Office for National Statistics

Wealth and Assets Survey User Guide

Wave two 'report' dataset: June 2012

CONTENTS

Overview

This guide has been produced to assist users in using the Wealth and Assets Survey wave two ‘report’ dataset effectively.

The Wealth and Assets Survey (WAS) is a longitudinal survey that interviewed across Great Britain; England, Wales and Scotland (excluding North of the Caledonian Canal and the Isles of Scilly). Respondents to wave one (July 2006 – June 2008) of the survey were invited to take part in a follow up interview two years later (July 2008 – June 2010) to identify whether their circumstances had changed. Interviews in waves one and two were conducted using Computer Assisted Personal Interviewing (CAPI). Wave one achieved approximately 30,000 household interviews; wave two achieved approximately 20,000 household interviews.

The economic well-being of households is sometimes measured by their income; this ignores the fact that a household's resources can be influenced by their stock of wealth. The increase in home ownership, the move from traditional roles and working patterns, a higher proportion of the population now owning shares and contributing to investment schemes as well as the accumulation of wealth of over the life cycle, particularly through pension participation, have all contributed to the changing composition of wealth. To understand the economic well-being of households it is increasingly necessary to look further than a simple measure of household income.

The WAS aims to address gaps identified in data about the economic well-being of households by gathering information on, among others, level of assets, savings and debt; saving for retirement; how wealth is distributed among households or individuals; and factors that affect financial planning.

The WAS is funded by a consortium of government departments: Department for Business Innovation and Skills; Department for Work and Pensions; HM Revenues and Customs; HM Treasury; Office for National Statistics; and, the Scottish Government. Fieldwork is undertaken by the Office for National Statistics.

To date, wave one cross-sectional data has been made available to users under a special license arrangement. Three wave two ‘report’ datasets has been deposited in the UK Data Archive alongside this user guide. These report datasets are a reduced versions of the full wave one/wave two linked datasets; it only includes variables included in publications to date. It does include all interviews conducted throughout wave one and two. A full wave one/wave two linkable dataset is currently being prepared and will be deposited with the UK Data Archive later this year.

ONS will use feedback on this user guide to improve the information provided for WAS data users. The second version of this user guide will provide information on how to link datasets across waves (not required for the report dataset as this is already linked).

Tips for using Wealth and Assets Survey data

Before conducting analysis with the Wealth and Assets Survey data, analysts should be aware of the the following details that may help you to understand and work with the data.

Content of data files

The interim release data is split into three linked files:

(1) a *household level file* containing all property and physical wealth component variables, as well as all derived variables (DV) used for the calculation of aggregated household wealth.
(2) a *financial data file on person level* consisting of all financial wealth component variables, including all financial wealth DVs and household level wealth DVs
(3) a *pension data file on person level* consisting of all pension wealth component variables, including all pension wealth DVs and household level wealth DVs

Variable naming conventions

- *Wave suffix*
  All three files provide linked data between the first and the second wave, whereby all variables for the first wave contain the suffix 'W1' and those for the second wave the suffix 'W2' in their variable name.
- *Imputation suffix*
  All variables used as components for wealth DVs were imputed. Therefore, for each of these variables an additional variable was created with a suffix '_i' including recorded responses at the interview stage as well as imputed responses.
- *Aggregation suffix*
  To calculate total household wealth all components DVs were aggregated to household level. To enable data users to use aggregated household level DVs on person level, relevant DVs are also provided on the two person level files. To distinguish household level aggregates from person level DVs, the aggregated variables have the suffix '_sum' in the variable name.

Weights

To carry out cross-sectional analysis based on wave one data, the appropriate weight to apply is the variable ‘XS_wgtW1’. For cross-sectional analysis with wave two data, the variable ‘XS_calwgtW2’ should be applied. Both weights can be used on household and on person level, whereby the cross-sectional household level weights are the aggregated mean of the person level weights.

As opposed to cross-sectional analysis, longitudinal analysis can only be carried out on person level. In this instance the variable ‘LONGIT_calwgtW2’ needs to be used for weighting the data.

Interview Outcome codes

All three datasets include responding households only. The variables ‘HOutW1’ and HOutW2’ give an indication of the type of interview outcome of the household:

Fully co-operating

110 Complete interview by required respondent(s) in person
120 Fully co-operating household: one or more interviews completed by proxy
121 HRP economic unit interviewed in person, one or more other interviews by proxy
122 HRP and/ or spouse/ partner interview by proxy
130 Complete interview by proxy

Partially co-operating

212 Non-contact with one or more respondents
213 Refusal by one or more respondents (all contacted)
214 All adults interviewed but one or more interviews was incomplete
222 Non-contact with one or more respondents and some proxy information
223 Refusal by one or more respondents (all contacted) and some proxy information
224 All adults interviewed but one or more interviews was incomplete and some proxy information
211 Full response in person from HRP economic unit – HRP (and spouse/ partner). One or more other interviews missing or incomplete.
212 Full response from HRP economic unit – one or both by proxy. One or more other interviews missing or incomplete.
220 HRP economic unit not complete (one of 2 eligible adults missed; either interview incomplete).
230 No individual interviews with HRP economic unit but household interview completed.

Although the dataset exclusively consists of responding households, not every individual in these households may necessarily be a responder. The variables 'IOut1W1' and 'IOut1W2' indicate the interview outcome of individuals:

1 Full interview (in person or by proxy)
2 Partial interview (in person or by proxy)
3 Ineligible for interview – child aged 0 to 15
4 Ineligible for interview – adult aged 16 to 18 in full-time education
5 Eligible adult – refused to be interviewed
6 Eligible adult – non-contact

Please note:
Although individuals with an outcome code of 5 or 6 did not give an interview they can still be included in the analysis because their values for wealth component variables have been imputed.

Also, analysts should be aware that although children have not been interviewed for this survey, the data on children assets has been recorded against their person number in the household, not against the adult who responded to the relevant questions in this section.

Longitudinal data linkage

The available interim release files are already linked longitudinally. The linkage exercise was carried out in two steps. In the first step person level data was linked by using the variable 'PIDNO', which is a unique person level identifier. This identifier remains the same over the survey life time of a sample unit.

For every individual there are also area, address and household codes that tell us to which household an individual belongs to. However, due to moves of individuals over time these codes may change and can therefore not be used to link households across waves. In order to link household level data, person level records were aggregated to household level (see illustrations below)

PIDNO	AreaW1	AddressW1	HholdW1	PersonW1	YearW1	MonthW1	AreaW2	AddressW2	HholdW2	PersonW2	YearW2	MonthW2
WAS060801234120101	1234	12	1	1	2006	8	1234	12	1	1	2008	8
WAS060801234120102	1234	12	1	2	2006	8	1234	12	1	2	2008	8
WAS060801234120103	1234	12	1	3	2006	8	1551	51	1	1	2008	8
WAS060801234120104	1234	12	1	4	2006	8	1552	51	1	1	2008	8
WAS071202345050101	2345	5	1	1	2007	12	2551	51	1	1	2009	12
WAS071202345050102	2345	5	1	2	2007	12	2551	51	1	2	2009	12
WAS081003456080101	3456	8	1	1	2008	1	3456	8	1	1	2010	1
WAS081003456080102	3456	8	1	2	2008	1	3456	8	1	2	2010	1

Figure 1: Example of linked person level records

Figure 1 provides an overview of what the household level identifiers (AreaWx, AddressWx, HholdWx, YearWx, MonthWx) would look like for scenarios where either some, all or none of the individuals move. The variable PIDNO remains the same for both waves, but the area and address codes (and in some situations the household code) changes for movers.

AreaW1	AddressW1	HholdW1	YearW1	MonthW1	AreaW2	AddressW2	HholdW2	YearW2	MonthW2
1234	12	1	2006	8	1234	12	1	2008	8
1234	12	1	2006	8	1551	51	1	2008	8
1234	12	1	2006	8	1552	51	1	2008	8
2345	5	1	2007	12	2551	51	1	2009	12
3456	8	1	2008	1	3456	8	1	2010	1

Figure 2: Example of linked household level records

Figure 2 illustrates how the linked file on household level would look like for the individuals show in Figure 1. In the example of the first household in wave one, two of the individuals moved to a different address. Therefore three different households in wave two link to the same household in wave one.

Therefore care needs to be taken when conducting cross-sectional analysis on household level. Due to linking records on household level, wave one household records include duplicates where households have split into two more more units in wave two. A dummy variable 'DupCheck' was created to help users exclude duplicates from their analysis. Where 'Dupcheck' does not equal to 1 the case should be excluded from cross-sectional analysis for wave one.

Longitudinal Flags

A number of longitudinal flags have been produced that may help to understand changes in the data when conducting longitudinal analysis with the linked data.

H_Flag1 – Flag for split households

1 = No split
All original sample members (OSMs) are still living in the same household (inc. households with entrants)

2 = Split
OSMs are in different households, some may not be linked if they were non-responders in W2 (inc. households with entrants)

H_Flag2 – Flag for entrant households

1 = EOSM household
New household at W2 (non-responding household at W1)

2 = OSM household

Longitudinal household that was interviewed at W1 and W2 (inc. movers and splits)

H_Flag3 – Flag for households with birth entrants or other secondary sample member (SSM) entrants

1 = OSM birth entrant(s)
Household with birth entrants borth to OSMs
2 = SSM birth entrant(s)
Household with birth entrants who where borth to SSMs
3 = Other SSM entrant(s)
Household with entrants other than births
4 = Any birth and other SSM entrant(s)
Households with entrants as in categories 1 to 3
5 = Household without entrants
Household where no individuals joined

H_Flag4 – Flag for the change of the household reference person (HRP)

1 = No split / same HRP
Household did not split and has still the same HRP as in W1
2 = No split / change of HRP
Household did not split but has now a different HRP compared to W1
3 = Split / same HRP
Split household where the HRP is the same as in W1
4 = Split / change of HRP
Split household with a new HRP compared to W1

H_Flag5 – Flag for mover households

1 = No mover
Productive household that remained at same address in W2
2 = Mover
Productive household that moved to a new address in W2

All of the above flags are included on the household level as well as the person level files. The following person level flags are only included on the person level datasets.

Type – Indicator for linkage status

1 = Linked individual
(regardless interview eligibility and response status)
2 = HAK joiner
Individual joined the household when keep-in-touch exercise was conducted
3 = HAD joiner
Individual joined the household when debtor survey was conducted
4 = W2 joiner
Individual joined the household when W2 interview was conducted
5 = W2 new household
Individual is part of a household that responded at W2 for the first time
6 = Individual no present at W2
This person was part of a responding household in W1 but left the

household at W2 and did not respond
7 = Household not present at W2
Individual was part of a responding household in W1 but the whole household did not respond at W2

P_Flag1 – Flag for wave member status

1 = LOSM
Longitudinal original sample member – individual was a member of a responding household in W1 and W2

2 = EOSM
Entrant original sample member – individual was a member of a responding household in W2, but household did not respond in W1

3 = SSM
Secondary sample member – individual was not a member of any household in W1 but joined a longitudinal household in W2

4 = NOSM
Non-responding original sample member – individual was a member of a responding household in W1 but left the sample at W2

P_Flag2 – Flag for wave entrant status

1 = OSM birth entrant
Child entrant (15years or younger) born to OSM household member

2 = SSM birth entrant
Child entrant (15years or younger) born to SSM household member

3 = Other SSM entrant
Adult entrant (16years or older)

P_Flag3W1/W2 – Flag for wave eligibility status

1 = Eligible adult
Aged 16 years or older and not in full-time education

2 = Ineligible adult
Aged 16 to 18 years in full-time education

3 = Ineligible child
Aged 15 years or younger

P_Flag4 – Flag for HRP status

1 = HRP in W1 & W2
Individual was HRP in both waves

2 = HRP in W1 only
Individual was HRP in first but not in second wave

3 = HRP in W2 only
Individual was HRP in second but not in first wave

4 = Never HRP
Individual was never the HRP (inc. children)

Variable specific notes

The following chapter aims to inform analysts about inconsistencies in the collected data and how certain variables subsequently were set up in order to treat them accordingly in the analysis.

Change of categories and ranges for band estimates between waves

The questionnaire asks respondents categorical questions in numerous instances during the interview. For some questions these categories may have been combined or slightly changed in wave two to improve the questionnaire flow.

Furthermore, the majority of questions that ask the respondent for a particular amount (e.g. current balance in their current account, current value of their pension) are followed by a questions with banded estimates, which prompt the respondent to give an answer if they cannot provide a point estimate. For the majority of banded estimate questions in the pension section, and some questions in other sections the ranges of bands changed between the two waves.

Analysts are advised to check the provided questionnaires and value labels on the datasets before using such variables for analysis in order to avoid confusion.

Personal Equity Plan (PEPs)

At the start of the financial section of the questionnaire respondents are asks what type of assets they have (FInvTy). The wave one questionnaire includes a separate answer option for PEPs, whereas the wave two questionnaire has a combined answer option for Individual Savings Accounts (ISAs) and PEPs.

New loans and hire purchase arrangements

The sections in the questionnaire asking respondents about loans and hire purchase arrangements have been extended slightly between wave one and wave two. In wave one all loans and hire purchases were recorded in the same way regardless as to whether the respondent has already started paying back the debt or whether the repayment will start some time in the future. To improve the way the debts were being captured, new questions were introduced in wave two that made it possible to record existing debt that was being paid back at the time separately from debt that will be repaid in the future.

Shared financial assets and debts

Some of the assets (current accounts and savings accounts) and debts (credit cards and store cards) collected through WAS can be held jointly with a partner. In order to avoid duplication of these assets or debts only one partner of the joint holders will be asked about the amount currently held in their shared current and/or savings account(s) and their current outstanding balance for jointly held credit and/or store cards.

Data users should be aware that in some instances, where for example partners were not interviewed at the same time, respondents gave inconsistent answers in regards to whether or not they had joint assets/debts and how many shared accounts or cards they had.

Since it was not possible to verify which of the two partners in these instances has given the correct answer, the responses remained unchanged. However, the derived variables for shared current and savings accounts split the wealth equally between the partners regardless as to whether the other partner said they had no or less shared accounts than the partner who provided the value of the assets. For the shared credit and store cards the derived variable records the liability for the respondent in the couple who has the responsibility for paying off the debt.

Survey design

Sampling strategy

The Wealth and Assets Survey collects information about private household wealth in Great Britain. The survey uses the small users Postcode Address File (PAF) as the sample frame for residential addresses in Great Britain, that is, England, Wales and Scotland; excluding North of the Caledonian Canal and the Isles of Scilly. The ONS copy of the PAF is updated twice a year to ensure that recently built addresses are included and demolished or derelict properties are removed quickly.

The survey estimates are designed to be representative of the GB population, therefore WAS, like most social surveys uses a ‘probability proportional to size’ or PPS method of sampling cases. This means that the probability of an address being selected is proportional to the number of addresses within a given geographic area, with a higher number of addresses being selected from densely populated areas.

WAS uses a two-stage or ‘clustered’ approach to sampling. Firstly, postcode sectors are randomly selected from the PAF. The postcode sectors are the primary sampling units (PSUs) for the survey. Within each of these postcode sectors, 26 addresses are randomly selected. The selection uses a stratified (ordered) PAF, where addresses are listed by postcode and street number. The list of 26 addresses is split into two quotas of 13 addresses to ease the allocation (to interviewers) and management of fieldwork.

The sampled PSUs were allocated to months at random. This was done 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.

Although the address selection within postcode sectors is random, some addresses have a higher probability of selection than others. This reflects the fact that wealth has a heavily skewed distribution with a relatively small number of addresses holding considerable wealth. This skewed distribution of wealth, alongside the fact that it is often harder to secure response from wealthier households (e.g. gated estates makes respondent contact harder), means that WAS decided to over-sample very wealthy households. For year one of wave one, addresses identified as having high wealth were 2.5 times more likely to be sampled than other addresses. This factor was increased to 3.0 for the second half of wave one in order to further increase the number of achieved interviews with high wealth addresses.

‘High’ wealth addresses are identified after the postcode sectors have been established. A limited amount of information is available about the type of household resident at a particular address on the PAF and what is generally available relates to the area around the address, rather than being specific to 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. Data from HMRC on tax returns at an address level, in conjunction with average FTSE350 dividend yields from the previous calendar year are used to estimate the value of share holdings at a household level. Those addresses estimated to be in the 90th percentile of shareholding value were then oversampled at a rate of 2.5 (wave one) or 3.0 (waves three and four – new cohort sample) relative to other addresses within a given postcode sector.

Sample sizes of each wave

The following table provides a summary of the sample sizes (rounded), both issued and achieved, for each of the first two waves of the Wealth and Assets Survey.

Wave	**Issued addresses**	**Achieved households**	**Achieved adults***
One	62800	30500	53300
Two	32200	20000	34500

*Respondents aged 16 and over.

In developing the survey, precision targets for change on key estimates were agreed in consultation with funding departments. From these, it was estimated that an overall achieved sample of approximately 32,000 households, spread evenly over the two years of wave one was required. In addition to the above precision targets there was a further target to achieve 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. Oversampling the wealthiest households allows for more detailed analysis of this group and gives more precise estimates of the levels of wealth across the whole population.

For wave two, the achieved wave one sample was issued, plus all of the non-contacts. A total of 32,200 addresses were issued for wave two.

Wave structure

The following diagram illustrates the longitudinal design of the Wealth and Assets Survey. Wave one started in July 2006 with fieldwork being spread over a two year period. Wave two, a follow up to wave one was conducted between July 2008 and June 2010. Information on waves three and four is also provided for information, ahead of the release of this data.

All interviews have a two yearly interval between waves, therefore providing estimates of change in relation to the same period of time. For example wave one interviews conducted during July 2006 would be repeated for wave two in July 2008. It is important that this gap remains constant so that estimates of change are comparable wave on wave.

In addition, a new cohort of addresses (shown in blue) were selected to increase the wave three sample size. The same approach is being taken for wave four, with the wave three achieved sample being issued for a follow up interview, and an additional new sample issued to 'top-up' the achieved wave four sample.

	July 06	July 07	July 08	July 09	July 10	July 11	July 12	July 13
Wave 1	Yr1	Yr2						
Wave 2			Yr 1	Yr2				
Wave 3					Yr 1	Yr 2		
Wave 3 new cohort					Yr 1	Yr 2		
Wave 4							Yr 1	Yr 2
Wave 4 new cohort							Yr 1	Yr 2

Mode of data collection

The Wealth and Assets Survey has two interview stages in the longitudinal panel design. The primary interview is where the WAS questionnaire is utilised; this is referred to as the 'mainstage' interview. The second is the Keeping in Touch Exercise (KITE) which is used to maintain respondent's contact details between waves.

Mainstage interview

The mainstage interview is conducted using Computer Assisted Personal Interviewing (CAPI). Face to face interviewing is the preferred choice for the Wealth and Assets Survey due to the complex subject matter of the survey and the need for the interviewer to support the respondent in answering the questions. The interviewer-respondent interaction is much greater on a face to face survey compared with other modes such as paper and telephone. Another reason for face to face interviewing is the need to interview everyone aged 16 and over in the household. This is more challenging with some alternative modes of data collection.

The interview length of the WAS questionnaire also means that CAPI is a good approach. Face to face contact with respondents allows interviewers to identify when respondents are becoming fatigued during the interviews. This allows interviewers to suggest a break from the interview, or perhaps for them to continue the interview at another time in some cases. Identifying respondent fatigue, picking up on body language, is best done when the interview is face to face.

CAPI was also considered the best approach to maximise cooperation with the survey. Response rates to face to face surveys always tend to be higher than telephone, paper and web alternatives.

Keep in Touch Exercise interview

Conversely, the KITE interview aims to collect much less information, and only from one person in each household. The questionnaire is set up to establish whether the household circumstances have changed and if so, how. In the vast majority of cases there isn't change in the household composition so the interview is only about five to ten minutes. The requirements of KITE are much simpler than the mainstage interview, therefore in order to reduce costs and maximise value for money, the interviews are conducted using Computer Assisted Telephone Interviewing (CATI).

Fieldwork procedures

The following provides a summary of interviewer training prior to starting a HAS quota of interviewing; how progress is monitored and performance benchmarked during data collection; and, how contact is maintained with HAS respondents between waves.

Interviewer training

Interviewers working on the Wealth and Assets Survey have received both generic field interviewer and survey specific training.

Generic interviewer training

New interviewers to ONS are placed on a six week training programme – the Interviewer Learning Programme (ILP) - where they are equipped with the skills required for social survey interviewing. The programme coordinates the activities of managers, trainers and interviewers into a structured programme that ensures all interviewers can meet the high standards expected of an ONS interviewer. The training adopts a blended learning approach. Methods used include: classroom training; instructional and activity based workbooks; instructional and activity based e-learning applications; activity based applications that test the interviewer's skills and knowledge base. At the end of the six weeks, interviewers continue to be supported in their personal development. This is done with the assistance of their field manager. They are also assigned a mentor who is an experienced interviewer. New interviewers shadow mentors as well as having a mentor accompany them when they begin working on a survey.

Interviewers also participate in specific training events such as Achieving Cooperation Training (known as ACT) and Achieving Contact Efficiently (ACE). Both of these training packages have been reviewed and rolled out to the entire field force (face to face and telephone interviewing). This is managed through training events and interviewer support group meetings. Quarterly meetings of field managers and their teams are held throughout the year where training issues and refresher training are regularly addressed. Telephone interviewers and ONS help desk operatives receive equivalent training and can very often convert refusals; following the receipt of an advance letter.

Survey specific training

Telephone interviewers

ONS telephone interviewers working on the Wealth and Assets Survey receive an annual briefing on how to administer the Keep in Touch Exercise (KITE) questionnaire. This briefing, delivered by research staff, covers the importance of the KITE interview; and, the importance of collecting contact details and ensuring these are reported correctly. KITE interviewers are trained to try and turn around refusals, should panel respondents express concerns over future involvement in the survey.

Face-to-face interviewers

Interviewers working on the Wealth and Assets Survey undergo training in two stages prior to starting any WAS interviews. Firstly they are provided with a home-study pack to work through which provides detailed information on the purpose and design of the survey as well as the questionnaire content. Following completion of the home study, interviewers complete an ‘electronic learning questionnaire’ or ELQ. This Blaise supported questionnaire is designed to test interviewer’s knowledge of the survey and identify areas where interviewers require further support. The results of the ELQ are submitted to the HQ field team for review ahead of a face to face briefing of up to 12 interviewers. This briefing reviews the content of the home study pack in more detail and offers the opportunity for interviewers to ask questions. The briefing day is tailored to address areas highlighted by results from the ELQ. The briefing is led by one or two field managers, sometimes with support from research and field team HQ staff.

Interviewers do not start WAS work until their field manager is assured that they are fully briefed and ready to undertake the survey.

Respondent contact

Once the sample has been selected, either from the small users Postcode Address File (new cohort), or by maintaining panel address details (old cohort), advance letters are issued to sampled households/respondents. Advance letters are issued approximately ten days prior to the start of the monthly fieldwork period. The advance letters are intended to inform eligible respondents that they have been selected for an interview; provide information on the purpose of the interview; explain the importance of respondent’s participation; and, to provide contact details in case eligible respondents want to find out more.

New cohort households are issued one advance letter addressed ‘Dear resident’ which assumes no prior knowledge or involvement in the survey. For the old cohort, each eligible respondent is sent an advance letter, addressed specifically to them, thanking for their help in the previous interview and inviting them to take part again. The exception to this is the old cohort where the respondent was a proxy interview in the previous wave – these respondents are sent a named advance letter, but the letter assumes no prior knowledge or participation in the survey.

ONS recognises that some sectors of the community can be difficult to contact. These include but are not limited to metropolitan areas, flats, London, ethnic minorities and gated estates. ONS recently reviewed and updated the interviewer guidance on calling patterns designed to maximise contact.

This strategy is known as Achieving Contact Efficiently and is underpinned by a Calling Checklist.

The calling strategy which achieves the highest contact rate at the lowest cost is to vary calling times. Many households will be easily contacted within the first couple of calls, but for those which are not it is important to make sure that successive visits are at different times of the day (including evenings) and on different days of the week.

ONS Methodology conducted a review of interviewer calling patterns and the success of these as the time of day, and day of week varied. This report recommended a set of calling patterns for interviewers to follow in order to maximise the likelihood of establishing contact with respondents[1]. 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 WAS.

Field sampling procedures

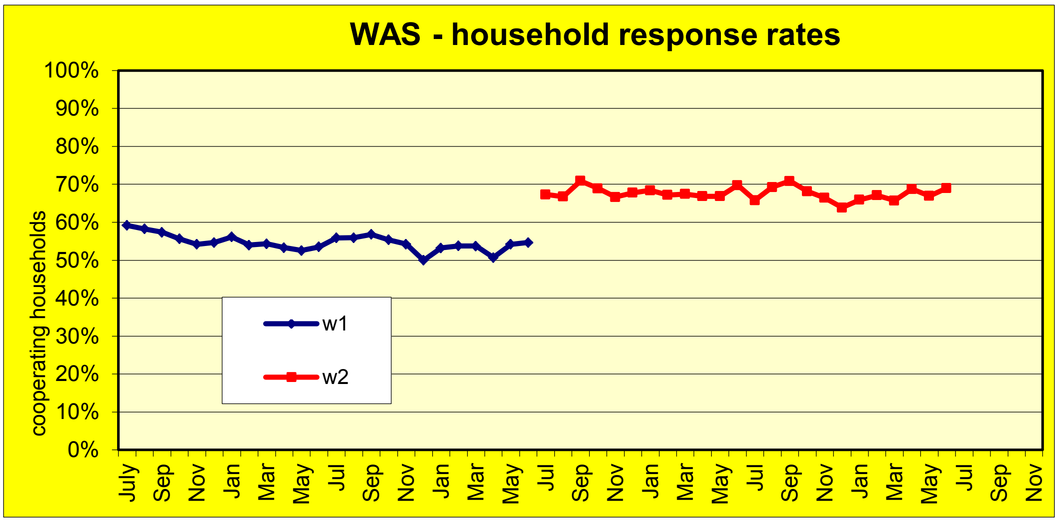
Where an interviewer discovered a multi-household address in England and Wales or a Scottish address with an multi-occupancy (MO) count less than two, up to a maximum of three randomly sampled households from the address were included in the sample. For Scottish addresses sampled with an MO count of 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, only some of which are listed. If the missing dwelling could be uniquely associated with a listed address, a divided address procedure was applied to compensate for the under-coverage. In these cases, the interviewer included the unlisted part in the sample only if the associated listed address had been sampled. Any sampled addresses identified by the interviewer as non-private or non-residential were excluded as ineligible.

---

[1] Hopper, N.: "An analysis of optimal calling pattern by Output Area Classification", ONS Working Paper, Methodology Division, 2008

Response rates

The following graph provides household response for waves one and two, by the monthly field periods.



WAS achieved an average response rate of 55 per cent for wave one, with fieldwork being conducted between July 2006 and June 2008. The achieved sample for wave one was issued for re-interview between July 2008 and June 2010, yielding an improved response of average response rate of 68 per cent.

The following table provides a detailed breakdown of the outcome of cases included in the set sample for both waves one and two.

Outcome	**Wave one**	**Wave two**
Issued cases	**61917**	**32195**
Eligible cases	55835	29584
Co-operating households	**30511**	**20009**
Non-contacts	3889	2717
Refusal to HQ	3805	1268
Refusal to interviewer	15397	4527
Total Refusal	*19202*	*5795*
Other non-response	1770	1063
Response rate	**55%**	**68%**
Non-contact	7%	9%
Refusal to HQ	7%	4%
Refusal to interviewer	28%	15%
Other non-response	3%	4%

Keeping in Touch

WAS is a longitudinal survey that follows all adults interviewed in wave one (original sample members, or OSMs). The survey is biennial, i.e. two years in-between each interview. WAS, like other longitudinal surveys, experiences attrition, which may occur for inevitable reasons such as death, or for reasons that can be minimised such as failure of tracing, failure of contact, or refusal.[2] The longitudinal design of WAS requires following OSMs over time in order to be able to measure changes of wealth. It is evident that tracing and following sample members becomes difficult when circumstances of sample members, in particular their location, change over time.[3] To minimise attrition caused by the loss of sample members due to the failure of tracking, WAS has a number of measures implemented in the survey design to maximise the likelihood of contact being made with the sample member at the next wave.

Firstly, the WAS questionnaire asks respondents at the interview to confirm their address details as well as further contact details such as phone numbers, email address, and contact details of two nominated persons (not resident at the same address) that are authorised to provide ONS with the respondent's new address in case the respondent has moved and cannot be traced. Secondly, a few weeks after the interview all respondents receive a 'Change of Address' card together with the posted incentive (alternatively this will be sent by email), which aims to encourage respondents to inform the ONS if their contact details change. Thirdly, a brief telephone interview is conducted prior to the next wave's interview. This telephone interview is referred to as the 'Keep in Touch Exercise', or KITE. During this interview information about household members as well as their address and contact details are confirmed or updated. It provides the opportunity to identify movers from the household, and their new contact details; as well to identify joiners to the household.

[2] Portanti, M.: "Attrition on Longitudinal Survey – Literature Review", ONS Working Paper, Social Survey Division, November 2009, pg. 2
Plewis, I., 2007. Non-Response in a Birth Cohort Study: The Case of the Millenium Cohort Study. *International Journal of Social Research Methodology*, 10(5), p. 325

[3] Laurie, H., Smith, R. & Scott, L., 1999. Strategies for Reducing Nonresponse in a Longitudinal Panel Survey. Journal of Official Statistics, 15, p. 269

Questionnaire Content

Overview

The Wealth and Assets Survey (WAS) collects data on a wide range of assets and liabilities that private individuals and households in Great Britain have. The primary aim of the survey is to derive overall estimates of wealth and monitor how these change over time. WAS broadly splits wealth into four categories:

1) Financial wealth
2) Pensions wealth
3) Physical wealth
4) Property wealth

The questionnaire is designed to collect relevant information across these four domains of wealth, to provide aggregated measures of wealth, but also to afford significant potential for analysis within these four domains. The questionnaire is therefore both broad and detailed in coverage, with a wide range of stakeholders interested in the data WAS provides.

The wave one questionnaire content was determined by the requirements of the WAS consortium of government departments; namely the Department for Business Innovation and Skills (BIS); Department for Work and Pensions (DWP); HM Revenues and Customs (HMRC); HM Treasury (HMT), and; the Office for National Statistics (ONS); the Department for Communities and Local Government (DCLG) and the Cabinet Office (CO) The primary focus of the questionnaire is to provide for estimates of wealth; however some additional information is collected on non-wealth topics such as socio-demographic characteristics and financial acuity. This allows for aggregate and component analysis of wealth with other factors.

Questionnaire changes

WAS is a longitudinal survey and therefore in order to measure change over time the questionnaire needs to be as stable as possible; so as to reduce discontinuities in the outputs. However, there is scope to make changes to the questionnaire between waves in order to adopt harmonised question standards and/or emerging information requirements.

Changes between waves are made with consortium agreement. Sponsoring departments provide their information requirements and specify any requested changes. These changes are discussed by the WAS Technical Group (TG), with recommendations for questionnaire changes being submitted to the WAS Steering Group (SG). The WAS SG is formed from senior representatives of the consortium departments. Recommended questionnaire changes have previously been subject to cognitive question testing and quantitative piloting. The cognitive question testing has the following objectives:

- ascertain whether the proposed questioning will address the information needs identified by key users and stakeholders, from the respondents' perspective
- establish what respondents understand the questions to mean and the terminology used
- understand how respondents formulate their answers and by so doing ensure that the questions are interpreted as key users and stakeholders intended
- ensure that response options are comprehensive
- ensure that respondents are willing to provide answers
- ensure that respondents are able to provide answers
- ensure that the order in which the questions are asked does not affect the answers given
- address issues relating to the collection of proxy data (if proxy information can be collected)

The quantitative piloting aims to provide a test run of the new questionnaire, and to identify any issues with the questionnaire before the next wave's data collection starts. An interviewer de-brief is held following the pilot to seek feedback on the questionnaire and any areas for improvement. The pilot also provides the opportunity to produce survey metrics such as interview length (broken down by topic area) and indicative response and data linkage consent rates.

Length of questionnaire

The table below shows the mean interview lengths for the first two waves of WAS, and for information the pilot timings for waves three and four.

WAS wave	Mean interview length (mins)*	75th percentile	90th percentile
WAS (wave 1)	88	103	135
WAS (wave 2)	85	104	137

The mean wave one interview length was 88 minutes and has remained relatively consistent for wave two of the survey.

However, the mean interview length is a slightly misleading metric when considering respondent burden. The WAS questionnaire uses extensive routing in order to ensure that respondents are only asked questions that are relevant to them. For example, a one adult household with no or little assets and liabilities would be routed to a relatively small number of questions and therefore have a short interview. Conversely, a two adult household with a lot of different assets and/or liabilities would be routed to a lot of questions and therefore have a much longer interview. This range is reflected in the variance of interview lengths. In wave one, ten per cent of all interviews lasted at least two and a quarter hours. This increased in wave two to over two hours 40 minutes.

Programming and testing

The Wealth and Assets Survey data is collected using Computer Aided Personal Interviewing (CAPI). The software, loaded into interviewer's laptops is called Blaise. All face to face ONS social surveys use Blaise for interviewing as ONS feel that it has the flexibility and technical capability to cope best with the complexity of social research surveys. Blaise's powerful programming language offers numerous features and its data entry program supports a variety of survey processing needs[4].

A number of features of Blaise are particularly advantageous for this survey:

- Blaise CAPI scripts have an in-built hierarchical block structure that effectively makes all questionnaires modular. The ability to handle the associated routing of a modular questionnaire is core to Blaise's architecture. In addition to its hierarchical block structure, Blaise also allows the creation of 'blocks' which can be accessed in parallel, allowing interviewers to switch out of one set of hierarchical blocks to another set. This provides valuable flexibility as it, for instance, allows an interviewer to pause an interview with one household member, initiate an interview with another household member (e.g. a household reference person), and then resume the interview with the original household member at a convenient time in the future.

- Blaise meets the requirement of being able to split the sample geographically or by sample identifiers. Separate questions can be allocated to these different sections of the sample or to randomly selected sub-samples of different sizes.

- Handling complex routing (including loops and repeated events), applying automatic logic and consistency checks in real time during the interview, and using text fills where required, are all core to Blaise's architecture. They are functions that we make extensive use of on the Wealth and Assets Survey.

Blaise allows interviewers to exit and restart interviews at any point which allows interviews to be suspended and resumed.

The Wealth and Assets Survey questionnaire records the length of time spent on different questions during interviews, by placing 'time stamps' at the start and end of different questions. We can use the session log file (called the audit trail in Blaise) to time individual questions. This method affords us the ability to monitor how different questions contribute to the overall length of the questionnaire, which is essential when conducting questionnaire content reviews.

Other features of Blaise which make it excellent for undertaking the Wealth and Assets Survey include:

[4] http://www.blaise.com/capabilities

- the ability for interviewers to back track in instances where later sections of an interview highlight an error made earlier
- flexibility over styles, fonts, font sizes and colours. Blaise allows these to be specified for all text or for individual words/questions etc. This helps ensure the screen seen by the interviewer is as well designed as possible, with effective interviewer prompts. This in turn helps promote interviewer-respondent rapport, thereby contributing to better data quality
- the ability to interact with a ‘question by question’ (QbyQ) help facility. This provides interviewers with real-time access to guidance on specific questions during the interview. This is an electronic programme that operates in conjunction with Blaise

The Wealth and Assets Survey questionnaire is tested extensively prior to being scattered to field interviewers. Currently, staff in the research team independently test the questionnaire; along with staff in ONS Survey Operations team. Questionnaire testing is done every month prior to the questionnaire scatter for the next fieldwork period.

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

The next stage involves checking that the routing of the questionnaire output is correct, using a process referred to as 'base checks'. SPSS programmes are run to emulate the routing performed in Blaise. This process is used to identify where Blaise has incorrectly routed respondents. This can either be corrected for by recoding data, or, where cases haven't been routed as they should have been; imputation requirements are specified. Where errors in routing are discovered, the Blaise questionnaire is corrected to enhance the quality of future data collection. The sooner base checks are performed; the sooner the Blaise questionnaire can be corrected; thus leading to lower levels of data imputation.

Editing and validation processes for the second wave of WAS were similar to those used for wave one: more details are provided in section 10.4 of the wave one report[5]. However, due to the longitudinal component of the survey design, part of the achieved sample size in wave two is linkable to wave one data. Therefore it was important to introduce longitudinal edit checks to the existing editing and validation processes.

The edit and validation checks were run in two stages, whereby first cross-sectional checks were carried out on the second wave to validate or edit outliers. As opposed to checks for the property and physical wealth data,

[5] http://www.ons.gov.uk/ons/rel/was/wealth-in-great-britain/main-results-from-the-wealth-and-assets-survey-2006-2008/report--wealth-in-great-britain-.pdf

checks for financial and pension wealth data were exclusively done on individual level because of the way the data had been collected. The investigation of outliers largely focused on the top and bottom ten per cent of the distribution of each wealth component, although for some variables this proportion was reduced if the number of cases highlighted for investigation was particularly high. When outliers were investigated in the pensions or the financial section, various variables within the same wealth component section or even different sections of the questionnaire were included to establish whether particularly large outliers could be explained by the circumstances of respondents. The majority of investigated cases proved to be genuine and only a small number of cases had to be edited, whereby data was only edited if sufficient information was recorded by interviewers to establish the correct response.

The second stage of checks was conducted after the linkage exercise was completed. At this stage the change of wealth components between the two waves was calculated and subsequently outliers of change were highlighted. To investigate these outliers, the circumstances of relevant respondents in both waves had to be considered to decide whether the value in either wave one or wave two was correct. As with the cross-sectional checks only a small number of corrections were made for each wealth component variable where sufficient information was available.

Imputation

General Methodology

In a way similar to all social surveys, data from the Wealth and Assets Survey contained missing values. Users of WAS data need to distinguish between item non-response, which typically occurs when a respondent does not know or refuses to answer a particular survey question, and unit non-response: missing units where an individual in a responding household refuses to be interviewed or a contact cannot be made. Item and unit non-response can be problematic in that many standard analytical techniques are not designed to account for missing data. More significantly, missing data can lead to substantial bias and inconsistencies in estimates and publication figures. Imputation is a statistical process that serves to counter these problems by replacing missing values with valid, plausible data. To avoid distorting the data through this process inappropriately the method applied must account for the survey question structure and the distributional properties of the observable data that structure yields. It must also take into account the possibility that unrecorded data is not missing completely at random. It is important to note that as the overarching aim of imputation is to improve the utility of the data, the key analytical aims of the survey should also be factored into the design of the imputation process.

Information about discrete assets or liabilities recorded by the Wealth and Assets Survey was collected through a relatively consistent question structure. Typically, an affirmative response to routing questions designed to determine; *do you have asset/liability x?* was followed by a question to specify the value; *what is the amount/income/expenditure of asset/liability x?* In cases where an exact amount was not known, participants were asked to provide a banded estimate from a range of bound values such as £0 to £100, £101 to £500, and so on.

For imputation, the structure of the survey questions gives rise to several important distributional properties in the data. Data from routing questions are categorical. Data from amount/income/expenditure questions can be highly skewed. Furthermore, distributions are often characterised by discrete steps or clustering. This can emerge through constraints imposed by implicit laws or regulations governing the absolute value of an asset or liability, or through respondents able only to provide a banded estimate. The key analytical aim of the survey was to provide longitudinal estimates of change over time as well as cross-sectional/single year estimates. To meet this aim the imputation must account not only for the distributional properties of the data associated discretely with each variable, but also the distributional properties of the rate of growth and/or decay over time.

At this point data users should be aware that the previously released wave one data only included imputation for item non-response. Over the course of processing wave two data the decision was made to also impute missing data from unit non-response to minimise the underestimation of household wealth. In order to make data records comparable on longitudinal level, all longitudinal records that were a unit non-response in wave one were also imputed in the recent imputation exercise for the wave two report. However, this means that when conducting cross-sectional analysis based on wave one data, only part of the data was imputed for unit non-response.

In general, because of the distributional properties of the data elicited by the Wealth and Assets question structure, missing data was best treated using a non-parametric imputation method. To this end, all item non-response and unit non-response was imputed using a Nearest-Neighbour approach (Bankier, Lachance, & Poirier, 1999; Durrent, 2005; Waal, Pannekoek, & Schltus, 2011). In this approach, missing data was replaced with plausible values drawn from other records in the data set referred to as 'donors'. For categorical data and skewed or clustered continuous data, donor-based methods are advantageous in that they use only values actually observed in the data. Significantly, this helps to avoid the distributional assumptions associated with parametric methods such as regression modelling. Importantly, if applied correctly, imputation will estimate the distributional properties of the complete data set accurately (Rubin, 1987; Chen & Shoa, 2000, Durrent, 2005).

Donor Selection

The key to a successful application of Nearest-Neighbour imputation is the selection of a suitable donor. In general, donors were selected based on information specified by other 'auxiliary' variables in the data. Typically, auxiliary variables are employed to constrain donor selection in two ways. Primarily, they serve to identify donors with similar characteristics as the respondent with missing data. Importantly, the auxiliary variables should be related with the data observed in the variable currently being imputed to help estimate accurately the missing value. Auxiliary variables can also be applied to ensure donor selection is tuned towards the key analytical aims of the survey and planned outputs. For all imputed variables in the Wealth and Assets Survey, appropriate auxiliary variables were identified through traditional regression-based modelling supplemented by guidance from experts familiar, not only with a particular subject domain, but also with the analytical program designed to provide outputs that meet customer needs.

Imputation was implemented in CANCEIS, a Nearest-Neighbour imputation tool designed and developed by Statistics Canada (Cancies, 2009). The CANCEIS platform was configured to select a suitable donor for each record needing treatment in two stages. In the first stage, a pool of potential donors was established through two nested processes. The first process divided all records in the survey into 'imputation classes' based on cross-classification of auxiliary variables chosen for this stage. Potential donors could only be selected from the sub-population of records in the same class as the record currently being imputed. The second process served to refine the potential donor pool by ranking all of the records within class. Ranking was determined by calculating the 'distance' between the potential donor and the recipient record based on a second set of auxiliary variables referred to as matching variables. Where appropriate, the calculation included differential weighting to account for cases were some auxiliary variables were more important than others. In general, one of two distance functions were used to calculate the distance between the potential donor and the recipient record, depending on the characteristics of each particular auxiliary variable:-

$x_f =$ the recipient record with *n* auxiliary variables
$x_d =$ the potential donor record with *n* auxiliary variables

$$D_{fd} = \sum_{i=1}^{n} \omega_i D_i$$

ω_i = the weight for the i^{th} variable
D_i = the individual distance for the i^{th} variable

For categorical data with no ordinal relationship between categories:-

$$(2) \quad D_i = \begin{cases} 0 \ \ where \ x_f = x_d \\ 1 \ \ where \ x_f \neq x_d \end{cases}$$

For categorical or continuous data with an ordinal and/or ratio relationship between categories or values:-

$$(2) \quad D_i = \begin{cases} 1 & if \quad |x_f - x_d| \geq y \\ 1 - \left(1 - \frac{|x_f - x_d|}{y}\right) & otherwise \end{cases}$$

y = desired minimum($|x_f - x_d|$) at which point and beyond D_i = 1

A subset of records with the smallest distance values were selected for the final potential donor pool as these were most similar to the record being imputed. For non-categorical data, extreme outliers were excluded from the donor pool to prevent propagation of values likely to have a significant impact on estimates derived from the data. These were identified through expert review and routinely represented values greater the 95th percentile of the observed data’s distribution. Table 1 shows a typical example of an auxiliary variable set. This particular set was used to impute an unknown value for a respondent’s private pension. All Wealth and Asset variables were treated in a similar way.

Table 1. Imputation Classes and Matching Variables used for imputing values for Private Pensions[1]

Imputation Class: Variable	Imputation Class: Classification	Matching Variable: Variable	ω	Matching Variable: Classification
Banded Estimate	1: Less than £2,500	Annual Gross Salary	0.3	Various amounts
	2: £2,500 > £4,999			
	3: £5,000 > £9,999	Employment Status	0.2	1: Employee
	4: £10,000 > £19,999			2: Self-Employed
	5: £20,000 > £49,999			
	6: £50,000 > £99,999	Age Group	0.1	1: 16-24
	7: £100,000 or more			2: 25-44
				3: 45-59 (Female) 45-64 (Male)
tSample	3 month sampling time frame			
				4: 60-74 (Female) 65-74 (Male)
				5: 75+
		Sex	0.1	1: Male
				2: Female
		NS-SEC	0.1	1: Professional
				2: Intermediate
				3: Routine
				4: Never worked
				5: Unclassified
		Employment Sector	0.1	1: Private
				2: Public
				3: Other
		Education	0.1	1: Degree level
				2: Other level
				3: Level unknown
				4: No qualifications

To impute missing values for private pensions donors were selected from an imputation class derived from the cross-classification of observed Banded Estimates and tSample. The Banded Estimate provided an important constraint on donor selection based on observed data. tSample was also significant as research had indicated that private pensions were particularly sensitive to economic trends over a short time frame.

The matching variable set consisted of variables related to the observed data identified through modelling and domain-expert review. Annual Gross Salary and Employment status were given higher weights when calculating the distance between the recipient record and the potential donor as the strength of association was stronger for these variables.

[1] Applied only to cross-sectional data where the respondent was new to the survey and did not have observed data for other waves.

Typically, the final potential donor pool was set to contain between 10 and 20 records. It is important to note that through the first stage of constructing a potential donor pool, the two nested processes used to establish this pool provide an implicit distributional model of the frequency, range, and variance of the set of discrete values observed in the data for records with characteristics similar the record being imputed. In the last stage of the process the final donor was selected at random. Consequently, the probability of a particular category or value being selected was proportional to the number of times that category or value was observed with respect to the total number of observation. This strategy served to support the aim of ensuring that the imputation did not have an unwarranted impact on the distributional properties of data.

Processing Strategy

The Wealth and Assets Survey data were processed in three Sections: Property & Physical, Pensions, and Financial. For all variables, imputation followed a basic processing strategy. First, missing routing was imputed against an appropriate set of auxiliary variables. Following that, where the routing indicated a missing value for the amount associated with a particular asset/liability, the value was imputed against its own set of auxiliary variables. To meet the key analytical aim of the survey; to provide longitudinal estimates of change over time as well as cross-sectional/single year estimates, the detail of the basic processing strategy varied for cross-sectional data belonging to respondents new to the survey, compared to the longitudinal data belonging to respondents who had been in the survey for both Wave1 and wave two.

In general, for respondents with cross sectional data only, processing focused on imputing a discrete category or value drawn from the range and distribution of categories/values observed directly in the data of records reaching the final potential donor pool. For these respondents, donors were selected against a set of auxiliary variables in a way similar to those outlined in Table 1. In contrast, for respondents with longitudinal data, the processing strategy was tuned more towards the observable interdependencies and rates of change in the data between wave one and wave two. To this end, when imputing each variable, respondents with longitudinal data were divided into four imputation groups as outlined in Table 2.

Table 2. Wave one and wave two longitudinal Imputation groups

Data Status		
Wave one	Wave two	Imputation Group
Observed	Observed	Potential donor (O:O)
Missing	Missing	Missing both Waves (M:M)
Missing	Observed	Missing Wave1 (M:O)
Observed	Missing	Missing Wave2 (O:M)

For each variable, potential donors were selected only from records with valid observations in both waves (O:O). When imputing values for respondents with data missing in both waves (M:M), discrete values for both waves were drawn from a single donor. This strategy served to preserve any implicit interdependencies between waves for categorical data and any implicit rates of growth and/or decay for data with continuous characteristics.

To maintain the principle of the longitudinal processing strategy when imputing missing data in records where data was observed in one wave but missing in the other (M:O or O:M) categorical data was treated slightly differently than continuous data. For categorical data, a discrete value observed in one wave was employed to serve as a constraint on donor selection in the same way as an imputation class when imputing the missing value in the other wave. For continuous data, an appropriately banded range was used in a similar way. However, instead of taking a discrete value from the donor, the ratio that described the rate of growth or decay in the donor between waves was transferred to the record to be imputed. The ratio was then used in conjunction with the observed value in one wave to calculate missing value in the other. This strategy is typically referred to as ratio-based roll-back (M:O) or roll-forward (O:M) imputation. Table 3 shows a typical example of a longitudinal auxiliary variable set used to impute a missing value for a respondent's private pension in Wave2 in the presence of observable data in wave one. Comparing Table 3 and Table 1 will help identify the subtle differences between cross-sectional and longitudinal imputation processing strategies.

Table 3. Imputation Classes and Matching Variables used for the longitudinal imputation of Private Pensions in wave two in the presence of observed data in wave two

Imputation Class		Matching Variable		
Variable	Classification	Variable	ω	Classification
Banded Value observed in Wave1	1: Less than £2,500	Annual Gross Salary Wave1 & Wave2	0.3	Various amounts
	2: £2,500 > £4,999			
	3: £5,000 > £9,999			
	4: £10,000 > £19,999	Employment Status Wave1 & Wave2	0.2	1: Employee
	5: £20,000 > £49,999			2: Self-Employed
	6: £50,000 > £99,999			
	7: £100,000 or more	Age Group Wave2	0.1	1: 16-24
	8: No Pension in Wave1			2: 25-44
				3: 45-59 (Female) 45-64 (Male)
Banded Estimate Wave2	1: Less than £2,500			
	2: £2,500 > £4,999			4: 59-74 (Female) 65-74 (Male)
	3: £5,000 > £9,999			
	4: £10,000 > £19,999			5: 75+
	5: £20,000 > £49,999			
	6: £50,000 > £99,999	Sex Wave2	0.1	1: Male
	7: £100,000 or more			2: Female
tSample Wave1 & Wave2	3 month sampling time frame	NS-SEC Wave2		1: Professional
				2: Intermediate
				3: Routine
				4: Never worked
				5: Unclassified
		Employment Sector Wave2	0.1	1: Private
				2: Public
				3: Other
		Education Wave2	0.1	1: Degree level
				2: Other level
				3: Level unknown
				4: No qualifications

To impute missing data in wave two based on rates of growth/decay between waves, donors were selected with reference points in wave one similar to the recipient record based on an imputation class derived from the cross-classification of observed Banded Values in wave one, observed Banded Estimates in wave two, and tSample in both waves. The category 'No Pension in wave one' helped differentiate between new and established pensions.

Topic expert review also indicated that changes in Gross Salary and Employment Status were likely to contribute to the variance in rates of change between waves. Consequently wave one and wave two data for these variables were included in the donor selection process.

Other notable variations in the processing strategy applied to the Wealth and Assets data described to this point were associated typically with samples too small to implement imputation classes based on complex multivariate cross-classification. In such cases, variables that would have been included in donor selection as an imputation class were included instead as a matching variable. Accordingly, the weights applied to the matching variables were adjusted to best suit a preferred priority order. In extreme cases, where for instance, a variable contained less than twenty observations and a small number of missing values, imputation was based on deterministic editing. The range and variance of values imputed this way was guided by topic expert review and was often based on the mean, median, or mode of the observable data.

Quality Assurance and Evaluation

Without exception, the imputed data for all Wealth and Asset variables was examined and tested before being formally accepted. The overarching aim of each evaluation was to ensure that the distributional properties of the observed data had not been distorted inappropriately by the imputation process. Fundamentally, evaluation was based on a comparing the observed data prior to imputation with the fully imputed data. In all cases, any notable departures from the observed data based on statistical measures such as shifts in central tendency or variance and/or the introduction of unexpected changes in the shape of the distribution had to be justified. Justification was based on the identification of sub-populations in the data with proportionally higher non-response rates that would correspond with an appropriate observable change in the properties of the data. This preliminary evaluation was supplemented by a more detailed review of the utility of the data by topic experts familiar, not only with the analytical aims of the survey, but also with expected data trends and characteristics inferred from other reliable external data sources.

Weighting

Overview

The weighting strategy embeds two important principles. The first principle is to maintain the link between the initial selection probability and the ongoing loss to follow up (LTFU) adjustments that remain for the evolving respondent subset over time. This is achieved through developing the longitudinal base-weight from the wave one cross-sectional weight. The second principle is that SSMs in the survey receive a temporary share of the base weight appropriate to their status at any given time point. These principles enable the weighting to refer back to the desired populations as closely as is possible with the current design.

Cross-sectional wave one weight

Survey data are routinely weighted to compensate for the different probabilities of individual households and people included in the analysis data 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 a design weight equal to the reciprocal of the address selection probabilities.

For wave one, the weights were constructed separately for each quarter. This enabled us to combine the quarterly weights as an average over the quarters for a year one or year two annual weight as well as a two year biennial weight. The design weight has the form:

$$w_{1qk}^{d} = w_{1qk}^{d1} w_{1qk}^{d2} \qquad (1)$$

w_{1qk}^{d} is the final Wave 1 design weight in the *qth* quarter for the *kth* respondent and is the product of the initial selection probability for the PSU (w_{1qk}^{d1}) and the selection probability for an address within the PSU (w_{1qk}^{d2}).

$$w_{qi}^{d1} = 1/f_{1qi} = \left(n_q N_i / N \right)^{-1} \qquad (2)$$

Where n_q is the number of selected psus (300 per quarter, 2400 overall), N_i is the number of addresses in the *i*th cluster and *N* is the total number of addresses on the frame.

$$w_{qi}^{d2} = 1/f_{2qih} \qquad (3)$$

Where, for a household at an address in the low (h=1) or high (h=2) wealth stratum, in the ith psu is:

$$f_{2qi1} = \begin{cases} \dfrac{26}{M_i^{lo} + M_i^{hi} c_j} & \text{low wealth stratum} \\ \dfrac{26\, c_j}{M_i^{lo} + M_i^{hi} c_j} & \text{high wealth stratum} \end{cases} \qquad (4)$$

Where M_i^{lo} is the number of addresses in the low wealth stratum, in the *k*th psu, M_i^{hi} is the number of addresses in the high wealth stratum for the psu and c_j is an over-sampling constant within year *j*, where *j* = 2.5 in the first year and *j* = 3 in the second year. The sampling rate between high and 'low' wealth stratum addresses was set at 2.5 in year one. However, this was increased to 3.0 for year two as the 2.5 over-sampling rate did not appear to be succeeding in gaining the requisite number of high wealth households.

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, differences in the survey outcomes between sampled households that do or do not respond to the survey would lead to non-response bias. For example, if wealthier households were less likely to take part in the survey, then there is a 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, a limited amount of information was 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.
This was done by estimating the response rate for different classes and weighting by the reciprocal of the observed response rate for each class. For a bias reduction on a survey estimate, the following information was required:

- 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 was the Financial ACORN code. 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 and 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 non-response weight was calculated as:

$$w_i^{nr} = \frac{1}{\hat{\phi}_i},\ i \in s^r \qquad (5)$$

Where $\hat{\phi}_i$ is the estimated probability of household *i* responding to the survey, derived from a logistic model .

The initial weight derived above can be used to produce estimated population counts for different groups defined by age, sex and region. ONS publishes regular population projections for different 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 that ensures that the survey estimates of the population match the population projections.

As the fieldwork was balanced on a monthly basis it was possible to divide the two-year fieldwork period into smaller time frames to provide estimates for 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 sum of the weights from the first four quarters was 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 by sex and for region derived from official mid-year population estimates. The weighting was carried out at the household level so that a single weight was produced at the household level that could be used for both individual-level and household-level analysis.

For a given quarter *q*, the Wave 1 cross-sectional calibration weight was constructed as:

$$^{cal}w_{1qk}^{xs} = w_{1qk}^{g} w_{1qk}^{d} w_{1k}^{nr} \qquad (6)$$

Where w^g is a calibration factor applied to the kth household, w_{qik}^{d} is the design weight and w_k^{nr} is the non-response weight (with coefficients derived from calculating response probabilities over the combined two year period).

The table shows a summary of the weight distribution at each stage of the weighting process. 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 to 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 to 1. 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 to1.

The weights are 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.

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

Great Britain

	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

Source: Office for National Statistics

Wave two weights

The first step in the wave two weighting process was to develop the attrition models for wave two. The product of the wave one weight and the attrition weights creates the longitudinal base weight. This base weight is the foundation for the development of both the wave two longitudinal and cross-sectional weights.

There are two separate steps that were used to adjust for attrition:

i. Unknown eligibility[6]
ii. Non-response/non-contact

In both cases logistic regression[7] was used to predict the propensity, first for known eligibility status and second for a response. This gives us an estimated

[6] This often, but not exclusively, occurs when interviewers are unable to trace people who have moved address (either whole households or household splits).

propensity for each case denoted by $\hat{\phi}$. Generically, i.e. ignoring subscripting, this is calculated as:

$$\hat{\phi} = \frac{\exp(\hat{\boldsymbol{\beta}}^T \mathbf{x})}{1 + \exp(\hat{\boldsymbol{\beta}}^T \mathbf{x})} \qquad (1)$$

The first model predicted the log-odds of known (to unknown) eligibility using a set of characteristics taken from the wave one survey data and using the wave one weight in the analysis. As both respondents and non-respondents to wave one have data from wave one, a rich set of response predictors is available.

The weights were then constructed as follows: where in (2) *e* represents the probability that outcome eligibility at time two is known, conditional upon the logit model *s* regressing known/unknown eligibility on various wave one individual and household level characteristics.

$$w_{2k}^{e} = \frac{1}{\hat{\phi}_k^{e}},\ k \in s_2^{e} \qquad (2)$$

s_2^e is the sample of k people enumerated within the households with known eligibility status at wave two, where the superscript e refers to the eligible sample, which excludes both cases with unknown eligibility and known ineligibles.

In (3) *r* represents the predicted probability of response from the known outcome eligible sample base again using a logit model with wave one individual and household level characteristics.

$$w_{2k}^{nr} = \frac{1}{\hat{\phi}_k^{r}},\ k \in s_2^{r} \qquad (3)$$

s_2^r is the sample of k individuals within a respondent household at wave two. The longitudinal base weight (w_{2k}^{long}) is product of the wave one weight (w_{1k}^{cal}) and the two loss to follow up (LTFU) adjustment weights for people in a respondent household and is the unknown outcome ineligibility adjusted wave one weight for those classed as known outcome ineligible at wave two.

$$w_{2k}^{long} = \begin{Bmatrix} w_{1k}^{cal} w_{2k}^{e} w_{2k}^{nr}, & k \in s_2^{r}, & \text{longitudinal} \\ w_{1jl}^{cal} w_{2jl}^{e}, & j \in s_2^{ie}, & \text{W2 outcome-ineligible} \end{Bmatrix} \qquad (4)$$

[7] The regression model accounts for the clustered survey design with the nesting of observations (people within households, households within PSUs) using the PSU as the ultimate cluster for the purposes of calculating standard errors of coefficients.

The base weight is trimmed at the 99th percentile of the unadjusted distribution of the weight and scaled to the wave one population values used in calibration[8].

The longitudinal base weight, as constructed above, would be sufficient for longitudinal analysis. However, a final calibration step was applied to take advantage of the calibration options available using the wave one population data. The two longitudinal sub-samples (eligible non-respondents and ineligible outflows) when pooled are, after adjustment for attrition, representative of the wave one population; so it is possible to calibrate the longitudinal base-weight to the wave one population totals. This procedure should have the advantage of further correcting for any attrition not already accounted for by adjusting to the wave one calibration control groups.
The calibration weights are calculated to sum to a set of known calibration totals t, minimising the distance between the calibrated weight (${}^{cal}w_{2l}^{long}$) and the pre-calibration weight (${}^{pre-cal}w_{2l}^{long}$). If the membership of the calibration groups is represented by a vector of auxiliary values x_l, then the problem can be represented as:

$$g_k = \min\left\langle \sum_l dist({}^{pre-cal}w_{2l}^{long}, {}^{cal}w_{2l}^{long}) \right\rangle \text{ such that } \sum_l x_{2l}\, {}^{cal}w_{2l}^{long} = t \quad (5)$$

which has a solution in the form of:

$$ {}^{cal}w_{2l}^{long} = g_l \, {}^{pre-cal}w_{2l}^{long} \quad (6)$$

The final longitudinal calibration weight is the product of the *g* weight and the initial longitudinal base weight, where the g-weight is defined as the solution to (5). The g-weight helps to rebalance the sample towards the population values of the variables included in the calibration model.

Basic descriptive statistics for the longitudinal calibration and base weights are provided in Table 2, along with their wave one cross-sectional counterpart weight descriptives,[9] for comparison. It is perhaps worth noting the comparatively high coefficient of variation seen at wave one, which was largely a result of oversampling wealthy households. Both weights are viable candidates for use with the data but calibration has added some extra variability which reflects some further bias adjustment additional to that undertaken in the attrition modelling.

[8] In fact, this population total is an average of the eight quarters used in the quarterly calibration of the Wave 1 weights.
[9] The wave 1 weight has been rescaled slightly because further data cleaning after production of the Wave 1 weights resulted in a small number of cases being dropped from the dataset. In order to provide a consistent reference point a scaling factor of 1.002 was used to equilibrate the weights.

Table 2: Summary of descriptive statistics comparing the longitudinal and wave one cross-sectional weights

Weight	n	Mean	Standard deviation	Coefficient variation	Minimum	Maximum
Wave 1	71,268	816	284	35	133	2250
Longitudinal base	43,338	1,341	602	45	214	3439
Longitudinal calibration	43,338	1,341	643	48	203	3900

A preliminary, albeit small scale, comparison was made from estimates using the longitudinal base weight and the final calibration weight on the longitudinal sample.[10] These longitudinal estimates were compared to estimates produced using the wave one weight on the full wave one cross-sectional sample. Both of the longitudinal weights produced estimates that compared well to the original wave one estimates; but, in general, the calibrated weight performed slightly better than the base weight.

It is apparent that calibration has made some minor adjustments to the weight for both the set of respondents and the set containing the ineligible population outflow. The sum of the respondents has been slightly downwardly adjusted and the weight for group of ineligible people has been correspondingly slightly upwardly adjusted (Table 3). For both groups, calibration has increased the range of the distribution. This suggests that calibration has adjusted for non-response differentials not otherwise adjusted for in the previous LTFU modelling adjustments.

Table 3: Summary of descriptive statistics comparing respondent and outflow sub-samples of the longitudinal weights

Group	n	Sum	Mean	Standard deviation	Coefficient variation	Minimum	Maximum
Respondent	41,331						
Base weight		56,048,842	1356	604	45	214	3439
Calibrated weight		56,028,929	1356	646	48	203	3900
Ineligible	2,007						
Base weight		2,086,996	1040	455	44	229	3439
Calibrated weight		2,106,909	1050	494	47	218	3900

Note: respondent here refers to all longitudinal OSM people enumerated in a respondent household whether eligible for an interview or not.

The cross-sectional weight was constructed using information from the following groups of people:

- Longitudinal OSMs - using the longitudinal base weight.
- SSMs – based on a weight share derived from the longitudinal base weight.

[10] Here including longitudinal respondents and population outflow ineligibles.

- Wave two entrant OSMs - based on their original design weight.

The aim was to create a single weight to cover both households and individuals. In order to achieve this aim an ‘integrative calibration’ (Lemaître & Dufour, 1987) approach was used simultaneously to create both household and person level wave two (pseudo) cross-sectional weights. This results in all people in the household having the same weight, which is also the household weight. The construction was based on counting the numbers of all wave two enumerated cases (i.e. those people eligible and ineligible for an interview) to calibrate to the population totals. The population totals were based on interpolations of ONS’ mid-year estimates taken from the midpoint of the wave two fieldwork period.

The first challenge for the wave two pseudo cross-sectional weight was to assign a weight to people entering the sample as SSMs. It is common to use a weight share method to approximate these probabilities (e.g. Huang 1984, Ernst 1989, Kalton & Brick 1995), rather than attempting to work out selection probabilities directly. A standard approach is to assign weight shares based on wave one household members to people in target wave two households. A variety of weight share algorithms exist (see e.g. Rendtel & Harms 2009). The WAS weight share was constructed following Kalton & Brick (1995), where the weight at time t_T for household w_i can be defined as the product of the initial weight and a constant:

$$w_i = \sum_j \sum_k \alpha_{ijk} w'_{ijk} \tag{7}$$

The ith household weight w_i at time t_T is the initial weight[11] w'_{ijk} summed over the k individuals in households j at time 1 contributing to membership of household i at time t_T. The constant (α) is defined in terms of the number of people in household i at time t_T who were in household (j) in the population at time t_1. As long as the sum of alpha within households equals unity, estimation will be unbiased (Kalton & Brick 1995).

$$\alpha_{ijk} = \begin{cases} 1/N_h & \text{if individual } k \text{ lives in household } i \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Finally, the weight w_i is assigned to all k household members of household i.

Ideally, in this scheme a population entrant at wave two is assigned a zero contribution to α and a zero initial weight (w'_{ijk}). However, a sample entrant who was in the population at wave one but only in the sample at wave two contributes to α but has a zero initial weight. Consequently, sample entrants in the population do not increase the sum of the weights; whereas population entrants do increase the sum of the population weights.

[11] This may be the Horvitz Thompson estimator or an adjustment of this, e.g. for non-response and/or through calibration. For WAS it is the longitudinal base weight.

This is the fair share method of Huang (1984) and also the weight share method of Ernst (1989). Lavallée (1995, 2007) also shows how this approach is a special case of his generalised weight share method. An alternative method, known as multiplicity, or equal household share, is possible but can be difficult to implement in practice because it requires knowing whether or not two (or more) sample entrants to a household *i* from the t_1 population came from the same household or not.

A key challenge for the weight share method is being able to distinguish between those SSMs who are new population entrants and those who were originally in the population but not originally in the sample. Unfortunately it is not possible to make this distinction with WAS data and consequently, excepting births, we treated all SSM entrants as if they were in the population at the time the sample was drawn. Births to OSM mothers were allocated their mother's weight.

Different surveys use different approaches to weight sharing. For WAS, weights were summed over and shared across all people in the household at times t and t+1. This is not universal practice. Some surveys restrict the sharing to adults or use other criteria, see Schonlau & Kroh (2010), who detail the methods used by key international longitudinal surveys. As WAS is concerned with enabling estimation for all population members and weighting is based on calibrated population totals, it seemed desirable and appropriate to ensure sharing was across all cases enumerated within households.
The weight share allows the longitudinal OSM and SSM sample members to be treated together as a single sample but the construction of the cross sectional weight requires an amalgamation with the group of entrant OSMs whose first interview was in wave two.

$$ {}_{pre-cal}w_{2i}^{xs} = \begin{Bmatrix} (1-\theta)\sum_j \sum_k \alpha_{ijk} w_{2ijk}^{long} & \text{longitudinal} \\ \theta \sum_i w_{1i}^{d} w_{1i}^{nr} & \text{entrant} \end{Bmatrix} \qquad (9) $$

For the longitudinal sample, the pre-calibration wave two cross-sectional weight is constructed through bringing forward the longitudinal weight for OSMs and sharing out between OSMs and SSMs in the wave two household. The entrant sample members have their design weight constructed as the inverse of the product of the selection probability and the non-response adjustment. This produces a household level weight which is constant for each individual in the household.
The final stage is to calibrate the cross-sectional weight to population totals at time t_T, using integrative calibration. Descriptive statistics of the resulting weight are given in Table 4. It is apparent that the bias adjustment is leading to a substantial increase in the coefficient of variation which will decrease the precision of estimates.

Table 4: Summary of descriptive statistics of the wave two cross-sectional weight

Weight	n	Sum	Mean	Standard deviation	Coefficient variation	Minimum	Maximum
Wave 2	46,347	59,191,698	1277	731	57	106	3700

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 develop and test 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.

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)
- net 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, less the value of non-mortgage debt.

Some points to note:

- 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;
- 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, Additional Voluntary Contributions (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. 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 per cent.

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

α_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

Contact details

For further information, or to provide feedback on the Wealth and Assets Survey documentation, please contact Tom Howe:

Phone: 01633 456330

Email: Thomas.howe@ons.gov.uk