wave 1 Technical Report illustrates the difference between true standard errors (i.e. those that take the design into account) and uncorrected standard errors. Design factors for key estimates tend to fall within the range 1.0 - 1.7, where values above 1.2 are commonly described as indicating sizeable variance inflation.

The age-sex distribution of the unweighted and weighted data for core sample members is given in Table 2 below. The weighted distribution is closer to the population distribution than the unweighted distribution. The most significant changes can be seen in the percentages of 50-54 year old men, 55-59 year old women and 85+ year old women.

**Table 2 Age-sex distribution (unweighted and weighted)**

Ageband (years) at W1	Unweighted			Weighted		
	Male %	Female %	Total %	Male %	Female %	Total %
50-54	20	20	20	23	20	22
55-59	19	18	19	18	16	17
60-64	15	14	15	15	14	15
65-69	15	14	15	14	13	13
70-74	13	13	13	12	12	12
75-79	9	10	9	9	11	10
80-84	6	8	7	5	7	7
85+	3	4	4	3	6	5
<b>Total</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

### SURVEY CONTENT

The ELSA Wave 1 interview 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 the case of couples, interviewed at the same time (in a single session) using concurrent interviewing techniques. In a concurrent session (see Survey Module Identifiers section below) the same block of questions was asked alternately of each person. The order in which the respondents answered the questions (i.e. who answered first) was randomly pre-set at the start of the concurrent session. Interviewers could override this order if they thought it necessary for the successful completion of an interview though this occurred rarely. Concurrent interviews tended to be quicker than two separate individual interview sessions, and were generally more convenient for respondents.

ELSA Wave 1 adopted the use of dependent interviewing in many of its modules. The technique is also 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. Earlier responses can either be provided to the individual before they respond to a question (proactive) or can be used after the individual has responded as a form of validation (reactive). Dependent interviewing was used to check some information collected during HSE, to determine changes in status, and to control routing within the questionnaire. Data from HSE has been combined with equivalent ELSA variables if the data was missing or not asked during the ELSA interview.

There were various modules each covering a different area of interest. The content and major routing of each module is described 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.

Four of the modules (Cognitive Function, Expectations, Psychosocial Health and Final Questions) form the 'private modules' block. Wherever possible, these modules were administered with no other household members present. The second responding individual was asked to fill in the self-completion in a separate room in the meantime. The two respondents then switched places.

The questionnaire modules are listed below in the order programmed in the CAPI interview. Most of the names of the variables (including derived variables) in each of the modules start with the same two-letter module prefix (shown in the brackets below). However, there are a small number of variable names that do not follow this convention.

#### Household Demographics ("DH")

The household demographics module was answered by one person on behalf of the household. It was used to collect basic demographic information about everyone living in the household. The composition of the household given at HSE was proactively fed forward and the interviewer checked if all the people were still present in the household. The ELSA program determined the eligibility for the ELSA interview for each person in the household (see sample design section).

Note that information provided by the respondent to these questions is copied onto the data for other members of the household.

### Individual Demographics (“DI”)

This module is at the start of the ELSA interview. Each respondent was asked details about their legal marital status, living children including adopted, foster and stepchildren, number of grandchildren and great-grandchildren, number of siblings and their own circumstances in childhood.

### Health (“HE”)

This module was administered to all respondents. It covered many different dimensions: self-reported general health: long-standing illness or disability; eyesight and hearing; specific diagnoses and symptoms; pain; difficulties with activities of daily living (ADLs) and instrumental activities of daily living (IADLs); and health behaviours. Respondents aged 60 and over were asked about falls and fractures.

### Social Participation (“SP”)

This module was administered to all respondents. It covered the frequency with which respondents participated in certain social activities, whether they were limited from participating in these activities, their care giving and use of public transport.

### Work and Pensions (“WP”)

This module was administered to all respondents. It covered respondents’ current work activities and any current or past pensions that they had. If the respondent was retired and was receiving a pension, details were collected about their pensions and any amounts they received.

### Income and Assets (“IA”)

At the start of the interview couples were asked whether they kept their finances together or separate (**Dhiasep** – this variable is not in the data). If kept together, they were considered to be a single financial unit that required only one respondent for the IA module (the “financial respondent”). The couple decided who the respondent would be. If their finances were kept separately, each person needed to answer the IA module separately and were treated as two separate financial units. The type of financial unit the respondent was in is stored in a variable called **Futype**. The variable that identifies who answered the questions in IA is called **lapid**. Note that the interviewer can choose to answer IA at a later time. This added flexibility for respondents and interviewers and inevitably means that a small number of people did not answer IA at all. **lapid** takes the value 0 when no one in the financial unit answered IA.

The information provided by the financial respondent is copied across to the other member of the financial unit (where this is appropriate). One important implication of this is that where a question refers to a person’s “partner” or “spouse” (e.g. **laspp** “how much does your husband/wife/partner receive from the state pension”), for individuals whose partners answered IA on their behalf, the data recorded in this variable will actually be referring to them (rather than their partner).

Details of the income that respondents received from a variety of sources over the last 12 months were collected including wages, state pensions, private pensions, and other annuity income and state benefits. In addition, this module covered the amount of financial and non-financial assets held, any income from these assets, regular transfers from non-household members and one-off payments in the last year.

### Housing (“HO”)

Only one eligible ELSA respondent in the household answered the housing module. Respondents decided themselves who the housing respondent should be. Information provided by the HO respondent is copied across to other members of the

household. The variable **Hopid** identifies who answered HO on behalf of the household.

This module collected information about current housing situation (including size and quality), housing-related expenses, ownership of durable goods and cars, and expenditure on food. House owners and people with mortgages were asked about the value of their property, and details of their mortgages, rent, etc.

#### Cognitive Function (“CF”)

This module was the start of the ‘private modules’ block, i.e. those which were administered with no other household members present. The CF module was asked of all respondents (except proxies – see Proxies section) and measured different aspects of the respondent’s cognitive function, including memory, speed, mental flexibility and numeracy.

#### Expectations (“EX”)

This module was asked privately of all respondents. It measured people’s expectations in a number of dimensions, the level of certainty respondents felt about the future, financial decision-making within households and the time frame they thought about when making financial decisions.

#### Psychosocial health (“PS”)

This module was asked privately of all respondents. It measured how respondents viewed their lives across a variety of dimensions and included a mental health scale and questions about respondents’ attitudes towards ageing.

#### Final questions (“FQ”)

This module was asked privately of all respondents. It asked for demographic information, a stable address contact, and consent to obtain health and economic data from administrative sources.

#### Walking speed (or Measurement – “MM”)

This module could be conducted at any point in the interview after the HE module. The test of walking speed was completed by respondents aged 60 and over when it was judged safe to do so. Respondents were asked to walk a distance of 8 feet (244 cm) at their usual walking pace. This walk was performed twice by respondents, and the interviewer recorded the time taken using a stopwatch (see protocol in interviewer instructions).

#### Self-completion (“SC”)

When respondents completed a full interview in a session alone, the self-completion questionnaire was usually left at the end to be returned by the respondent by post. However, when two respondents completed the interview in a concurrent session, the self-completion questionnaire was completed by one respondent while the other carried out the “private” modules of the personal interview (i.e. CF, EX, PS, and FQ).

### **PROXIES**

If an eligible respondent was physically or cognitively impaired, or in hospital or temporary care for the whole of the fieldwork period, a proxy interview was permissible. Interviewers were asked to identify a proxy informant (i.e. a person who could answer the interview on behalf of the eligible respondent). The proxy informant was a 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, close family members such as a partner or son or daughter fulfilled this role. Proxy interviews were not conducted for people found to be in institutions.



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 (the asterisked modules were not asked for all respondents – see below):

HD*	Household grid
ID	Individual demographics
HE	Health
WP	Work and Pensions
IA*	Income and Assets
HO*	Housing
FQ	Final questions and consents

All proxy informants completed ID, HE, WP and FQ. Some specific questions for proxies were included in the Health module, HE.

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 by 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.

**The variable Askpx1 identifies proxy interviews. 158 proxy interviews were for core members, 6 were for new partners, and 11 were for younger partners.**

### **DATASET INFORMATION**

The core dataset has been archived for ELSA Wave 1. This lists cases at an individual level.

Data for ELSA respondents collected during HSE can be used for longitudinal analysis (HSE data is defined as Wave 0). The Wave 0 dataset will be available from the data archive shortly, and can be linked to ELSA Wave 1 data using an Index file (to be deposited simultaneously).

The ELSA Wave 1 core dataset contains data in the following order:

- Key variables not in the questionnaire (e.g. serial number, weighting variable)
- Variables in the questionnaire (in the order they appear in the CAPI interview). A small number of additional variables that are associated with particular questionnaire variables are located alongside these variables in the data. These are annotated in the questionnaire.
- Other variables not in the questionnaire (including administrative variables, other derived variables and variables fed forward from HSE)

See Appendix for the full list of variables not included in the questionnaire.

The main group of respondents for analysis is the core members. Data on partners (younger partners and new 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.

## **SERIAL NUMBERING**

### **Wave-specific household serial number**

The five digit household serial number (**Idahhw1**) was randomly generated for the archived dataset, and does not relate to the serial number used during interviewing.

### **Wave-specific individual serial number**

**Idaindw1** is the individual serial number. This serial number is independent of the household serial number and the serial number used during interviewing. Individuals who had joined the household after the HSE interview were given a number following on from the highest number recorded at the HSE interview.

### **Constant serial numbers**

When deposited, the dataset for Wave 0 (HSE) and future ELSA waves will each contain a different set of individual (**Idaindw0**, **Idaindw2** etc.) and household (**Idahhw0**, **Idahhw2** etc.) serial numbers. An index file will be provided which will enable data users to link all these serial numbers in order to compare data for each respondent at different waves. There is also a unique individual analytical serial number (**Idauniq**) included in the Wave 1 core dataset, which will be added to other ELSA datasets so that they can be linked.

### **Person number**

Each person within the household was given a number, starting from 1, 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 indicate which one of the other people in the household is the respondent's partner (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 **lapid**).

## **INTERVIEW OUTCOME**

Individual and household level interview outcomes are given in **Indoc** and **Rthout** respectively.

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).

## **SURVEY MODULE IDENTIFIERS**

Some modules in the survey interview were answered by all interviewed individuals whereas others were not. Each of the modules that could be answered by one individual on behalf of others has a variable that identifies the person who answered the module:

- Household Demographics = **Hhresp**
- Income and Assets = **lapid**
- Housing = **Hopid**

Where modules are answered by one person on behalf of others, information is copied directly to other member of the household or financial unit.

The variable **Nump** indicates whether or not the respondent was interviewed alone (**Nump=1**) or concurrently with another respondent in the household (**Nump=2**). The person number of the individuals allocated to each concurrent session and the order in which people were interviewed are provided in **Allocp1** and **Allocp2**.

### **SAMPLE TYPE**

**Finstat** indicates the status of the individual, that is, whether they were interviewed as a core sample member (C1CM), younger partner (C1YP) or new partner (C1NP1). Note that the data set contains other status variables, for example the status at sampling (**Elsa** and **Partner**), but **Finstat** is the final version that should be used for analysis.

### **AGE**

**Dhdobyr** and **Dhager** provide the date of birth and age of respondents recorded in the household grid. Note that a given respondent may not have provided this information themselves (as anyone in the household can complete the household grid).

**Didbny** and **Diag** are the date of birth (year only) and age given in the individual interview.

**Indobyr** is derived from the date of birth variables from the household grid (**Dhdob**) and individual interview (**Didbn**). If the variable from the individual session was answered (**Didbny**) then this code was imputed, otherwise the variable from the household grid (**Dhdob**) was used in the derivation. **Indobyr** provides the year of date of birth only; the day and month of birth have been dropped from the dataset to retain confidentiality.

**Indager** is computed from date of birth (**Indob** – dropped from dataset) and date of interview (**Intdat** – dropped from dataset). All respondents over age 90 have been classified as 99 years old for confidentiality reasons.

For age analysis, **Indobyr** and **Indager** are the suggested variables to use.

Please note the derived variables used to create the following age tables are not included in the archived dataset. These variables will be provided as part of a derived variable dataset at a later date.

**Table 3 Age of Core Members**

<b>Age</b>	<b>N (unweighted)</b>	<b>N (weighted)</b>
50-54	1981	2211
55-59	2185	2035
60-64	1688	1686
65-69	1711	1552
70-74	1471	1397
75-79	1094	1164
80+	1262	1347
<b>Total</b>	<b>11392</b>	<b>11392</b>

**Table 4 Age of Partners (no weighting assigned to partners)**

<b>Age</b>	<b>N Younger partners</b>	<b>N New partners</b>	<b>Total</b>
20-34	9	5	14
35-39	41	5	46
40-44	119	4	123
45-49	388	6	394
50-54	79	14	93
55-59	0	17	17
60-64	0	8	8
65-69	0	7	7
70-74	0	6	6
<b>Total</b>	<b>636</b>	<b>72</b>	<b>708</b>

**GENDER**

**DiSex** was the sex given in the respondent's individual questionnaire session.

**Dhsex** was given in the household demographics module. **Indsex** is derived from these two variables. If the variable from the individual session was answered (**Disex**) then this code was imputed, otherwise the variable from the household grid (**Dhsex**) was used in the derivation.

**Table 5 Sex of Core Members**

	<b>N unweighted</b>	<b>N weighted</b>
Male	5187	5281
Female	6205	6111
<b>Total</b>	<b>11392</b>	<b>11392</b>

**Table 6 Sex of Partners (no weighting assigned to partners)**

	<b>Younger partners</b>	<b>New partners</b>	<b>Total</b>
Male	120	29	149
Female	516	43	559
<b>Total</b>	<b>636</b>	<b>72</b>	<b>708</b>

### **SOCIAL OCCUPATIONAL AND SOCIAL INDUSTRY CODING (SOC/SIC)**

In the Work and Pensions module of ELSA Wave 1, we only ask job details and code to SOC2000/NS-SEC and SIC92 if respondents' current or most recent job is not their job at the HSE interview, or if we did not have SOC coding from HSE.

Job details collected in the ELSA interview are coded to SOC2000/NS-SEC and SIC92. The corresponding variables produced are **Exs2000** and **Esic92** (this variable is not in the core dataset). The key variable is **Enssec**.

In HSE 1998 and 1999 job details were originally coded to SOC90. For comparability with ELSA, these were subsequently recoded to SOC2000/NS-SEC and coded to SIC92. The variables produced are **Axsc2000** and **Ahsic92** (this variable is not in the core dataset). The key variable is **Anssec**.

In HSE 2001 the job details were coded to SOC2000/NS-SEC, and SIC92. These codes have been merged into the **Axsc2000** and **Ahsic92** (see above). The key variable is **Anssec**.

Therefore:

**Enssec** = coding for job details collected in the ELSA interview.

**Anssec** = coding for job details collected during the HSE interview.

Shorter versions of NS-SEC (e.g. 8 category classification) are not included in this dataset, but will be provided as part of a derived variable dataset at a later date.

### **EDUCATION**

If a respondent's qualifications had been collected during the HSE interview, only additional qualifications gained since the last interview were recorded at ELSA Wave 1. Those not interviewed at HSE, were asked for all qualifications.

The variable **Aqual** indicates if the respondent had qualifications at HSE. The qualifications collected during the ELSA Wave 1 interview are **Fqquzm1** to **Fqquzm3**. The same qualification codes were used in HSE (1998, 1999, 2001) and ELSA Wave 1 (apart from 'other diploma' which was added to the ELSA variables after the interview). The grouped highest education qualification at ELSA Wave 1 (based on responses given at HSE and at variables **Fqquzm1** to **Fqquzm3** at ELSA Wave 1) are in **Edqual**.

### **HOUSEHOLD TYPE**

Benefit unit classification is based on all individuals in an ELSA household. The derivation identifies households with a single benefit unit comprised of eligible ELSA respondents only, and households with multiple benefit units (with eligible or ineligible ELSA respondents).

The variable **Buclass** shows the how each ELSA respondent has been classified according to benefit unit.

Other derived variables include **Bubreak**, which can be used to aggregate individual level file to benefit unit level.

**Buclass** and **Bubreak** are not available in the core dataset – they will be released at a later date as part of a derived variables dataset.

**Table 8 Benefit unit (BU) classification (estimate)**

	<b>No. of respondents</b>
Single person BU (ELSA eligible)	2912
Two person BU (ELSA eligible)	6828
Multiple BUs in HH (ELSA eligible/ineligible)	2360
<b>Total</b>	<b>12100</b>

### **FINANCIAL VARIABLES**

ELSA contains very detailed information on all aspects of finances. Summary variables have been derived and are available in a separate dataset (“Financial derived variables”). To derive these variable, some imputation was carried out and the methods used are described in “Financial DVs and imputation procedures.doc”. Key variables contained in this supplementary dataset are:

Totinc_bu	benefit unit	total income
Empinc_bu	benefit unit	employment income
Seinc_bu	benefit unit	self-employment income
Ppinc_bu	benefit unit	private pension income
Spinc_bu	benefit unit	state pension income
Beninc_bu	benefit unit	benefit income
Assinc_bu	benefit unit	asset income
Othinc_bu	benefit unit	other income
Eqtotinc_bu	benefit unit	equivalised total income
Eqempinc_bu	benefit unit	equivalised employment income
Eqseinc_bu	benefit unit	equivalised self-employment income
Eqppinc_bu	benefit unit	equivalised private pension income
Eqspinc_bu	benefit unit	equivalised state pension income
Eqbeninc_bu	benefit unit	equivalised benefit income
Eqassinc_bu	benefit unit	equivalised asset income
Eqothinc_bu	benefit unit	equivalised other income
Savings_bu	benefit unit	total savings
Invests_bu	benefit unit	total investments
Grossfw_bu	benefit unit	total gross financial wealth
Debt_bu	benefit unit	total debt
Netfw_bu	benefit unit	total net financial wealth
Netpw_bu	benefit unit	total net physical wealth
Grosstotnhw_bu	benefit unit	total gross non-housing wealth
Nettotnhw_bu	benefit unit	total net non-housing wealth
Ghwealth_bu	benefit unit	total gross housing wealth
Nhwealth_bu	benefit unit	total net housing wealth
Nettotw_bu	benefit unit	total net (non-pension) wealth

## **DROPPED VARIABLES**

All variables in the questionnaire documentation with a @ symbol next to their name have been deleted from the archived dataset (or have been recorded in multiple variables instead).

The following types of variables have been deleted in order to reduce the potential to identify individuals and for other reasons (specified below):

1. Those containing text
2. Those which contained a personal identifier (e.g. name/address)
3. Those considered to be disclosive, such as:
  - Detailed ethnicity
  - Specific country of birth
  - Full interview date
  - Full date 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
5. Variables that only contain missing values – excluded because not useful.

There are no geographical variables in the archived dataset. A separate geographical dataset with broad identifiers is available under secure arrangements

## **CODING AND EDITING**

Additional coding and editing tasks were performed after the interviews were conducted. The ELSA Wave 1 Code Book and Edit Instructions document (part of the main User Guide) provides details of the tasks that were conducted.

### **CODING**

The coding of responses was mostly dealt with by the CAPI questionnaire through the use of fully closed questions. There were, however, a number of questions (e.g. **Wphlwho**) where a code frame was not used in order to capture all responses to a particular question verbatim. The responses to these 'open' questions were coded into separate variables after the interview was conducted (e.g. open responses to **Wphlwho** were coded into **Wphlwz**). The coded responses to open questions are in the core dataset, but as mentioned above the original text responses have been removed to reduce the potential to identify individuals.

Other questions in the CAPI questionnaire had code frames, which 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**).

If these other-specify questions were 'single coded', i.e. when only 1 option could be chosen (e.g. **Dikliv**) then the text answers were coded and incorporated into the original coded variable (i.e. **Diklo**).

For all multi-coded<sup>4</sup> 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. **Spcab1** to **Spcab4**). The second set of variables contain the original coding plus the codes assigned to “other answers” (e.g. **Spcam01** to **Spcam04**). Note that suffix ‘m’ added to these variable names means that they contain ‘merged’ original and coded answers. This naming convention was followed consistently so the final merged variables can be identified by name. The merged variables should be used instead of the original variables.

The original variables provide multi-coded answers in the order in which they were mentioned by the respondent. Please note that in some cases the order of the respondent’s answers in the merged variables is not the same as in the original variables.

### **EDITING**

As with the coding, most of the editing for the ELSA Wave 1 questionnaire was carried out by the interviewers in the field. However, there were a couple of additional checks that related to inconsistencies in the data that were carried out after the interview.

### **MISSING VALUES**

For most questions there are the following missing values:

- 1 Not applicable
- 8 Don’t know
- 9 Refusal

For some questions, a response of ‘don’t know’ or ‘refusal’ was not permitted. This is indicated in the questionnaire.

For various reasons, some respondents did not complete the questionnaire. Where possible, for these ‘partial’ interviews, the questions that were not asked were coded as –1 (not applicable). This will be checked and refined in future versions of the dataset.

---

<sup>4</sup> A variable is multi-coded if more than one response can be given e.g. favourite activities are reading *and* going to the cinema.



## **APPENDIX: Variables not included in the questionnaire**

<b>Variable name</b>	<b>Variable label</b>	<b>Added information</b>
<b>BEGINNING OF DATASET</b>		
idaindw1	Analytical wave-specific individual serial number	
idahhw1	Analytical wave-specific household serial number	
idauniq	Unique individual serial number	
perid	Person ID (same as person number in household grid)	
finstat	Post-field final type of sample member (including cohort number added)	
indoc	Final individual outcome code	
w1wgt	W1 weight for all core SM to account for non-response	
<b>END OF DATASET</b>		
intdatm	Month of Household Interview	
intdaty	Year of Household Interview	
iintdtm	Month of Individual Interview: Month of date: TODAY~S DATE	
iintdty	Year of Individual Interview: Year of date: TODAY~S DATE	
rthhout	Final household outcome code	
eligw1	Eligibility at W1	
elsa	HSE Feed Forward: ELSA Sample Member	
partner	HSE Feed Forward: Partner of ELSA Sample Member	
eligat	Total Eligible for interview based on HSE feedforward	
eligbt	Total Eligible for interview AFTER grid	
eligct	Total Eligible for interview AFTER DHPProxy	
hhsel	Number of respondents in HH selected for Individual Questionnaire	Age variable combined info from HH grid and individual demographics collapsed at 90 plus
nofiq	Number of interviewing sessions in household	
indno	Individual questionnaire session number.	
nump	Number of respondents in interviewing session	
allocp1	Person number of first person in interviewing session: PLEASE ENTER THE PERSON NUMBER OF THE PERSON TO BE INTERVIEWED IN THIS SESSION	
allocp2	Person number of second person in interviewing session: PLEASE ENTER THE PERSON NUMBER OF THE PERSON TO BE INTERVIEWED IN THIS SESSION	
adresp1	Person answering on behalf of first person in interviewing session (if proxy): WHO IS ANSWERING	

	ON BEHALF OF [RESPONDENT]?	
adresp2	Person answering on behalf of second person in interviewing session (if proxy): WHO IS ANSWERING ON BEHALF OF [RESPONDENT]?	
cpid	Person number of partner	
askpx1	Whether interviewed by proxy	
indrs	INTERVIEWER: CODE REASON WHY PERSON [Person number] WAS UNABLE TO COMPLETE INDIVIDUAL INTERVIEW	
w1mover	Whether moved between HSE and ELSA W1	
hopid	Person who answered HO for household	
iapid	Person who answered IA for Financial Unit	
askpay	Person number of IA respondent to answer questions about income of non-eligible household members (IaPayW - IaOm - asked once per household)	
futype	Financial unit type	
indsex	Sex - Priority: DiSex, DhSex	Dhsex and Disex are combined and recorded in Indsex – IF Disex was answered then Indsex=Disex, ELSE Indsex=Dhsex
indobyr	Year of birth combined HH grid and individual demographics collapsed at 90 plus	Year of birth from Dhdob and Didbn are combined and recorded in Inddoby – IF Didbn was answered then Indobyr=Didbn, ELSE Indobyr=Dhdob.
indager	Age variable combined info from HH grid and individual demographics collapsed at 90 plus	Age computed from IndDob + Intdat is recorded in Indager
anssec	FROM HSE: NS-SEC - long version	
axsc2000	FROM HSE: SOC2000 (without dots)	
enssec	ELSA NS-SEC	
exs2000	ELSA SOC2000 (without dots)	
hhtot	Number of people in ELSA household	
chinh1	Whether or not has a child in household (derived from household grid)	
chouth1	Whether or not has a child outside household (derived from household grid)	
gcinh1	Whether or not has a grandchild in household? (derived from children grid)	
mainhh1	Whether or not has a mother in household? (derived from household grid)	
painhh1	Whether or not has a father in household? (derived from household grid)	
couple1	Relationship status (derived from household grid)	
whoso1	INTERVIEWER: WAS THERE ANYONE OTHER THAN YOU AND RESPONDENT(S) IN THE ROOM DURING THE INTERVIEW SO FAR? (Before CF)	

whoso2	INTERVIEWER: WAS THERE ANYONE OTHER THAN YOU AND RESPONDENT(S) IN THE ROOM DURING THE INTERVIEW SO FAR? (Before CF)
whoso3	INTERVIEWER: WAS THERE ANYONE OTHER THAN YOU AND RESPONDENT(S) IN THE ROOM DURING THE INTERVIEW SO FAR? (Before CF)
scnosc	CODE WHY RESPONDENT DID NOT COMPLETE SELF-COMPLETION
scnoscc	Whether answer to Scnosc was recoded post-interview from text answer
sc_rec	Self-completion receipt
cogrec	Cognitive function booklet received
ahseyear	HSE Year
asampsta	Type of sample member at HSE
ahsecls2	Data Archive HSE clustering variable
astratif	HSE stratification variable
ahseint	HSE Feed Forward: Whether interviewed at HSE
ahsest	HSE Feed Forward: HSE Status
ahhsize	HSE Feed Forward: HSE Household size.
apersno	HSE Feed Forward: HSE Person number in Household Grid.
amintb	HSE Feed Forward: DATE OF HSE INTERVIEW (MONTH)
ayintb	HSE Feed Forward: DATE OF HSE INTERVIEW (YEAR)
asex	HSE Feed Forward: HSE SEX.
arelto01	HSE Feed Forward: HSE Relationship to other persons in household
arelto02	HSE Feed Forward: HSE Relationship to other persons in household
arelto03	HSE Feed Forward: HSE Relationship to other persons in household
arelto04	HSE Feed Forward: HSE Relationship to other persons in household
arelto05	HSE Feed Forward: HSE Relationship to other persons in household
arelto06	HSE Feed Forward: HSE Relationship to other persons in household
arelto07	HSE Feed Forward: HSE Relationship to other persons in household
arelto08	HSE Feed Forward: HSE Relationship to other persons in household
arelto09	HSE Feed Forward: HSE Relationship to other persons in household
arelto10	HSE Feed Forward: HSE Relationship to other persons in household
alivemab	HSE Feed Forward: HSE Natural mother still alive
aagemab	HSE Feed Forward: HSE Age of natural mother
alivepab	HSE Feed Forward: HSE Natural father still alive
aagepab	HSE Feed Forward: HSE Age of natural father
aageangi	HSE Feed Forward: HSE Age of diagnosis of angina
aagehart	HSE Feed Forward: HSE Age of diagnosis of heart attack
aagestro	HSE Feed Forward: HSE Age of diagnosis of stroke

aagedi	HSE Feed Forward: HSE Age of diagnosis of diabetes	
anactiv	HSE Feed Forward: HSE Activity last week	
aeverjob	HSE Feed Forward: HSE Ever had a job	
aemploye	HSE Feed Forward: Are you ...{an employee or self-employed}	
asoccls	HSE Feed Forward: HSE Social Class	
aeconact	HSE Feed Forward: Economic Status	
astwork	HSE Feed Forward: Did you do any paid work in the seven days ending [last Sunday~s date], either as an employee or self-employed?	
awklook	HSE Feed Forward: Thinking now of the four weeks ending [last Sunday~s date]. Were you looking for any paid work or Government training scheme at an +	Full question: HSE Feed Forward: Thinking now of the four weeks ending [last Sunday~s date]. Were you looking for any paid work or Government training scheme at any time in those four weeks?
awkstrt	HSE Feed Forward: If a job or a place on a Government training scheme had been available would you have been able to start within two week?	
aothpaid	HSE Feed Forward: Apart from the job you are waiting to take up, have you ever been in paid employment or self-employed?	
aftptime	HSE Feed Forward: Working full-time or part-time?	
adirctr	HSE Feed Forward: Can I just check, in this job [are/were] you a Director of a limited company?	
aempstat	HSE Feed Forward: Are you a ...? {manager/foreman or supervisor/other employee}	
anemplee	HSE Feed Forward: Including yourself, about how many people are/were employed at the place where you usually work?	
asnemple	HSE Feed Forward: [Do/did] you have any employees?	
ahhldr01	HSE Feed Forward: In whose name is the accommodation owned or rented? (1st mention)	
ahhldr02	HSE Feed Forward: In whose name is the accommodation owned or rented? (2nd mention)	
ahhldr03	HSE Feed Forward: In whose name is the accommodation owned or rented? (3rd mention)	
ahhldr04	HSE Feed Forward: In whose name is the accommodation owned or rented? (4th mention)	
atenureb	HSE Feed Forward: HSE Tenure	
aqual	HSE Feed Forward: HSE Qualifications	
aeducend	HSE Feed Forward: HSE Age education ended	
apobr	HSE Country of birth collapsed into UK and elsewhere to avoid disclosure	
aethnic	HSE ethnic group collapsed into White and Non-white to avoid disclosure	