# WAS User Guide Volume 1

## Contents



## Background

Although the need for a reliable source of comprehensive information on the wealth of households and individuals had long been recognised, the beginnings of the present survey date back only to 2000 when the Wealth and Assets Steering Group was set up. Following a pre-feasibility study in 2004, a feasibility study in 2005 and a pilot study in early 2006, the full survey, known in the field as the Household Assets Survey (HAS), commenced in July 2006.

The survey has many objectives given the large number of research and policy questions, which relate to the topic. However, the key objective can be summarised as:

The survey should provide representative data for households and individuals in Great Britain covering:

- The level, distribution nature and type of assets (including savings) and debts of all types.
- Attitudes to financial planning, saving and financial advice.
- Change in the above over time.

In order to achieve the third option, the survey was designed to be carried out over a number of waves, with the original panel of households being revisited at two yearly intervals, and newly formed households being visited if a household split between waves.

The survey is managed and funded by a consortium made up of the ONS and other government departments, namely the Department for Work and Pensions, Department for Business Innovation and Skills, HM Revenue and Customs, Department of Communities and Local Government, the Scottish Government, and the Financial Services Authority.

The report 'Wealth in Great Britain' based on the household and person-level datasets, which were produced from the wave 1 cases, was published in December 2009 and is available at:

http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=15074

## Sampling

### Sample design

The Wealth and Assets Survey (WAS) is a longitudinal survey, which commenced with a first wave of interviews carried out over two years from July 2006 to June 2008. Consenting, responding households from the first wave are being approached for a second wave interview two years on from their initial interview, which covers the period July 2008 to June 2010. A third wave is due to begin in July 2010.

In the first wave, the survey aimed to sample all private households in Great Britain. Consequently, people in communal establishments such as retirement homes, prisons, barracks, halls of residence and hotels, as well as homeless people, were not included in the sample.

In developing the survey, precision targets for change on key estimates were agreed in consultation with funding departments. It was estimated that an overall achieved sample of approximately 32,000 households, spread evenly over the two years of the first wave, was required.

In addition to the above precision targets there was a further target of achieving a two-year sample of 4,500 households above the top wealth decile for Wave one. This was well above the 3,200 households that would be above the top wealth decile for an equal probability sample[1]. Oversampling the wealthiest households both allows for more detailed analysis of this group and will give more precise estimates of the levels of wealth across the whole population.

### Sampling frame

The Wave one WAS sample was drawn from the Postcode Address File (PAF), which is the Royal Mail's database of all addresses in the UK. The sample was restricted to Great Britain and excluded Scotland north of the Caledonian Canal, the Scottish Islands and the Isles of Scilly.

In common with all ONS general population samples, the extract of the PAF used for sampling was restricted to those addresses defined as being small user addresses[2]. In addition, those small user addresses with an organisation name, indicating a small business, were excluded. The ONS copy of the PAF is updated twice a year to ensure that recently built properties are included and demolished or derelict properties are removed quickly.

The ONS PAF sampling frame is held in a hierarchical structure, with addresses grouped into primary sampling units (PSUs) each comprising a postcode sector, or smaller sectors grouped together. The postcode sector is defined as all addresses sharing all but the last two characters of the postcode.

**Sample structure**

The sample for the first wave of WAS had two stages. At the first stage, a stratified sample of PSUs was drawn, followed by a second-stage sample of 26 addresses from each sampled PSU. For each year of the first wave of the survey, 1,200 PSUs were drawn, giving a set sample of 31,200 addresses per year. It was estimated that once ineligible addresses were excluded, along with survey non-contacts and refusals, there would be a sample of approximately 16,000 achieved households per year.

In the first stage of sampling the annual sample of 1,200 PSUs was drawn using a form of systematic random sampling, with probability proportional to size, from an ordered sampling frame of PSUs[3]. The annual samples were drawn separately, creating independent samples.

The sampled PSUs were allocated to months at random. This was achieved using a repeating random permutation which ensured that PSUs allocated to the same quarter and month were evenly spread across the original sample, while still ensuring that each sampled PSU had an equal chance of being allocated to each month. This even spread meant that monthly and, particularly, quarterly samples were balanced with respect to the regional and census-based variables used in the stratification.

The distribution of wealth variables in the population was known to be highly skewed, with relatively few people owning many times the average level of wealth. Consequently, to make the sample more efficient, addresses more likely to contain wealthier households were sampled at a higher rate than the rest. This higher sampling rate for wealthier households also led to a higher proportion of the sample above the top wealth decile, as required by the funding consortium.

Such an emphasis in the sampling requires prior knowledge of the likely wealth characteristics of sampled households. There is a limited amount of information available on the types of households at addresses on the PAF and what is generally available relates to the area around the address, rather than being specific to those at an address. However, HMRC collects data on income and certain components of wealth in order to administer the tax system and the Self-Assessment regime.

By combining this information, HMRC was able to identify those addresses where at least one person was likely to have total financial wealth above a certain threshold. The addresses meeting this criterion and falling within the PSUs sampled for WAS were flagged on the ONS PAF database. This was carried out using anonymised

address records and no tax or personal information was used directly in the sampling process that followed. The addresses meeting the high wealth criterion and falling within the PSUs sampled for WAS were flagged on the ONS PAF database.

In the second stage, from each sampled PSU, 26 addresses were sampled using systematic random sampling from the list of addresses sorted by postcode and street number. This sampling was carried out in such a way that the flagged addresses had two and a half times the probability of being sampled as non-flagged addresses. In the second year of fieldwork, flagged addresses had three times the selection probability as non-flagged addresses

A small proportion of addresses on the PAF have more than one resident household. The PAF provides a multiple occupancy (MO) count field, which has been shown to be a useful indicator of the number of households present in Scottish addresses, but less accurate in England and Wales. Because of this, Scottish addresses with an MO count of three or more were sampled with probability proportional to the MO count.

Generally, the sampled addresses from a PSU would be allocated to one interviewer to complete within a month. However, given the expected long interview times for WAS, it was judged that this would often be too much work for one interviewer. As a result, the sampled addresses from a PSU were split into two batches of thirteen addresses which could be independently allocated to one or two interviewers.

**Field sampling procedures**

Where an interviewer discovered a multi-household address in England and Wales or a Scottish address with an MO count less than two, up to a maximum of three randomly sampled households from the address were included in the sample. For Scottish addresses sampled with an MO count of three or more, a single household was sampled if the MO count equalled the actual number of households present. If the number found differed from the MO count, the number of households sampled was adjusted, but again to a maximum of three. The number of additional households that could be sampled was subject to a maximum of four per PSU.

Some occupied dwellings are not listed on the PAF. This may be because a house has been split into separate flats but only some of these are listed. If the missing dwelling can be uniquely associated with a listed address a divided address procedure can be applied to compensate for the under-coverage. In these cases, the interviewer included the unlisted part in the sample only if the listed address with which it could be associated had been sampled.

Any sampled addresses identified by the interviewer as non-private or non-residential were excluded as ineligible.

## Data Collection

### Field procedures

WAS is one of ONS' largest population surveys in terms of sample size and each month a total sample of 2,600 addresses was assigned to the ONS interviewer panel across Great Britain. Selected addresses were initially sent a letter in advance informing them of their selection, briefly outlining the purpose of the survey and advising them that an ONS interviewer would be making contact to arrange a suitable time to conduct the survey interview.

Interviewers were required to attempt to complete each monthly quota of 13 addresses within five visits to the area and up to 28 working hours, excluding travel time. Best practice procedures whereby interviewers varied their calling times and days in the area were also employed in an attempt to maximise response to the first wave of WAS.

### Computer assisted personal interviewing

Once contact had been established, the WAS questionnaire was administered using Computer Assisted Personal Interviewing (CAPI). There are a number of advantages to this approach over traditional paper interviews which include:

- the ability to use complex sequencing to define specific populations for questions;
- the automatic routing of respondents to those questions relevant to them;
- the inclusion of in-field edit checks which allow seemingly inconsistent responses to be confirmed at the time of the interview;
- the automatic application of alternate question wording according to each respondent's characteristics and prior responses;
- automated questionnaire coding which dispenses with the need for a separate data input stage and reduces the potential for input error; and
- the reduction in cost, timing and security issues around the transport and safe storage of paper forms as interviewers receive and transmit work via a secure modem in their own homes.

### Questionnaire content

The WAS questionnaire, consisting of several thousand questions, was programmed in the Blaise language, developed by Statistics Netherlands. The topics covered are:

- Demography
- Property owned – value and mortgage
- Financial assets – holdings and bills & debt
- Pensions – present value
- Physical assets

- Business assets
- Income – wages, benefits, …
- Attitudes to spending, saving, borrowing and retirement planning

All adults aged 16 years and over (excluding those aged 16-18 currently in full-time education) who consented were interviewed in each responding household.

The questionnaire was divided into two parts. The first part was the household schedule which was completed by one person in the household (the 'household reference person' or HRP, usually the head of household or their spouse) and predominantly collected household level information such as the number, demographics and relationship of individuals to each other, as well as information about equity release, the ownership, value and mortgages on the residence and other household assets.

The second part of the questionnaire was the individual schedule which was administered to each adult in the household and asked questions about economic status, education and employment, numerical ability, business assets, benefits and tax credits, saving attitudes and behaviour, attitudes to debt, major items of expenditure, retirement, attitudes to saving for retirement, pensions, financial assets, non-mortgage debt, investments and other income. Proxy interviews were allowed for those with a good understanding of the person's details.

It should be noted that the values that were collected in WAS are as reported by the respondents. However, interviewers were asked to record whether documentation was consulted after key questions had been asked in the pensions, business assets and income from employment modules.

The questionnaire was designed in consultation with representatives from each of the funding departments with additional advice being provided by analysts with expert knowledge in particular topic areas. Factors considered in designing the questionnaires included the required output, the length and complexity of individual questions, the use of easily understood words and concepts, the sensitivity of topics, the number of topics covered and the overall length of the questionnaire.

The WAS questionnaire was also fully field tested prior to the main enumeration to ensure that:

- it was adequately addressing the data requirements from the survey and that it obtained data in the most effective and efficient way, with respondents being clear as to the meaning of the questions;
- there was minimal respondent concern about the sensitivity or privacy aspects of the information sought and there were acceptable levels of respondent load;
- there was effective respondent/interviewer interaction; and

- the operational aspects of the survey such as the arrangement of topics, sequencing of questions, adequacy and relevance of coding frames were satisfactory.

The average interview length for the first wave of WAS was approximately 80 minutes, with the time varying according to the size of the household and its circumstances. Because the distribution was skewed, the median interview length was about 70 minutes with approximately 25 per cent of interviews lasting 96 minutes or longer and 10 per cent lasting two hours and six minutes or longer.

**WAS interviewers**

All interviewers who worked on WAS were recruited from the ONS trained interviewer pool and had previous experience conducting ONS household surveys.

Prior to commencing WAS fieldwork interviewers were briefed on the background, features and importance of the survey as well as the content of the questionnaire. Interviewer training also promoted an understanding of the survey specific concepts and definitions, and the necessary procedures to ensure a consistent approach to data collection. Interviewers were further required to complete a number of assessed training exercises prior to commencing work in the field.

During an interview, interviewers initially utilised paper-based interviewer instructions for additional information and clarification about the concepts covered by individual questions. However, for most of the survey period, interviewers were able to access this information via Question-by-Question help, which had been programmed into the CAPI instrument. To ensure consistency of approach however, interviewers were instructed to ask the interview questions exactly as worded in the questionnaire.

To further enhance data quality, interviewers encouraged respondents to consult relevant documentation such as bank statements to ensure that the information that was being collected was as accurate as possible. Whether or not documents were consulted was recorded for some questions to assist users of the data in assessing the accuracy of the results.

The proportion of households judged by interviewers to have given 'fairly' or 'very' accurate financial information was consistently high at between 91 per cent and 97 per cent per month for the first wave of WAS.

Where a respondent was not sufficiently fluent in English to undertake the WAS interview other persons in the household could act as an interpreter if this was suggested by the respondent. Otherwise, arrangements were made, where possible, for the interview to be conducted either by an ONS interviewer fluent in the respondent’s language or with assistance from an interpreter service.

## Response rates

Of the 62,823[5] selected households in the Wave one WAS sample, 55,834 were found to be within the scope of the survey and therefore eligible for an interview to be conducted. Of these, 30,595 households were either fully or partially responding and were included as part of the estimates in the Wave one report, see Table 1.

**Table 1 Household response rates for wave one: 2006/08**

| | Number | Percentage |
|---|---|---|
| Sample | 62,800 | |
| Eligible cases | 55,829 | 100 |
| Fully responding households | 28,957 | 52 |
| Partially responding households | 1,638 | 3 |
| Non-contact | 4,135 | 7 |
| Refusal to office | 3,759 | 7 |
| Refusal to interviewer | 15,451 | 28 |
| Other non-response | 1,889 | 3 |
| Wave one household response rate | 30,595 | 55 |

In total, 28,957 (52 per cent) households fully co-operated, 1,638 (3 per cent) households partly cooperated, giving an overall response rate of 55 per cent[6]. Of the 19,210 (34 per cent) households refusing to be interviewed, four fifths of these were refusals to the interviewer, as opposed to refusals to the ONS office. The interviewer could not make contact with 4,135 (7 per cent) households, while a further 1,889 (3 per cent) households were coded to other non-response, which includes things such as illness, physical or mental inability and language difficulties.

**Table 2 Household response rates for wave one: by regions, 2006/08**

| Region | Eligible cases Count | Responding households | Non-contacts | Refusals | Other non-response |
|---|---|---|---|---|---|
| North East | 2,588 | 53 | 7 | 36 | 3 |
| North West | 6,680 | 54 | 7 | 36 | 3 |
| Yorkshire and Humberside | 4,851 | 58 | 7 | 33 | 3 |
| East Midlands | 4,187 | 58 | 6 | 35 | 2 |
| West Midlands | 5,091 | 54 | 7 | 36 | 2 |
| South West | 4,849 | 54 | 8 | 35 | 3 |
| London | 6,672 | 48 | 12 | 34 | 5 |
| South East | 7,823 | 56 | 7 | 34 | 4 |
| East of England | 5,416 | 55 | 5 | 36 | 3 |
| Scotland | 4,867 | 58 | 7 | 32 | 3 |
| Wales | 2,805 | 59 | 6 | 31 | 4 |

Regionally, response rates did not vary a great deal, with the highest response rate of 59 per cent for Wales and all but London reaching at least 53 oer cent, as Table 2 illustrates. The lowest regional response rate was for London with only 48 per cent, this relatively low response rate mirroring the experience of other surveys, including the Family Resources Survey (FRS), the Labour Force Survey (LFS) and the Living Costs and Food Survey (LCF).

## Data Processing

### Editing

An extensive range of computer edits were applied to both the household and individual questionnaires during data entry in the field and to the aggregate data file in the office. These edits checked that:

- logical sequences in the questionnaire had been followed
- all applicable questions had been answered
- specific values lay within valid ranges
- there were no contradictory responses
- that relationships between items were within acceptable limits.

Edits were also designed to identify cases for which values, although not necessarily erroneous, were sufficiently unusual or close to specified limits as to warrant further examination.

Once an interview had taken place, the WAS data were transmitted back to ONS and were aggregated into monthly files. Further editing occurred at this stage and included:

- recoding text entries if an appropriate response category was available
- investigating interviewer notes and utilising the information where applicable
- confirming that overridden edit warnings had been done correctly
- broad data consistency checks.

Once this had occurred validation checks on key input and output variables were conducted to ensure that the data had been correctly converted to the new format. Following this, additional validation, including routing checks, detailed internal consistency checks, range checks and extreme value checks, were undertaken.

Variables on the WAS file were either formed directly from information recorded at individual survey questions or derived from answers to several questions. During validation, data for both types of variables were output as frequency counts and tables containing cross-classifications of selected variables were created for further checking purposes. These processes were intended to identify any problems in the input data which had not previously been identified as well as errors in derivations and other inconsistencies between related items.

In the final stages of validation, comparative checks were undertaken to ensure that WAS estimates conformed to known or expected patterns and were broadly consistent with data from other external data sources, allowing for methodological and other factors which might impact on comparability. This phase of validation is discussed more comprehensively under Data Quality.

**Imputation**

Imputation is an adjustment process which is used to determine and assign replacement values to resolve problems of missing, invalid or inconsistent data. For ease of reference, such values are referred to as 'erroneous'. Imputation of WAS data was achieved by changing as few of the responses as possible to ensure that plausible, internally consistent records were created. Whenever possible, erroneous records were amended during the face-to-face interview with the respondent. However, with a complex household survey it was not usually possible to resolve all such items in this way. Hence, there was a requirement for some form of imputation process to correct for the remaining erroneous records.

The problem of erroneous data in WAS was approached in two stages: firstly a deductive imputation method followed by a statistical method:

Deductive imputation was applied where a missing or inconsistent value could be deduced with certainty. For example, if an individual's total annual income was known to be £35,000 comprising £30,000 from earnings and an unknown amount of interest from savings it can be deduced that the annual interest from savings was £5,000. Deductive imputation was applied wherever possible before applying statistical methods.

There are many differing methods of statistical imputation available. Since WAS collects mainly quantitative data, it was preferable to use a nearest-neighbour imputation method where information from a donor record that had no errors or missing values was used to replace the erroneous values for a recipient record. Nearest-neighbour imputation selects a donor record based on a set of matching variables or predictors. With this method of imputation, the goal is not necessarily to find a donor that matches the recipient exactly on the matching variables. Instead, the goal is to find the donor that is 'closest' to the recipient in terms of the matching variable within the imputation class - i.e. to find the nearest neighbour. This statistical closeness is defined by calculating a distance measure between two observations using a set of matching variables. For example, if the gross earnings variable was missing the variables used to search for the nearest neighbour might be based on the reported net earnings and pay period.

Within ONS, the generalised statistical edit and imputation system CANCEIS (CANadian Census Edit and Imputation System) which was developed by the Canadian Statistical Office and implements a highly efficient nearest-neighbour

imputation methodology was used. CANCEIS performs the simultaneous imputation of categorical and numeric variables. The software has been extensively tested at ONS and analysis confirms that CANCEIS consistently preserves the variance and complex relationships among the variables.

Progress through the WAS questionnaire was governed by a complex routing architecture. Routing variables were indicators of whether the associated subset of questions should contain responses or should be set to 'No Code Required'. For example, it was necessary to establish whether a respondent was an employee before asking their amount of earnings from paid employment. This led to two complications which occur when applying imputation on household surveys:

i. Completeness of routing variables

Where the responses to routing variables contained erroneous values, it was necessary to impute the routing variables before considering the target variables. The imputation of routing variables was often more complex than the imputation of the target variables themselves as there were associated sets of rules defining when a respondent was eligible to enter the variable subset.

ii. Number of variables to be imputed

Often the requirement was to impute a single target variable However, other variables within, or preceding, the subset also needed imputation. As a simple example consider earnings from main employment: in order to impute a value for the target variable (main employment), whether the respondent worked and, if so, whether they worked full time or part time was also required.

Other issues affecting imputation were:

i. Level at which imputation applies

Information collected in WAS will be analysed and presented at both person level and household level. There is, for example, interest in the variation between individuals in pension scheme membership and in holdings of different types of accounts and investments but these data are aggregated to household level as contributors to total household assets. Thus, although the imputation was applied at the person-level for items asked in the individual questionnaire, the methodology also covered the relevant characteristics of the Household or of the Household Reference Person.

ii. Treatment of proxy information

Proxy respondents on WAS used the standard questionnaire, so a proxy response does not, in itself, add a further tier of item non-response. Proxy interviews were only allowed in situations where the person interviewed was likely to have detailed knowledge of the missing person's financial affairs, usually the spouse or partner of the named person. This approach sought to ensure that the number of missing items in proxy interviews was minimised. Proxy response was therefore not a major complication for imputation, as it can be on some surveys.

iii. Missing values in banded variables

The main focus of imputation was on missing values at individual questions. A large number of the variables were collected in a banded form when the response to the continuous variable was unknown or declined. The widths of the bands were necessarily uneven so as to obtain a roughly even spread of the sample across the different categories. Where the response to an individual variable was missing and a banded value was present, the individual value was imputed using the banding as a hard imputation class i.e. by selecting a donor record with a continuous response which was within the relevant band.

iv. Use and release of imputed values

For WAS, the main requirement was to provide imputed values for key survey estimates. These estimates were derived using a large number of collected variables, so imputation was applied across all of the variables that formed the building blocks for key outputs.

The final WAS dataset includes imputed values for variables feeding into derived estimates, but ONS would caution against the use of the contributing variables for a number of reasons, chiefly:

- the greater effect of small sample numbers and skewness of the data at the level of individual variables
- the false impression of accuracy in the imputed values at this level

**Derived variables**

A large number of derived variables was specified by various stakeholders. Lists of these variables, their specifications and the syntax for their production are provided in the documentation accompanying this volume. Some of the concepts and definitions involved in the derived variables for wealth and pensions are provided at the end of this volume.

**Weighting**

Survey data are routinely weighted to compensate for the different probabilities of each household and individual being included in the dataset and to help reduce the random variation in survey estimates. Some of the variation in the inclusion probabilities can be controlled as, for example, WAS has been designed to give those addresses predicted to have higher wealth a higher chance of selection than others. If this were not compensated for in the weighting, estimates of wealth from the collection would be biased upwards. Therefore, the initial step in weighting WAS data was to create, for each case, a design weight equal to the reciprocal of the address selection probabilities.

If it were possible to achieve complete response, the design weight alone would be sufficient to give unbiased estimates from the collected survey data. However, if there were differences in the survey outcomes between sampled households that do

or do not respond to the survey then this will lead to non-response bias. For example, if wealthier households were less likely to take part in the survey, then there was the risk that wealth estimates will be biased downwards.

It was not possible to directly test whether response rates were different for different wealth levels as WAS data were only recorded for the responding households. However, there was a limited amount of information available for both the responding and non-responding households. This can be used in sample-based non-response weighting to compensate for non-response bias[14].

This was achieved by estimating the response rate for different classes and weighting by the reciprocal of the observed response rate for each class. For non-response bias weighting to lead to a bias reduction on a survey estimate the following characteristics need to apply:

- The weighting classes have different response rates
- The survey variable used in the estimate has a different mean in different weighting classes
- The mean of the survey variable was similar for responders and non-responders within each weighting class.

The key available information for both responding and non-responding households on WAS was the Financial ACORN code[7]. This uses Census and survey information to segment the UK population according to financial 'sophistication' into 11 groups and then 49 types. The Financial ACORN code was attached through the postcode of the sampled address.

Using a logistic regression analysis, the Financial ACORN type variable was found to be a significant predictor of household response to WAS. The response rate was calculated, weighted using the design weight, for each of the Financial ACORN types. The reciprocal of this response rate was used as a weight factor to compensate for non-response to the survey. The original design weight was multiplied by this non-response weight factor to produce an initial weight, taking account of both the design and non-response adjustment.

The initial weight derived in this way can be used to produce estimated population counts for different groups defined by age, sex and region. ONS publishes regular population projections for groups based on the Census and information about births, deaths and migration. The estimates from WAS using the initial weight will differ from these population projections because of non-response not yet accounted for and because of random variation. The initial weight was adjusted using a process called calibration to produce a final weight which ensures that the survey estimates of the population match the population projections.

As the fieldwork for WAS was balanced on a monthly basis it was possible to divide the two year fieldwork period into smaller time frames to provide estimates pertaining to those particular time points. Consequently, the sample was conceived as

permitting the following sets of estimates: eight quarterly, two annual and one biennial. This process necessitated the creation of a set of 11 weights. The eight quarterly weights were constructed independently, as described below. The weights from the first four quarters were then divided by four to get an annual weight for Year one. This averaging process was used again to create a year two weight from quarters five through eight. Finally, the two annual weights were averaged to produce a biennial weight.

Each of the quarterly weights was calibrated to fixed population totals of the number of residents living in private households for age group[8] by sex and for region[9] derived from official mid-year population estimates. The weighting was carried out at the household level, so that a single weight was produced at that level, which can be used for both individual and household-level analysis.

Table 3 shows a summary of the weight distribution at each stage of the weighting process[10]. For ease of presentation, only the biennial weight is shown. At the first stage, the range of design weights is due to the oversampling of the predicted high wealth addresses. The ratio of the 95th percentile to the 5th percentile is 3.1.

At the second stage, the design weights were multiplied by the non-response weighting factor to produce the initial weight. The ratio of the 95th percentile to the 5th percentile increased a little to 3.3. The final WAS weight includes the impact of calibration. This tends to increase the range of weights and in particular it can be seen that there were a few outlying weights to the right of the distribution. The ratio of the 95th percentile to the 5th percentile has increased to 4.1.

**Table 3 Summary of weight distribution at each stage in the weighting: 2006/08**

| | Percentile points | | | | | | |
|---|---|---|---|---|---|---|---|
| | min | 0.01 | 0.05 | 0.5 | 0.95 | 0.99 | max |
| Design weight | 137 | 146 | 166 | 434 | 520 | 571 | 716 |
| Initial weight | 237 | 267 | 301 | 802 | 975 | 1,105 | 1444 |
| Final weight | 133 | 239 | 293 | 831 | 1,212 | 1,432 | 2245 |

The weights contribute only part of the impact of outlying values on the variance of the survey estimate. The overall impact can be summarised by the product of the weight and the survey variable contributing to the estimate. If this contribution is considered to be too large, it is possible to reduce the weight to reduce volatility in the estimates while accepting a small bias.

## Confidentiality measures

Much of the data collected in the survey is of a sensitive nature, and a written pledge was made to respondents regarding disclosure of the data. Consequently, the

datasets which are being released to the UK Data Archive have had certain variables removed; including those relating to date of birth and text fields.

## Data Quality

All reasonable attempts have been made to ensure that the data are as accurate as possible. However, there are two potential sources of error which may affect the reliability of estimates and for which no adequate adjustments can be made. These are known as sampling and non-sampling errors and should be kept in mind when interpreting the WAS results.

### Sampling error

Sampling error refers to the difference between the results obtained from the sample population and the results that would be obtained if the entire population were fully enumerated. The estimates may therefore differ from the figures that would have been produced if information had been collected for all households or individuals in Great Britain.

One measure of sampling variability is the standard error which shows the extent to which the estimates should be expected to vary over repeated random sampling. In order to estimate standard errors correctly, the complexity of the survey design needs to be accounted for, as does the calibration of the weight to population totals (see Weighting). WAS has a complex design in that it employs a two-stage, stratified sample of addresses with oversampling of the wealthier addresses at the second stage and implicit stratification in the selection of PSUs.

Although data users should produce standard errors with the outputs of their analysis, with the WAS datasets available at UKDA this is not possible without design information (details of weights, stratification, clustering and calibration). Such information could not be provided with the datasets for statistical disclosure reasons. However, methodologists in ONS are planning to consult with members of the User Group to determine the best way forward in order to facilitate the generation of appropriate standard errors.

Note that some initial estimates of standard errors for key variables are available in the supporting tables to the report referred to above, but imputation effects need to be taken account of, so these should be treated as preliminary: more accurate estimates would be likely to be larger.

### Non-sampling error

Additional inaccuracies, which are not related to sampling variability, may occur for reasons such as errors in response and reporting. Inaccuracies of this kind are

collectively referred to as non-sampling errors and may occur in a sample survey or a census. The main sources of non-sampling error are:

- response errors such as misleading questions, interviewer bias or respondent misreporting
- bias due to non-response as the characteristics of non-responding persons may differ from responding persons
- data input errors or systematic mistakes in processing the data.

Non-sampling errors are difficult to quantify in any collection. However, every effort was made to minimise their impact through careful design and testing of the questionnaire, training of interviewers and extensive editing and quality control procedures at all stages of data processing. The ways in which these potential sources of error were minimised in WAS are discussed below.

Response errors generally arise from deficiencies in questionnaire design and methodology or in interviewing technique as well as through inaccurate reporting by the respondent. Errors may be introduced by misleading or ambiguous questions, inadequate or inconsistent definitions or terminology and by poor overall survey design. In order to minimise the impact of these errors the questionnaire, accompanying documentation and processes were thoroughly tested before being finalised for use in the first wave of WAS.

To improve the comparability of WAS statistics, harmonised concepts and definitions were also used where available. Harmonised questions were designed to provide common wordings and classifications to facilitate the analysis of data from different sources and have been well tested on a variety of collection vehicles.

WAS is a relatively long and complex survey and reporting errors may also have been introduced due to interviewer and/or respondent fatigue. While efforts were made to minimise errors arising from deliberate misreporting by respondents some instances will have inevitably occurred.

Lack of uniformity in interviewing standards can also result in non-sampling error, as can the impression made upon respondents by personal characteristics of individual interviewers such as age, sex, appearance and manner. In ONS, thorough training programmes, the provision of detailed supporting documentation and regular supervision and checks of interviewers' work are used to encourage consistent interviewing practices and maintain a high level of accuracy.

One of the main sources of non-sampling error is non-response, which occurs when people who were selected in the survey cannot or will not provide information or cannot be contacted by interviewers. Non-response can be total or partial and can affect the reliability of results and introduce a bias.

The magnitude of any bias depends upon the level of non-response and the extent of the difference between the characteristics of those people who responded to the

survey and those who did not. It is not possible to accurately quantify the nature and extent of the differences between respondents and non-respondents. However, the level of non-response bias was mitigated through careful survey design and compensation during the weighting process, the latter having been discussed earlier. To further reduce the level and impact of item non-response resulting from missing values for key items in the questionnaire, ONS undertook imputation (see Imputation).

Non-sampling errors may also occur between the initial data collection and final compilation of statistics. These may be due to a failure to detect errors during editing or may be introduced in the course of deriving variables, manipulating data or producing the weights. To minimise the likelihood of these errors occurring a number of quality assurance processes were employed which are outlined elsewhere in this guide.

### External source validation

In the final stages of validating the WAS data, comparative checks were undertaken to ensure that the survey estimates conformed to known or expected patterns and were broadly consistent with data from other external sources. This work was undertaken by ONS and analysts from the funding departments as well as a number of academics who had expertise in the various topics included in WAS. The following guidelines were recommended by ONS when undertaking the external source validation process:

- identify alternate sources of comparable data
- produce frequencies and cross tabulations to compare proportions in the WAS dataset to those from external sources
- if differences were found, assess whether these were significant
- where significant differences were found ensure that reference periods, populations, geography, samples, modes of collection, questions, concepts and derivations were comparable.

Results from these analyses indicated that estimates from the Wealth and Assets Survey were broadly in line with results from other administrative and survey sources. Further work to produce more detailed analyses and comparisons is ongoing and any data quality issues which are identified with WAS variables will be fully documented and made available on the ONS website.

## Concepts and definitions

The ways in which some of the wealth and pensions variables were derived are explained below.

**Wealth estimates**

The wealth estimates in this report are derived by adding up the value of different types of asset owned by households, and subtracting any liabilities. Total wealth with pension wealth is the sum of four components:

- net property wealth,
- physical wealth,
- net financial wealth, and
- private pension wealth.

Total wealth without pension wealth is the sum of the first three of these components.

The components are, in turn, made up of smaller building blocks:

- Net property wealth is the sum of all property values minus the value of all mortgages and amounts owed as a result of equity release
- Physical wealth is the sum of the values of household contents, collectibles and valuables, and vehicles (including personalised number plates).
- Gross financial wealth is the sum of the values of formal and informal financial assets, plus the value of certain assets held in the names of children, plus the value of endowments purchased to repay mortgages.

Some points to note:

- While all other wealth variables in the dataset are imputed, the value of financial assets held in the names of children are not imputed;
- Informal financial assets exclude very small values (less than £250);
- Money held in Trusts, other than Child Trust Funds, is not included.
- Financial liabilities are the sum of current account overdrafts plus amounts owed on credit cards, store cards, mail order, hire purchase and loans plus amounts owed in arrears.
- Net financial wealth is gross financial wealth minus financial liabilities
- Private pension wealth is the sum of the value of current occupational pension wealth, retained rights in occupational pensions, current personal pension wealth, retained rights in personal pensions, AVCs, value of pensions expected from former spouse or partner and value of pensions in payment. Note that, while net property wealth, physical wealth and net financial wealth are calculated simply by adding up the value of assets (minus liabilities, if applicable) for every household in the dataset, private pension wealth is more complicated because modelling is needed to calculate the value of current occupational pension wealth, retained rights in occupational pensions etc for each household. As with all models, the results depend on the assumptions made.

**Private pension wealth measures**

Nine separate components of private pension wealth were calculated based on the WAS survey responses. There were four categories of pension to which respondents were making (or could have made) contributions to at the time of the survey:

- Defined benefit (DB)
- Additional Voluntary Contributions (AVCs) to DB schemes
- Employer-provided defined contribution (DC)
- Personal pensions.

The distinction between employer-provided DC pensions and personal pensions is as reported by the respondent. So, for example, if an individual had a Stakeholder Pension facilitated by their employer and chose to report that as an 'employer-provided/occupational scheme', this is counted as an employer-provided DC pension. Conversely, if an individual reported this simply as a 'Stakeholder Pension', it would be included in personal pensions.

In addition to these four categories of current pension scheme, wealth from five other types of pension was calculated:

- Pensions already in receipt
- Retained rights in DB-type schemes
- Retained rights in DC-type schemes
- Pension funds from which the individual is taking income drawdown
- Pensions expected in future from a former spouse

How the wealth for each of these components was calculated is described in detail in the following sections.

**Current defined benefit occupational pension scheme wealth**

Individuals could report up to two current defined benefit pensions[13]. The wealth in each of these schemes was calculated separately (as described below) and then summed to derive total wealth in current defined benefit (DB) occupational schemes.

Wealth in these schemes was defined as:

$$W_i = \frac{A_R Y_i^p + L_i}{(1+r)^{R-a}}$$

Where:

$A_R$ is the age- and sex-specific annuity factor at normal pension age, R, based on (single life) annuity rates quoted by the Financial Services Authority, assuming average age- and sex-specific life-expectancies (as estimated by the Government Actuary's Department) and a discount rate of 2.5%.

$Y_i^P$ is annual pension income, defined as $Y_i^P = \alpha_i n_i s_i$

$\alpha_i$ is the accrual fraction in the individual's scheme

$n_i$ is the individual's tenure in the scheme

$s_i$ is the individual's gross pay at the time of interview

$L_i$ is the lump sum that the individual expects to receive at retirement

r is the real investment return (assumed to be 2.5 per cent per annum)

R is the normal pension age in the pension scheme

a is the individual's age at interview

Since these are individual, not household, pension wealth measures, and due to the complexity of the calculations and the information that would have been required from respondents, survivor benefits are not modelled. In practice, this would lead to a underreporting of pension wealth for women, since the expected future survivor's benefits that they will receive when they (on average) outlive their husbands will not be measured. To the extent these survivors benefits will be sometime in the future for most women, their omission will have only a small effect on the calculations.

**Definition of wealth from Additional Voluntary Contributions (AVCs)**

Individuals who reported being members of an occupational DB scheme were asked whether they had made any AVCs and, if so, what the value at the time of interview of their AVC fund was. Current AVC wealth is, therefore, simply defined as the fund value reported by the respondent at the time of the interview.

**Definition of current defined contribution occupational pension scheme wealth**

Individuals could report up to two current defined contribution pensions. The wealth in each of these schemes was calculated separately (as described below) and then summed to derive total wealth in current defined contribution (DC) occupational schemes. This procedure was also followed for those who reported that their employer-provided scheme was a hybrid scheme or that they did not know the type of scheme.

Individuals were asked to report the value of their fund at the time of the interview and were encouraged to consult recent statements where available. Current occupational DC pension wealth is, therefore, simply defined as the fund value reported by the respondent at the time of the interview.

**Definition of current personal pension wealth**

Individuals could report up to two current personal pensions; current being defined as schemes to which the individual was (or could have been) contributing at the time of interview. The wealth in each of these schemes was calculated separately (as described below) and then summed to derive total wealth in personal pensions.

Individuals were asked to report the value of their fund at the time of the interview and were encouraged to consult recent statements where available. Current personal pension wealth is, therefore, simply defined as the fund value reported by the respondent at the time of the interview.

**Retained rights in defined benefit occupational pension scheme**

Individuals could report up to three pensions in which rights have been retained. These could be either DB of DC schemes. The wealth in each DB retained scheme was calculated separately (in much the same way as for current DB schemes described above) and then summed to derive total wealth held as retained rights in defined benefit (DB) occupational schemes.

Wealth in these schemes was defined as: $$W_i = \frac{A_R Y_i^p + L_i}{(1+r)^{R-a}}$$

Where:

- $A_R$ is the age and sex-specific annuity factor at retirement age, R (see above)
- $Y_i^P$ is expected annual pension
- $L_i$ is the lump sum that the individual expects to receive at retirement
- r is the real investment return (assumed to be 2.5 per cent a year)
- R is assumed to be 65, or the individual's current age if he/she was already aged over 65
- a is the individual's age at interview

**Retained rights in defined contribution occupational pension scheme**

The wealth in each DC retained scheme was calculated separately (in much the same way as for current DC schemes described above) and then summed to derive total wealth held as retained rights in DC schemes. Specifically, individuals were asked to report the value (at the time of interview) of their retained DC fund.

**Rights retained in schemes from which individuals are drawing down**

Individuals could also report that they were already drawing down assets from a retained pension scheme. In these cases, individuals were asked to report what the remaining fund value for their scheme was at the time of interview. The wealth in each of these schemes was then summed to derive total wealth held in schemes of this type.

**Pensions expected in future from former spouse/partner**

Individuals were asked to report in total how much they expected to receive in the future from private pensions from a former spouse or partner. Respondents were given the choice to report this either as a lump sum wealth figure, or as an expected annual income. Two slightly different approaches were followed, depending on how the respondent answered.

For those who reported a total lump sum value, this figure was taken as the relevant wealth measure and discounted back to the time of the interview. For those who reported an expected future annual income, wealth was calculated in much the same way as for DB schemes described above:

$$W = \frac{A_R Y^p}{(1+r)^{R-a}}$$

Where:

- $A_R$ is the age- and sex-specific annuity factor at retirement age, R (see above)
- $Y_i^P$ is expected annual pension
- r is the real investment return (assumed to be 2.5 per cent a year)
- R is assumed to be 65, or the individual's current age if he/she was already aged over 65
- a is the individual's age at interview

**Definition of wealth from pensions in payment**

In order to calculate the value of the future stream of income provided by pensions from which the individual was already receiving an income, the lump sum which the individual would have needed at the time of interview to buy that future income

stream from a pension provider was calculated. Wealth from pensions in payment was therefore defined as:

$$W = A_a Y^p$$

Where

$A_a$ is the age- and sex-specific annuity factor based on respondent's current age, a

$Y^P$ is reported current annual private pension income

For those age groups for whom no market annuity factor was available (ages 75 and over), we predicted a hypothetical annuity factor based on the information from those ages where annuity prices were available.

References

1. This figure also assumes an even response profile across the wealth distribution.
2. All addresses on the PAF are assigned to a postcode. A large user postcode is one that is assigned to a single address due to the large volume of mail received at that address. The remainder are small user postcodes, each with on average fifteen addresses assigned, although this can vary between 1 and 100. See Royal Mail PAF Digest for more details: http://www.royalmail.com/link/download?catId=4200004&mediaId=5800028
3. The sample at the first stage was not formally stratified as the sampling within different strata was not independent. However, the impact on precision is similar.
4. The estimate will, in general, understate total financial wealth held outside of tax-exempt savings vehicles and it will exclude wealth on which no income accrued, eg shares not paying a dividend. It will also understate wealth to the extent that taxpayers have under-reported their income.
5. This number exceeds the 62,400 addresses issued because it includes incidences of multiple households at the same address.
6. The response rate was calculated prior to rounding and therefore discrepancies may occur between sums of the component items and the total.
7. The Financial ACORN codes were supplied by CACI. See www.caci.co.uk for details.
8. The age groups used were: 0-9, 10-15, 16-24, 25-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-74, 75 and above.
9. The regional variable was the Government Office Regions for England: North East, North West, Yorks. & Humberside, East Midlands, West Midlands, East of England, London, South East, South West, plus Wales and Scotland.
10. The distribution of weights is shown for the responding households only.
11. Where there was an odd number of PSUs within a region the last PSU was added to the previous pair, giving a stratum containing three PSUs.
12. See: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=9220
13. Respondents can report a maximum of two current pension schemes. For example, an occupational DB scheme and a personal pension.
14. However, if response were dependent upon other, unrelated, characteristics, then non-response bias may persist.

# User Guide Annex - Release of total wealth DVs.

## Introduction

The WAS wave 1 household level dataset released in March 2010 contained a number of derived variables. However, excluded from the dataset were the DV's reflecting the total wealth in the four wealth categories and the overall total. Although syntax for generating these figures was provided in Sarah Levy's paper, '*Building estimates of total wealth in the 2006/8 Wealth in Great Britain report – A user guide*', there was a divergence between the results obtained using the syntax on the release file and those presented in the *Wealth in Great Britain* report. Because of this divergence, it was suggested that ONS generate these DVs and provide these variables in a dataset which could be matched onto the exiting dataset, along with documentation providing a comparison of the summary figures for each of these wealth categories presented in the report with the parallel figures obtained from the DV's for the release dataset. This paper is intended to accompany the total wealth DV dataset in order to provide information on the degree of divergence and its causes.

## Wealth DV's compared

Nine variables are being added to the datasets. Please note, the report values are expressed, as in the report, rounded to the nearest 100, whereas the release file figures are expressed to two decimal places. The latter will enable users to confirm the results in their own dataset, but will only be an approximate comparison of the two sets of means to be made.

*Table 1 Total household wealth and its four main components*

| Variable label | Variable name | Report value | Release file value | Release file - valid cases |
|---|---|---|---|---|
| | | mean | mean | |
| Total household wealth including pension wealth | TOTALWPEN | 367,600 | 367,131.78 | 13,901,282 |
| Total household wealth excluding pension wealth | TOTALW | 223,200 | 222,707.10 | 13,901,282 |
| Household physical wealth | HPHYSW | 39,700 | 39,665.74 | 13,901,282 |
| Household private pension wealth | TOTPEN | 141,900 | 141,889.94 | 24,583,701 |
| Net household property wealth | HPROPW | 143,200 | 143,218.54 | 24,583,701 |
| Household property wealth (gross) | HPROPWGR | 182,200 | 182,245.74 | 24,583,701 |
| Net household financial wealth | HFINWNT | 40,000 | 39,431.65 | 24,583,701 |
| Household financial wealth (gross) | HFINW | 43,500 | 43,012.97 | 24,583,701 |
| Household financial liabilities | HFINL | 3,500 | 3,581.32 | 24,583,701 |

## Differences explained

The differences in the summary statistics of the variables derive from differences in a relatively small number of cases in two areas, namely, children's assets and equity release.

*Children's assets*
Children's assets are made up of Child Trust Funds and other children's assets. The report and release figures differ because, in the cleaning process which continued after the production of the datasets for the report, a small number of errors were discovered and corrected.

Although imputed values were generated for missing variables, children's assets were the exception because of the limited information collected for children. Furthermore, in those cases where respondents reported having such assets, but no value was given for them, a zero value was assigned. Thus, since the missing values in those cases reporting possession of children's assets have not been subject to imputation, but have instead been assigned a value of zero, their inclusion in the gross and net financial wealth DV's and the total wealth DVs (including and excluding pension wealth) should be treated with caution. Nevertheless, since they were included in this form in the report calculations, they have also been included in the derived variables to be added to the release file. Should users wish to exclude these variables form the totals, they should subtract the variables ,CaCTV_sum and CASSET_sum from the total financial and overall wealth DV's.

*Equity release*
Differences occur for two reasons:

1. Quality checks revealed two cases with erroneous values which were subsequently corrected
2. Imputation was carried out for formal equity release arrangements, but problems arose in cases of private and other arrangements. Thus, in approximately 60 cases out of a total of 192 reporting equity release arrangements, imputed values are missing. Consequently, the net household property wealth and property liabilities and the total household wealth, including and excluding pension wealth, in both the report figures and the release datasets, do not include private and other arrangement equity release amounts because of the problem of imputing values for these.

## Other issues

Physical wealth: As noted elsewhere, only approximately half of households were questioned about physical wealth in wave 1. Those cases are flagged in the dataset by the variable GCPREAM (coded 1 for included households). Thus, analysis involving physical wealth should only be carried out using those cases which include this asset. Furthermore, the total wealth variable will only include physical wealth for those cases, so analysis of total wealth should only be carried out for either the subset asked about physical wealth or those not asked. In the above table, the 'valid number of cases' indicates the size of the full sample (weighted) as 24583701 households,

whereas the weighted sample of those of whom the physical wealth question was asked numbered 13,901,282.

Note also that although two means of adjusting the aggregate wealth figure by a 'rating up factor' are only applicable to the aggregate figures: they should not be used to adjust individual cases.

*ONS, October 2010*