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

PREFACE

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 datafile (including key variables).

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

Title	Format
ELSA Wave 1: Dataset	SPSS
ELSA Wave 1 : CAPI Questionnaire Documentation (National Centre for Social Research)	Word
ELSA Wave 1: Code Book and Edit Instructions (National Centre for Social Research)	Word
Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing. Technical Report (National Centre for Social Research). To be deposited at a later date.	Word
ELSA Wave 1: Interviewer instructions (National Centre for Social Research) To be deposited at a later date.	Word
Health, wealth and lifestyles of the older population in England: The 2002 English Longitudinal Study of Ageing. Report on Wave 1 data.	Link to website: www.ifs.org.uk

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 (to be released shortly). ELSA Wave 2 data will be available approximately six months after the completion of fieldwork**.**

ELSA is managed by a team based at 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).

SURVEY 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 they 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. This boost sample was discarded for ELSA. 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 recontacted at some time in the future. Individuals in the remaining households provided the basis for the ELSA sample (11,580 households containing 18,822 eligible individuals).

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 29th 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* 29th 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 sample and are not included in analyses as individuals in their own right. 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 not included in analyses as individuals in their own right. Their information has been collected to make the data as complete as possible.

In the Wave 1 deposited datafile there are 11,392 Core Members, 636 Younger Partners, and 72 New Partners.

SERIAL NUMBERING

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.

Individual serial number

idaindw1 is the individual serial number. 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.

When deposited, the dataset for wave 0 (HSE) and future ELSA waves will each contain a different set of individual and household 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.

ETHICAL CLEARANCE

Ethical approval for ELSA Wave 1 was granted from the Multicentre Research and Ethics Committee.

DEVELOPMENT AND PILOTING

Two computer-assisted personal interviews (CAPI) pilots were conducted in August and November 2001. All pilot respondents lived in a household which 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 mitigate against non-response problems on financial variables.

ACHIEVED SAMPLE

The ELSA wave 1 fieldwork produced 12100 productive interviews. 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 variable **rthhout**). In addition 636 productive interviews were conducted with younger partners, and 72 with new partners.

Of the 11,580 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.

A responding household is defined as one where at least one eligible person was interviewed. Using this criteria, 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 (see Preface).

ELSA SURVEY CONTENT

The ELSA 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 variable **Dhnow**) 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, but interviewers could override this order if necessary. 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 the majority of its modules. Dependent interviewing was used to check 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, there was flexibility 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.

The questionnaire modules are listed below in the order programmed in the CAPI interview. Most of the variables in each of the modules have been included in the dataset with a module prefix (shown in the brackets below).

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 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).

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 the amount they received.

Income and Assets ("IA")
At the start of the interview couples were asked whether they kept their finances together or separate. If kept together, they were considered to be a single financial unit which required only one respondent for the IA module. 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.

Details of the income that respondents received from a variety of sources over the last 12 months was collected including wages, state pensions, private pensions, 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.

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) 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 (variant on main module)
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.

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 for younger partners.

DATASET INFORMATION

One 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 using an Index file (to be deposited.

The 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 are in in the CAPI interview – see above)
- Other variables not in the questionnaire (including administrative variables, derived variables and variables fed forward from HSE)

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

WEIGHTING

The variable in the dataset to be used for weighting is **w1wgt.** Weights were calculated for core sample members only, as this is the sample of interest. All other individuals that were interviewed (i.e. new and younger partners) have a weight of zero. The data for partners can be used as characteristics of the core members (i.e. supplementary information) but partners should not be analysed as individuals in their own right.

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, eliminate 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. HSE sample design is described in the ELSA technical report (see preface).

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 as having significant levels of non-response that justified calculating a non-response weight (to account for differences between respondents and non-respondents). These were:

- Households that did not contain an age-eligible individual who agreed to be re-interviewed beyond HSE.
- Household level non-response at ELSA wave 1.

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) was appropriate to use in this situation (for further details see Deville and Sarndall (1992)). 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. A key advantage of the 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.

Age-sex distribution

The age-sex distribution of the unweighted and weighted data for core sample members is given in Table 1 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 1 Age-sex distribution unweighted and weighted

Ageband (years) at W1	Unweighted Male %	Unweighted Female %	Unweighted Total %	Weighted Male %	Weighted Female %	Weighted 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
Total	100	100	100	100	100	100

Key estimates

The extent of the effect that the weights have on the data is likely to differ by each data item and each estimate. The effect of weighting on key estimates is illustrated in Table 2 by comparing unweighted and weighted estimates of variables, and by looking at the design factor. Columns 3 and 4 show the unweighted and weighted sample sizes, columns 5 and 6 show the unweighted and weighted estimates, column 7 shows the standard error for the weighted estimate, column 8 shows the 95% confidence interval for the weighted estimate, and the final column shows the design factor for the weighted estimate.

The design factor, 'deft', is the factor by which the standard error of an estimate from a simple random sample has to be multiplied to give the true standard error of the complex design. In other words, it is the ratio of the standard error of the complex sample to that of the simple random sample of the same size. The defts were calculated in STATA. If the deft=1 this implies that the estimates are the same as the estimate would be from a simple random sample. A value of less than 1 implies that the weighting has improved the efficiency of the estimate, and a value of more than 1 that the weighting has introduced some inefficiency to the estimate. The design factors are all relatively close to 1.

Table 2 Key estimates - comparison of unweighted and weighted estimates[1]

Key estimate description	Survey variable names	n unweighted	n weighted	Value unweighted	Value weighted	Standard error	95% CI	DEFT
Income	totinc*	11,135	11,123	348*	347*	4.0	339 - 355	1.01
Net wealth	nettotw*	11,135	11,123	158434*	155663*	3071	149644 - 161683	0.99
Self-assessed general health (good)	hehelf, hegenh	11,220	11,208	91.14	91.26	0.27	90.74 - 91.79	1.00
Non-limiting longstanding illness (1)	heill, helim	11,383	11,383	21.35	20.91	0.38	20.16 - 21.66	1.00
Limiting longstanding illness (2)		11,383	11,383	35.50	35.04	0.45	34.16 - 35.93	1.01
Difficulty walking quarter mile (1)	hefunc	11,216	11,204	70.64	70.66	0.44	69.81 - 71.52	1.01
Difficulty walking quarter mile (2)		11,216	11,204	13.28	13.11	0.32	12.48 - 13.74	1.01
Difficulty walking quarter mile (3)		11,216	11,204	5.98	5.87	0.22	5.44 - 6.31	1.01
Difficulty walking quarter mile (4)		11,216	11,204	10.09	10.35	0.30	9.77 - 10.93	1.01
Diagnosed heart attack	hedia01-hedia10	11,385	11,385	5.99	5.93	0.22	5.49 - 6.36	1.01
Walking speed (in seconds)	mmwlka, mmwlkb	6,193	6,077	5.35*	5.47*	0.37	4.75 - 6.18	1.04
Feeling depressed	psceda	11,041	11,024	17.92	17.95	0.37	17.22 - 18.68	1.01
Self-reported memory (good/v good/excellent)	cfmetm	11,093	11,078	67.61	67.68	0.45	66.80 - 68.57	1.01
Memory test -mean number of animals	cfani	11,034	11,017	19.18*	19.14*	0.06	19.02 - 19.27	1.02
Completing Self-Completion section	scmiss*	11,234	11,221	7.97	8.31	0.27	7.78 - 8.85	1.04
SC - taking a holiday abroad	scptpa5	10,275	10,223	47.00	46.53	0.50	45.56 - 47.51	1.01
SC - not being a member of any organisation	scorg09	9,871	9,810	30.65	30.83	0.47	29.90 - 31.75	1.01

* These variables are not part of the ELSA Wave 1 core dataset. They will be released at a later date as part of a derived variable dataset.

[1] All estimates are percentages unless asterisked; asterisked estimates are means.

Estimating errors in complex sample designs

The ELSA sample used a stratified multi-stage design which was clustered within postal sectors. An effect of using this complex design is that standard errors for survey estimates are generally higher than they would be derived from a simple random sample of the same size. Key estimates from across the range of topics covered in the interview have been identified to illustrate the effect on standard errors.

Table 3 shows the key estimates and the related standard errors and design factors associated with each. Column 3 show the size of the sample on which it is based, column 4 shows the weighted sample size, column 5 shows the mean/proportion estimated, column 6 shows the estimated true standard error, column 7 shows the 95% confidence interval for the estimate, and the final column shows the design factor. The design factors vary by estimate. The estimates most affected are income, net wealth and walking speed.

Table 3 Key estimates - true standard errors[2]

Key estimate description	Survey variable names	n unweighted	n weighted	Value	True standard error	95% CI	DEFT
Income	totinc*	11,135	11,123	347*	5.5	336 - 358	1.40
Net wealth	nettotw*	11,135	11,123	155663*	4407	147013 - 164314	1.42
Self-assessed general health (good)	hehelf, hegenh	11,220	11,208	91.26	0.29	90.69 - 91.83	1.09
Non-limiting longstanding illness (1)	heill, helim	11,383	11,383	20.91	0.42	20.08 - 21.74	1.11
Limiting longstanding illness (2)		11,383	11,383	35.04	0.52	34.01 - 36.08	1.17
Difficulty walking quarter mile (1)	hefunc	11,216	11,204	70.66	0.52	69.65 - 71.68	1.20
Difficulty walking quarter mile (2)		11,216	11,204	13.11	0.35	12.43 - 13.79	1.09
Difficulty walking quarter mile (3)		11,216	11,204	5.87	0.24	5.40 - 6.35	1.09
Difficulty walking quarter mile (4)		11,216	11,204	10.35	0.33	9.70 - 10.99	1.14
Diagnosed heart attack	hedia01-hedia10	11,385	11,385	5.93	0.23	5.48 - 6.37	1.02
Walking speed	mmwlka, mmwlkb	6,193	6,077	5.47*	0.57	4.35 - 6.58	1.62
Feeling depressed	psceda	11,041	11,024	17.95	0.41	17.15 - 18.75	1.12
Self-reported memory (good/v good/excellent)	cfmetm	11,093	11,078	67.68	0.49	66.72 - 68.65	1.10
Memory test -mean number of animals	cfani	11,034	11,017	19.14*	0.08	19.00 - 19.29	1.24
Completing Self-Completion section	scmiss*	11,234	11,221	8.31	0.32	7.68 - 8.95	1.24
SC - taking a holiday abroad	scptpa5	10,275	10,223	46.53	0.60	45.35 - 47.71	1.22
SC - not being a member of any organisation	scorg09	9,871	9,810	30.83	0.54	29.76 - 31.89	1.17

* These variables are not part of the ELSA Wave 1 core dataset. They will be released at a later date as part of a derived variable dataset.

[2] All estimates are percentages unless starred, starred estimates are means.

KEY ANALYSIS VARIABLES

The main group for analysis are the Core Members. Weights have been assigned to Core Members only. Data on partners (younger partners and new partners) can be used as characteristics of the Core Members (i.e. to provide supplementary information), but the 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.

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 and individual interview. If the variable from the individual session was answered (Di) then this code was imputed, otherwise the variable from the household grid (Dh) 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 due to confidential reasons.

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 4 Age of Core Members

Age	n (unweighted)	n (weighted)
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
Total	11392	11392

Table 5 Age of partners (no weighting assigned to partners)

Age	n Younger partners	n New partners	Total
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
Total	636	72	708

SEX

DiSex is the sex given at the respondent's individual questionnaire session. Dhsex is given at the household grid. **IndSex** is derived from these two variables. If the variable from the individual session is answered (Di) then this code is imputed, otherwise the variable from the household grid (Dh) is used in the derivation.

Table 6 Sex of Core Members

	n unweighted	n weighted
Male	5187	5281
Female	6205	6111
Total	**11392**	**11392**

Table 7 Sex of Partners (no weighting assigned to partners)

	Younger partners	New partners	Total
Male	120	29	149
Female	516	43	559
Total	**636**	**72**	**708**

SOCIAL OCCUPATIONAL AND SOCIAL INDUSTRY CODING

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 (long version).
Anssec = coding for job details collected during the HSE interview (long version).

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 qualifications collected at HSE are in variable **aqual**.

The qualifications collected during the ELSA Wave 1 interview are **fqquzm1** to **fqquazm3**.

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).

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.

Table 8 Benefit unit (BU) classification (estimate)

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

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 variable dataset.

CODING AND EDITING

The wave 1 data has been coded and edited. The coding and editing instructions document shows what was done (ELSA Wave 1: Code Book and Edit Instructions – to be deposited at a later date). For all multi-coded variables that were coded, there are two sets of variables. The first are the originals which contain the answer recorded by the interviewer (e.g. spcab1-spcab4). The second set of variables contain the original coding plus the codes assigned to "other answers" (e.g. spcam01 to spcam04). Note the addition of the letter 'm' to the variable names. 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 variable provides 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 are not the same as in the original variables.

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. For these 'partial' interviews, the questions that were not asked will be coded as –8 (don't know) or –9 (refusal).

DATA CLEANING

Editing was done in the Blaise programme in CAPI mode. Errors in the data were indicated for the editor to action as they moved through the questionnaire.

However, most of the editing was carried out by the interviewers in the field. Additional checks in the CAPI editing program which relate to inconsistencies in the data are noted in the coding and editing manual.

DROPPED VARIABLES

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

In order to reduce the potential to identify individuals, the following types of variables have been deleted:

1. Those containing text

2. Those which contained a personal indentifier (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

There are no geographical variables in the archived dataset. In the future a separate geographical dataset with broad identifiers will be made available under secure arrangements.

Variable not included in the questionnaire

Variable name	Variable label	Added information
BEGINNING OF DATASET		
idaindw1	Analytical wave-specific individual serial number	
idahhw1	Analytical wave-specific household 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	
END OF DATASET		
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 hh 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 ffwd	
eligbt	Total Eligible for interview AFTER grid	
eligct	Total Eligible for interview AFTER DHProxy	
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 Inddobyr – IF Didbn was answered then indsex=Didbn, ELSE indsex=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	
chinhh1	Whether or not has a child in household (derived from household grid)	
chouthh	Whether or not has a child outside household (derived from household grid)	
gcinhh1	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
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
aethnicr	HSE ethnic group collapsed into White and Non-white to avoid disclosure