



Office for  
National Statistics

# Wealth and Assets Survey User Guide Wave 5

## CONTENTS

<b>OVERVIEW .....</b>	<b>3</b>
<b>USING WEALTH AND ASSETS SURVEY DATA .....</b>	<b>4</b>
CONTENT OF DATA FILES .....	4
VARIABLE NAMING CONVENTIONS .....	4
WEIGHTS.....	4
INTERVIEW OUTCOME CODES .....	5
LONGITUDINAL DATA LINKAGE.....	6
LONGITUDINAL FLAGS .....	6
<b>VARIABLE SPECIFIC NOTES .....</b>	<b>7</b>
<b>SURVEY DESIGN .....</b>	<b>9</b>
SAMPLING STRATEGY .....	9
SAMPLE SIZES OF EACH WAVE .....	9
WAVE STRUCTURE.....	11
MODE OF DATA COLLECTION.....	11
<i>Mainstage interview.....</i>	11
<i>Keep in Touch Exercise interview.....</i>	12
<b>FIELDWORK PROCEDURES .....</b>	<b>13</b>
INTERVIEWER TRAINING .....	13
<i>Generic interviewer training.....</i>	13
<i>Survey specific training .....</i>	13
RESPONDENT CONTACT .....	14
FIELD SAMPLING PROCEDURES .....	15
<b>RESPONSE RATES .....</b>	<b>17</b>
KEEPING IN TOUCH .....	18
<b>QUESTIONNAIRE CONTENT .....</b>	<b>20</b>
QUESTIONNAIRE CHANGES .....	20
LENGTH OF QUESTIONNAIRE .....	21
PROGRAMMING AND TESTING.....	21
<b>EDITING .....</b>	<b>24</b>
<b>IMPUTATION.....</b>	<b>26</b>
GENERAL METHODOLOGY.....	26
DONOR SELECTION .....	27
PROCESSING STRATEGY .....	28
QUALITY ASSURANCE AND EVALUATION .....	30
<b>WEIGHTING.....</b>	<b>32</b>
<b>DATA QUALITY .....</b>	<b>32</b>
SAMPLING ERROR.....	40
NON-SAMPLING ERROR .....	40
EXTERNAL SOURCE VALIDATION.....	40
<b>WEALTH ESTIMATES.....</b>	<b>41</b>
<b>CONTACT DETAILS.....</b>	<b>47</b>

## **Overview**

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 wave two follow up interview two years later (July 2008 – June 2010). Respondents to wave 2 were then invited to take part in a wave three follow up interview another two years later (July 2010 – June 2012). In addition to these follow up interviews, a new random sample of addresses was also added at wave 3, wave 4 and wave 5. Interviews in all waves were conducted using Computer Assisted Personal Interviewing (CAPI). Wave one achieved approximately 30,000 household interviews, wave two achieved approximately 20,000 household interviews, wave three achieved approximately 21,000 household interviews, wave four achieved approximately 20,000 household interviews and wave five achieved approximately 18,000 household interviews.

The economic well-being of households is often measured by its income, and yet a household's resources are composed of its stock of wealth as well as its flow of income. To fully understand the economic well-being of households it is necessary to look beyond measures of household income.

The WAS addresses this gap in the economic well-being of households data by gathering information on the ownership of assets (financial assets, physical assets and property), pensions, savings and debt.

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

# Using Wealth and Assets Survey data

## ***Content of data files***

The data are split into two 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 and income.
- (2) A *person level file* consisting of all person level financial wealth, pension wealth and income component variables and DVs.

## ***Variable naming conventions***

- *Wave suffix*  
Variables in both datasets are given the suffix 'W5' to indicate that they contain values collected in wave 5 and respectively a 'W4' suffix to indicate values collected in wave 4. There are a few variables with the suffix 'W3' 'W2' or 'W1' and these contain values from wave 2 or wave 1. Most of the W2 and W1 variables are present to allow the datasets to be matched to those from previous waves of the survey, e.g. HH SERIAL W2.
- *Imputation suffix*  
All variables used as components for wealth DVs were subject to imputation. Variables that have had missing data imputed appear in the datasets in two versions. The version that contains only the values observed at interview will end with the suffix W5 as described above, e.g. FSInValW5. The version that contains both observed and imputed values will end with the suffix '\_i', e.g. FSInValW5\_i.
- *Aggregation suffix*  
To calculate total household wealth all component 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 person level file, e.g. DVFBondVW5\_aggr at household level and DVFBondVW5 at person level.

## ***Weights***

To carry out cross-sectional analysis based on the individual wave data, the following table has the appropriate variable weight to apply for cross-sectional analysis.

Wave	Cross-sectional Calibration Weight
1	XS_wgtW1
2	XS_calwgtW2
3	W3xswgt

4	W4xperswgt
5	W5xperswgt

As opposed to cross-sectional analysis, longitudinal analysis can only be carried out on person level. The following table has the longitudinal variable weight to apply for longitudinal analysis.

Wave	Longitudinal Calibration Weight
W1W2 Longitudinal Weight	Longit_calwgtW2
W1W3 longitudinal weight	w1w3wgt
W2W3 longitudinal weight	w2w3wgt
W1W4 Longitudinal Weight	W1W4_longwgt
W3W4 longitudinal weight	W3W4_longwgt
W4-W5 longitudinal weight	W4W5_longwgt
W1-W5 longitudinal weight	w1w5wgt

### ***Interview Outcome codes***

The datasets include responding households only. The variable HOutW5 gives 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

##### Partial- information given by required respondent(s) only

- 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

##### Partial- some information given by proxy

- 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 incomplete and some proxy information

##### Partially co-operating . Complete household questionnaire , one or more individual interviews missing or incomplete

- 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 every household responds. The variable **IOut1W5** indicates 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***

All final data files are linked files which have a single variable to use for linking cases between waves. For household linking, there are separate variables for each wave; each case may have up to three variables with a valid code. For person level there is one variable used for matching a case in any wave.

- Always used the linked file as a base when matching variables across waves.
- Use HHSerialW3, HHSerialW4 and/or HHSerialW5 for household linking.
- Use PIDNo for person level linking, this remains the same over the survey life time of a sample unit.
- When you GET the files, only KEEP the variables you need to add to the file (including the one needed to match cases). This makes it easier when matching.

To add W4 variables to the W5 person file, keeping only W5 cases:

- Sort W1W2W3W4W5 person level linked files by PIDNo (W1W2W3W4W5\_UKDA.sav)
- Sort W4 Person file by PIDNo
- Sort W5 Person file by PIDNo
- Match W4and W5 files with the linked file being used as a look up TABLE; use PIDNo to MATCH.
- This will add W4 variables to W4 cases and W5 variables to W5 cases
- Select the required cases e.g. for W5 cases (including linked W4 cases) use HHSerialW5>0.

To add W5variables to the W4 household file, keeping all cases:

- Sort the linked file by HHSerialW4(W1W2W3W4\_UKDA.sav)
- Sort W4 household file by HHSerialW4

- Match files using HHSerialW4
- Sort the new file by HHSerialW5
- Sort W4 household file by HHSerialW5
- Match files using HHSerialW5
- This will produce a linked W4W5 file with W4 and W5 variables.

#### Linking within a wave

To add household variables to the W5 person file:

- Sort both files by HHSerialW5 and use this variable to MATCH
- Use the household file as a look up TABLE. This will add the household variables to each person in the household.

Note: Person level variables cannot be added to the household file unless they are aggregated first.

#### Linking End User Licence (EUL) data

For W5 and W4, the Government Office Region, date and month of the interview will be added to the EUL, to assist with the analysis of the dataset.

As the EUL datasets are anonymised the variables HHSerial and PIDNo have also been anonymised. To link household files the variable **Case** replaces HHSerial, therefore Case will need to be used when linking cases. To link person files the variable **Person** replaces PIDNo, therefore variables Case and Person will need to be used when linking cases.

Linking using EUL datasets:

- Use CaseW1, CaseW2, CaseW3, CaseW4 and/or W5 for household linking.
- Use PersonW1, PersonW2, PersonW3 PersonW4 and/or PersonW5 and CaseW1, CaseW2, CaseW3 CaseW4 and/or CaseW5 for person level linking.

To add W4 variables to the W5 person file, keeping only W5 cases:

- Sort W4 by CaseW4 and PersonW4.
- Sort W5 by CaseW4 and PersonW4.
- Match W4 and W5; using PersonW4 and CaseW4 to MATCH.
- This will add linkable W4 cases to the W5 file and add W4 variables to W4 cases.
- Select the required cases e.g. for W5 cases (including linked W4 cases) use CaseW5 > 0.

To add W5 variables to the W4 household file, keeping all cases:

- Sort W4 household file by CaseW4
- Sort W5 household file by CaseW4
- Match files using CaseW4
- This will produce a linked W4W5 file with W4 and W5 variables.

#### ***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.

The following person level flags are only included on the person level datasets.

### **Type – Indicator for linkage status**

- 1 = W3 – W5 Linked cases  
(regardless interview eligibility and response status).
- 2 = W4 – W5 Linked cases  
(regardless interview eligibility and response status).
- 3 = W5 HAK Joiner  
Individual joined the household when keep-in-touch exercise was conducted.
- 4 = W5 HAD Joiner  
Individual joined the household when debtor survey was conducted.
- 5 = W5 New respondents  
Individual joined the household when W5 interview was conducted.
- 6 = W5 New Household  
Individual is part of a household that responded at W5 for the first time.
- 7 = Individual no present at W5  
This person was part of a responding household in W3 but left the household at W5 and did not respond.
- 8 = Household not present at W5  
Individual was part of a responding household in W4 but the whole household did not respond at W5.

### **P Flag1W5 –Flag for wave member status**

- 1 = LOSM  
Longitudinal original sample member – individual was a member of a responding household in W5 and W4.
- 2 = EOSM  
Entrant original sample member – individual was a member of a responding household in W5, but household did not respond in W4.
- 3 = SSM  
Secondary sample member – individual was not a member of any household in W4 but joined a longitudinal household in W5.
- 4 = NSM  
Non-responding sample member – individual was a member of a responding household in W3 but left the sample at W4.

### **P Flag2W5 – Flag for wave entrant status**

- 1 = OSM birth entrant  
Child entrant (15 years or younger) born to OSM household member.
- 2 = SSM birth entrant  
Child entrant (15 years or younger) born to SSM household member.
- 3 = Other SSM entrant  
Adult entrant (16 years or older)

### **P Flag3W5 – 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 Flag4W5 – Flag for HRP status**

- 1 = HRP in W4 & W5
  - Individual was HRP in both waves
- 2 = HRP in W4, not W5
  - Individual was HRP in fourth but not in fifth wave
- 3 = HRP in W5, not W4
  - Individual was HRP in fifth but not in fourth wave
- 4 = Never HRP
  - Individual was never the HRP (inc. children)

### **Variable specific notes**

Refer to the wave 5 paper questionnaire for notes on specific variables.

### **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, and the fact that it is often harder to secure response from wealthier households is the reason for the over sampling of wealthy addresses. 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 90<sup>th</sup> percentile of shareholding value were then oversampled at a rate of 2.5 (wave one) or 3.0 (waves three, four and five – 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 three waves of the Wealth and Assets Survey.

Wave	Issued addresses	Achieved households	Achieved adults*
One	62,800	30,500	53,300
Two	32,200	20,000	34,500
Three	37,900	21,450	40,400
Four	35,300	20,200	38,300
Five	32,700	18,400	35,600

\*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. In wave three, follow-up of the respondents and non-contacts at wave one and wave two was supplemented by the introduction of a new random sample of around 12,000 addresses. In wave four, the follow-up of the respondents and non-contacts at wave two and three was again supplemented by the introduction of another new random sample of around 8,300 addresses.

For wave five, the follow up of households who responded in either wave three or wave four was supplemented by a random sample of approximately 6,000 addresses.

### **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. The introduction of a new cohort of addresses in wave three is shown in blue.

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.

***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.***

	Jul-06	Jul-07	Jul-08	Jul-09	Jul-10	Jul-11	Jul-12	Jul-13	Jul-14	Jul-15	Jul-16	Jul-17
Wave 1		Yr 1	Yr 2									
Wave 2				Yr 1	Yr 2							
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				
Wave 5									Yr 1	Yr 2		
Wave 5 new cohort									Yr 1	Yr 2		
Wave 6										Yr 1	Yr 2	
Wave 6 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 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. In the vast majority of cases there is no change to the household's address or composition so the interview is very short (about five 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 Interviewer 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 Interviewer 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 attend a face to face briefing of up to 12 interviewers.

The briefing day is tailored to provide an overview of the survey highlighting the importance of administering the filter and seal screens. The briefing is led by one or two Interviewer Managers, sometimes with support from an experienced interviewer/mentor.

The interviewers complete a post briefing online questionnaire (currently via Survey Monkey) testing the interviewers knowledge of the survey. Any concerns are passed onto the interviewer's manager.

Interviewers do not start WAS work until their Interviewer 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 has produced the interviewer guidance on calling patterns designed to maximise contact as efficiently as possible. This strategy 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<sup>1</sup>.

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***

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.

Where an interviewer discovers a concealed multi-household address in England, Wales or Scotland the same procedure is followed. The interviewer lists all the addresses found using a standard method, and then a kish grid is used to select one address to be sampled. The interviewer uses a grid row value assigned to each address in conjunction with the number of addresses found at the sampled address. If 6 concealed addresses are found, and the sampled address has a grid row value of 12, address 2 is selected for interview. Any sampled addresses identified by the interviewer as non-private or non-residential are coded as ineligible.

---

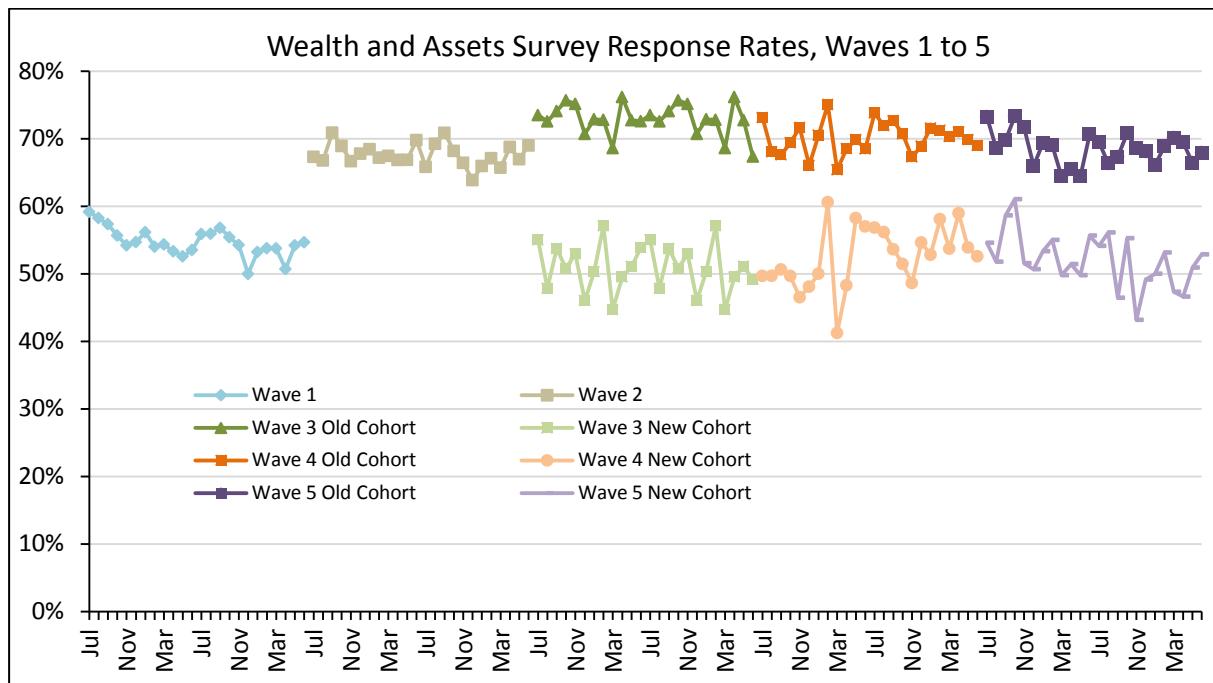
<sup>1</sup> Hopper, N.: “An Analysis of optimal calling pattern by Output Area Classification”, ONS Working Paper, Methodology Division, 2008

## Kish Grid

		NUMBER OF HOUSEHOLDS AT ADDRESS														
PLEASE RING		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	<b>1</b>	1	1	1	2	3	<b>4</b>	5	5	5	5	5	5	6	6	6
	<b>2</b>	1	2	3	4	5	<b>6</b>	7	8	9	10	11	12	13	14	15
	<b>3</b>	1	2	2	3	4	<b>5</b>	6	7	8	9	10	11	12	13	14
	<b>4</b>	1	1	1	1	1	<b>1</b>	1	1	1	1	1	1	2	2	2
	<b>5</b>	1	2	3	3	2	<b>3</b>	4	4	4	4	4	4	5	5	5
	<b>6</b>	1	1	2	1	1	<b>2</b>	2	2	2	2	2	2	3	3	3
	<b>7</b>	1	1	1	2	2	<b>1</b>	3	3	3	3	3	3	4	4	4
<b>GRID ROW VALUE</b>	<b>8</b>	1	2	3	4	5	<b>6</b>	7	6	7	8	9	10	11	12	13
	<b>9</b>	1	2	2	3	4	<b>4</b>	5	6	6	7	7	8	9	10	11
	<b>10</b>	1	1	2	2	3	<b>3</b>	4	5	5	6	6	7	8	9	10
	<b>11</b>	1	2	3	4	5	<b>5</b>	6	8	9	8	8	9	10	11	12
	<b>12</b>	1	1	1	1	3	<b>2</b>	2	3	3	3	3	6	7	7	8
	<b>13</b>	1	1	1	1	1	<b>1</b>	1	1	1	1	1	1	1	1	1
	<b>14</b>	1	2	3	3	4	<b>3</b>	3	4	4	4	4	3	3	8	9
	<b>15</b>	1	1	2	2	2	<b>2</b>	1	2	2	2	2	2	1	1	7
	<b>16</b>	1	2	3	4	3	<b>6</b>	4	7	8	7	8	7	7	7	7
	<b>17</b>	1	1	1	1	1	<b>4</b>	2	2	6	5	6	5	5	5	5
	<b>18</b>	1	2	2	2	2	<b>5</b>	3	3	7	6	7	6	6	6	6
	<b>19</b>	1	2	3	4	5	<b>4</b>	7	6	7	10	11	10	11	12	12
	<b>20</b>	1	1	2	3	4	<b>3</b>	6	5	5	9	10	9	9	9	9
	<b>21</b>	1	2	1	2	3	<b>2</b>	5	4	4	5	9	8	8	8	8
	<b>22</b>	1	1	1	1	2	<b>1</b>	3	1	3	3	5	4	4	4	3
	<b>23</b>	1	2	3	4	5	<b>6</b>	7	8	9	10	11	12	13	14	15
	<b>24</b>	1	1	2	3	4	<b>5</b>	6	7	8	8	9	11	12	13	13
	<b>25</b>	1	1	1	1	1	<b>1</b>	1	1	1	1	2	3	2	2	1
	<b>26</b>	1	2	3	3	3	<b>4</b>	4	4	6	6	5	7	10	10	10
	<b>27</b>	1	1	2	2	1	<b>2</b>	2	2	2	2	3	4	5	3	2
	<b>28</b>	1	2	2	4	4	<b>5</b>	5	6	6	7	7	9	10	11	11
	<b>29</b>	1	2	3	4	5	<b>6</b>	6	8	8	9	10	11	12	13	14
	<b>30</b>	1	1	1	1	2	<b>3</b>	2	3	3	4	4	5	6	6	4

## Response rates

The following graph provides household response for waves one, two and three, four and five 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.

In wave 3, interviews were attempted with the responding households and non-contacts from waves one and two. In addition to this a new random sample of around 12,000 addresses was added. In wave 4, interviews were attempted with the responding households and non-contacts from wave two and three. In addition to this a new random sample of around 8,300 addresses were added.

In wave 5, interviews were attempted with households that had responded in either wave 3 or wave 4. In addition, a new random sample of 6,000 addresses were added. Response rates for these “old” and “new” cohorts in wave three and four are shown separately.

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

<b>Outcome</b>	<b>Wave one</b>	<b>Wave two</b>	<b>Wave 3 old</b>	<b>Wave 3 new</b>	<b>Wave 3 all</b>	<b>Wave 4 old</b>	<b>Wave 4 new</b>	<b>Wave 4 all</b>	<b>Wave 5 old</b>	<b>Wave 5 new</b>	<b>Wave 5 all</b>
Issued cases	<b>61,917</b>	<b>32,195</b>	<b>25,234</b>	<b>12,683</b>	<b>37,917</b>	<b>27,062</b>	<b>8,269</b>	<b>35,331</b>	<b>26,739</b>	<b>6,002</b>	<b>32,741</b>
Eligible cases	55,835	29,584	21,397	11,297	32,694	23,199	7,417	30,616	22,795	5,371	28,166
Co-operating households	<b>30,511</b>	<b>20,009</b>	<b>15,517</b>	<b>5,734</b>	<b>21,251</b>	<b>16,238</b>	<b>3,894</b>	<b>20,132</b>	<b>15,622</b>	<b>2,793</b>	<b>18,415</b>
Non-contacts	3,889	2,717	1,503	988	2,491	1,648	614	2,262	1,783	483	2,266
Refusal to HQ	3,805	1,268	809	876	1,685	930	561	1,491	782	423	1,205
Refusal to interviewer	15,397	4,527	2,868	3,296	6,164	3,435	2,060	5,495	3,487	1,449	4,936
Total Refusal	19,202	5,795	3,677	4,172	7,849	4,365	2,621	6,986	4,269	1,872	6,141
Other non-response	1,770	1,063	700	403	1,103	948	288	1,236	1,121	223	1,344
<b>Response rate</b>	<b>55%</b>	<b>68%</b>	<b>73%</b>	<b>51%</b>	<b>65%</b>	<b>70%</b>	<b>53%</b>	<b>66%</b>	<b>69%</b>	<b>55%</b>	<b>65%</b>
Non-contact	7%	9%	7%	9%	8%	7%	8%	7%	8%	9%	8%
Refusal to HQ	7%	4%	4%	8%	5%	4%	8%	5%	3%	8%	4%
Refusal to interviewer	28%	15%	13%	29%	19%	15%	28%	18%	15%	27%	18%
Other non-response	3%	4%	3%	4%	3%	4%	4%	4%	5%	4%	5%

## ***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.<sup>2</sup>

---

<sup>2</sup> 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 Millennium Cohort Study. *International Journal of Social Research Methodology*, 10(5), p. 325

The longitudinal design of WAS requires following OSMs over time in order to be able to measure changes in wealth. It is evident that tracing and following sample members becomes difficult when circumstances of sample members, in particular their location, change over time.<sup>3</sup> 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.

---

<sup>3</sup> 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 at that time; namely the Department for Business Innovation and Skills (BIS); Department for Work and Pensions (DWP); HM Revenues and Customs (HMRC); HM Treasury (HMT); 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, income 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 five waves of WAS.

<b>WAS wave</b>	<b>Mean interview length (mins)*</b>	<b>75<sup>th</sup> percentile</b>	<b>90<sup>th</sup> percentile</b>
WAS (wave 1)	88	103	135
WAS (wave 2)	85	104	137
WAS (wave 3)	75	91	119
WAS (wave 4)	72	86	112
WAS (wave 5)	76	91	118

The mean wave one interview length was 88 minutes and has remained relatively consistent for wave two of the survey. The mean wave three interview length reduced to 75 minutes and then again in wave four to 72 minutes

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 decreased in wave four to one hour and 52 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<sup>4</sup>.

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:

- 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

---

<sup>4</sup> <http://www.blaise.com/capabilities>

- 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 fourth wave of WAS were similar to those used in previous waves: more details are provided in section 10.4 of the wave one report<sup>5</sup>. However, due to the longitudinal component of the survey design, part of the achieved sample size in wave five is linkable to the previous 4 waves of 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 fourth wave to validate or edit outliers. As opposed to checks for the property and physical wealth data, 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

---

<sup>5</sup> <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>

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 longitudinal outliers, the circumstances of relevant respondents in both waves had to be considered to decide whether the value in either wave three or wave four was correct. As with the cross-sectional checks, only a small number of longitudinal 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 (WAS) contains missing values. Item non-response occurs when a respondent does not know or refuses to answer a particular survey question. Unit non-response refers to cases 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 bias, error, and inconsistencies in estimates and publication figures. Imputation is a statistical process that serves to counter these problems by estimating the statistical properties of the missing data. These estimates are used to replace missing data with valid, plausible values.

Information about discrete assets or liabilities recorded by the Wealth and Assets Survey are collected through a relatively consistent question structure. Typically, an affirmative response to routing questions designed to determine; *do you have asset/liability x?* is 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.

The key analytical aim of the WAS survey is to provide longitudinal estimates of change over time as well as cross-sectional/single year estimates. To meet this aim the imputation strategy is designed to estimate the longitudinal and cross-sectional properties of the missing data depending on the availability of data from previous waves. A significant feature of the WAS imputation strategy is that the accuracy of longitudinal survey estimates is enhanced by re-imputing data imputed in a previous wave relative to any new observation in the current wave. The strategy is also designed to preserve relationships in the data defined by implicit laws or regulations governing the absolute value of an asset or liability. Important or significant relationships between variables in the end-to-end question set and between persons in a household are also taken into account.

In general, the WAS imputation strategy is based on non-parametric Nearest-Neighbour donor-based imputation methodology (Bankier, Lachance, & Poirier, 1999; Durrent, 2005; Waal, Pannekoek, & Schiltus, 2011). In this framework, missing data is replaced with plausible values drawn from other records in the data belonging to respondents with similar characteristics. Donor based methods serve to avoiding the distributional assumptions associated with parametric methods. Consequently, they are robust; imputed values are always consistent with values actually observed in the data; and imputation can be applied under strict edit constraints ensuring relationships between variables are appropriately maintained. Significantly, if applied correctly, donor-based methods will preserve the conditional statistical distributions in the observed data and/or adjust them in the presence of a non-response bias (Rubin, 1987; Chen & Shao, 2000, Durrent, 2005).

## **Donor Selection**

The key to a successful application of Nearest-Neighbour imputation is the selection of a pool of suitable ‘potential’ donors. In general, selection is based on information specified by other ‘auxiliary’ variables in the data. The set of auxiliary variables is typically referred to as the imputation model. The imputation model ensures that imputed values are drawn from a representative distribution of plausible values from respondents with similar characteristics. 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.

For a discrete imputable record, the pool of potential donors is determined by calculating the ‘distance’ between a record that need repair and other fully observed respondent records, keeping only those that match with minimum distance. Where appropriate, auxiliary variables in the imputation model are given a higher weight to account for cases where some auxiliary information is more important. 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:-

$$x_f = \text{the recipient record with } n \text{ auxiliary variables}$$

$$x_d = \text{the potential donor record with } n \text{ auxiliary variables}$$

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

$\omega_i$  = 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:-

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

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

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

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

The final imputed value is selected from the donor pool based on the probability distribution associated with the range of plausible values.

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<sup>1</sup>**

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

<sup>1</sup> Applied only to cross-sectional data where the respondent was new to the survey and did not have observed data for other waves.

## 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 previous and current wave.

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 previous and current wave. To this end, when imputing each variable, respondents with longitudinal data were divided into four imputation groups as outlined in Table 2.

**Table 2. Previous and current wave longitudinal Imputation groups**

Data Status		
Previous wave	Current wave	Imputation Group
Observed	Observed	Potential donor (O:O)
Missing	Missing	Missing both waves (M:M)
Missing	Observed	Missing previous wave (M:O)
Observed	Missing	Missing current wave (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, the value observed in one wave was included as a heavily weight auxiliary variable in the imputation model. 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 change 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 the current wave in the presence of observable data in the previous wave. 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 current wave in the presence of observed data in previous wave**

Imputation Class		Matching Variable		
Variable	Classification	Variable	$\omega$	Classification
Banded Estimate 1: Less than £2,500 observed in	2: £2,500 > £4,999	Gross Salary (An.)	0.3	Various amounts
		Both waves		

Previous wave	3: £5,000 > £9,999 4: £10,000 > £19,999 5: £20,000 > £49,999 6: £50,000 > £99,999 7: £100,000 or more 8: No Pension in Wave1	Employment Status Both waves	0.2 1: Employee 2: Self-Employed
Banded Estimate	1: Less than £2,500	Age Group	0.1 1: 16-24
Current wave	2: £2,500 > £4,999 3: £5,000 > £9,999 4: £10,000 > £19,999 5: £20,000 > £49,999 6: £50,000 > £99,999 7: £100,000 or more	Current wave	2: 25-44 3: 45-59 (Female) 45-64 (Male) 4: 59-74 (Female) 5: 65-74 (Male) 5: 75+
tSample	3 month sampling time frame	Sex	0.1 1: Male 2: Female
Both waves		NS-SEC	0.1 1: Professional 2: Intermediate 3: Routine 4: Never worked 5: Unclassified
<i>To impute missing data in current wave 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 previous wave, observed Banded Estimates in current wave, and tSample in both waves. The category 'No Pension in previous wave' helped differentiate between new and established pensions.</i>		Employment Sector	0.1 1: Private 2: Public 3: Other
<i>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 previous wave and current wave data for these variables were included in the donor selection process.</i>		Education	0.1 1: Degree level 2: Other level 3: Level unknown 4: No qualifications

## **Quality Assurance and Evaluation**

For all Wealth and Asset variables, the imputed data was examined and tested before being formally accepted. The overarching aim of the 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 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 methodology of WAS data enables it to be used for estimating wealth across the whole of the GB population, as well as enabling more detailed analysis such as longitudinal analysis.

This chapter will discuss the survey weighting methods applied to calculate the wave five (W5) longitudinal and cross sectional weights. Methods used to account for attrition; non-response and the multi-panel complex design are discussed. The properties of the final weights are also presented.

WAS is designed to follow the same people over time; this is achieved through consecutive waves of interviews. This longitudinal perspective of the survey allows for estimation of gross change over time.

The cross sectional perspective of the survey is another important feature as it allows for estimation of wealth at certain time points. Top-up panels have been introduced in wave 3 (W3), wave 4 (W4) and wave 5 (W5) to firstly increase the sample size and to secondly update the sample.

## **Different Types of Weights**

As the survey develops there are numerous longitudinal weights that could be calculated from the many different combinations of the waves. It was proposed from W3 onwards to produce three types of weights calculated at each wave; these are as follows:

longitudinal weight for the survivors from wave 1 to wave T (W1-W5)  
longitudinal weight from wave (T-1) to wave T (W4-W5)  
cross-sectional weight for wave T (W5)

At W5, the longitudinal weight for the survivors consists of responders to all five waves, i.e. W1, W2, W3, W4 and W5. The longitudinal weight for the latest two consecutive waves is applied to all responders in W4 and W5. This includes respondents sampled from the original panel, joiners in waves 2, 3 and 4 and the W3 and W4 top up panels. Finally, the cross-sectional weight incorporates all responders to W5. This includes respondents sampled from the original panel, joiners since W1, the W3 top-up panel, joiners in W4, the W4 top-up panel, joiners in W5 and the W5 top-up panel.

## **Longitudinal Weights**

The weighting strategy is based on a principle of maintaining the link between the initial selection probability and the ongoing loss-to-follow-up adjustments that remain for the evolving respondent subset over time. This is achieved through developing the longitudinal base weight (see e.g. Verma *et al.* 2007). This principle enables the weights to refer back to the desired population as closely as is possible with the current sample design and respondent follow-up procedures.

Different longitudinal base weights are used to construct the two longitudinal weights. The product of the relevant W4 weight and the W5 attrition weights creates the W4 longitudinal base weight. The *relevant* W4 weight is different for the two longitudinal weights, it is

- 1) the W1-W4 longitudinal weight for the survivors, and
- 2) the W4 cross-sectional weight for the (T-1) to T cases (W4-W5).

As detailed above, the W4-W5 longitudinal weight consisted of two panels. Each panel was weighted separately and then combined with respect to effective sample size of each panel.

The first step in the weighting process is to develop the attrition models for W5. Two separate steps were used to adjust for attrition:

- a model for unknown eligibility status<sup>6</sup>
- a model for non-response/non-contact

In both cases logistic regression was used to predict follow-up propensity, first for known eligibility status and second for a response. This gives us an estimated propensity for each case denoted by  $\hat{\phi}$ . Generically, i.e. ignoring subscripting, this is calculated as:

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

where  $\hat{\beta}$  is a vector of coefficients estimated by the regression model, and  $\mathbf{x}$  is a vector of response predictors in the regression model.

The first model predicted the log-odds of known to unknown eligibility status, using a set of characteristics taken from the W4 survey data and using the W4 weight in the analysis. In many cases, both respondents and non-respondents to W5 have data from W4, so a rich set of response predictors is available. The ‘unknown eligibility status’ weights were calculated as follows:

$$2. \quad w_{5i}^o = \frac{1}{\hat{\phi}_i^o}, \quad i \in s_5^o$$

where  $\hat{\phi}_i^o$  is the predicted probability that eligibility status at W5 is known, from the logit model regressing eligibility status on various W4 individual and household level characteristics for person  $i$ .  $s_5^o$  is the sample of people who have a known eligibility status at W5. This is those who are 1) W1-W5 longitudinal cases for the survivor weights or 2) W4-W5 longitudinal cases for the (T-1) to T weights. These are the cases which are weighted up to represent the cases with an unknown eligibility status at W5.

---

<sup>6</sup> This often, but not exclusively, occurs when interviewers are unable to trace people who have moved address (either whole households or household splits), as it is not known whether they remain in the target population or not.

In equation (3) below,  $\hat{\phi}_i^r$  represents the predicted probability of response from the known eligibility status sample base, again using a logit model with W4 individual and household level characteristics as predictor variables.

$$3. \quad w_{5i}^{nr} = \frac{1}{\hat{\phi}_i^r} \quad , \quad i \in s_5^r$$

$s_5^r$  is the sample of individuals within a responding household at W5, this is those who are 1) W1-W5 longitudinal cases for the survivor weights or 2) W4-W5 longitudinal cases for the (T-1) to T weights. These are the people who are weighted up to represent the non-respondents at W5.

For individuals in a respondent household at

- 1) both W4 and W5 or
- 2) W1, W2, W3, W4 and W5 (for the survivor weights)

the longitudinal base weight ( $w_{5i}^{long}$ ) is the product of the W4 weight ( $w_{4i}^{rel}$ ) and the two loss to follow up adjustment weights. For individuals known to be part of movements out of the target population at W5 ( $Out_4$ ), ( $w_{5i}^{long}$ ) is the product of ( $w_{4i}^{rel}$ ) and the ‘unknown eligibility status’ weight:

$$w_{5i}^{long} = \begin{cases} w_{4i}^{rel} w_{5i}^o w_{5i}^{nr} & , \quad i \in s_5^r \\ w_{4i}^{rel} w_{5i}^o & , \quad i \in Out_5 \end{cases} \quad \begin{matrix} \text{longitudinal responders} \\ \text{loss to population at W5} \end{matrix}$$

The base weight ( $w_{5i}^{long}$ ) is trimmed and scaled to the W4 population total for the W4-W5 weights and the W1 population total for the survivor weights.

The two longitudinal sub-samples (eligible respondents and ineligible outflows) are, after adjustment for attrition, representative of the populations to which they relate back to (W4 time point for the W4-W5 dataset and W1 time point for the survivors), so it is possible to calibrate the longitudinal base weight to the relevant population totals.

The calibration weights are calculated to minimise the distance between the pre-calibration weight ( $w_{5i}^{long}$ ) and the calibrated weight (which we write as an adjustment of the pre-calibration weight,  $g_i w_{5i}^{long}$ ). While summing to a set of known calibration totals, the g-weight helps to rebalance the sample towards the population values of the variables included in the calibration model.

Basic descriptive statistics relating to the W5 longitudinal weights for both the W4-W5 dataset and the survivors dataset are provided below. Descriptive statistics for weights from previous waves are also provided for comparison purposes.

## Descriptive statistics for the longitudinal weights

Weight	n	Mean	Standard deviation	Coefficient of variation	Minimum	Maximum
<b>W1-W2-W3-W4-W5 longitudinal “survivors”</b>	15,847	3,669	2,811	76.6	246	28,592
<b>W1-W2-W3-W4 longitudinal</b>	21,247	2,736	1,869	68.3	271	13,289
<b>W1-W2-W3 longitudinal</b>	28,696	2,026	1,208	59.6	267	7,000
<b>W4 - W5 longitudinal</b>	32,093	1,941	1,417	73.0	55	9,068
<b>W3 – W4 longitudinal</b>	33,525	1,834	1,327	72.4	87	10.093
<b>W2 - W3 longitudinal</b>	31,472	1,870	1,236	66.1	155	7,000

The sample size for the survivor W1-W5 group is smaller than the previous survivors' sample size, as we see more people dropping out of the survey. This group of people has the largest mean weight as the smaller sample size results in each individual representing a higher proportion of the population.

The W4 top-up panel has caused the W4-W5 longitudinal sample size to remain similar to the W3-W4 sample size. This has resulted in a small decrease in the mean weight. It is interesting to note that the variation in the weights is higher for the W4-W5 longitudinal cases than it is for the survivors, even though the survivors are a smaller subset. As the W4-W5 longitudinal weight consists of two panels, this is to be expected.

## Cross-sectional Weights

A W5 pseudo cross-sectional weight has been created; pseudo because the data used contains any W5 respondent regardless of the panel they belong to. Therefore, the samples that were selected in W1 or W4 may not be representative of the W5 time point population. The original panel, the W3 panel, the W4 panel and the W5 panel have had weights calculated separately. The three panels were then combined with respect to effective sample size.

### Cross-sectional weights for the original panel

There are several subgroups within the cross-sectional population and so there were many different methods implemented when producing the cross-sectional weights. Those who had previously responded had their most recent weight as their base weight. For cases responding for the first time, the weights were created as follows:

- W5 joiners to households (including births) - based on a weight share derived from the base weights for individuals within their household.

- W5 first time responders (from W4 or W5 top-up panel) - original design weights, constructed as the inverse of the selection probabilities.

The final stage for each subgroup is a rescaling of the weights to a specified total.

The first stage for constructing the W5 cross-sectional weight was to assign all previously responding cases a base weight. The most recent weight for each of these respondents was used. For respondents who were in:

- W4 and W5: W4-W5 longitudinal weights were used
- W3 and W5: – W3 cross-sectional weights were used
- W2 and W5: – W2 cross-sectional weights were used
- W1 and W5: – W1 cross-sectional weights were used

This ensures that, where possible, weights which have been previously adjusted for non-response and unknown eligibility status were used as base weights for W5.

The next challenge was to assign a base weight to people entering the sample at W5 (joiners). The WAS weight share was constructed following Kalton & Brick (1995). This standard approach is based on the W5 household member's weights not including the joiners, and sharing these weights between all associated W5 household members.

A key challenge for the weight share method is being able to distinguish between those joiners who are new population entrants and those who were in the original population but not originally in the sample. Unfortunately it is not possible to make this distinction with WAS data and consequently, except for births, we treated all joiners as if they were in the population at the time the sample was drawn.

First, we sum the base weights of the individuals  $i$ , in each household  $j$  excluding the W5 joiners then we divide this value by the number of individuals in the associated W5 household minus the number of births, as shown in the formula immediately below, where  $bw_{ij}$  is the base weight and  $N_j^i$  is the number of individuals in household  $j$ :

$$4. \quad w_{ij} = \frac{\sum_{i \in j=1}^{N_j^i} bw_{ij}}{N_j^i - births_j}$$

This ensures that all respondents within a household have a weight but that the sum of the weights does not increase when joiners enter the sample<sup>7</sup>. On the other hand, a birth is a true increase in the population and so, in this case, the sum of the weights does increase. Births were assigned the mother's weight.

---

<sup>7</sup> This is a consequence of the assumption that population entrants were in the original sample, which we made because the data do not allow us to distinguish population entrants.

First time responders' starting weights are their original design weights, constructed as the inverse of the selection probabilities. These are then adjusted for non-response. This then requires the weights of those who had previously responded and joiners to households to be scaled down, in order to make way for the first time entrants (we do not want to over adjust for non-responders).

The W5 cross sectional weights of the original panel are then calibrated to the W5 time point population totals. The aim of the cross sectional weights is 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 W5 pseudo cross sectional weights. This results in all people in the household having the same weight, which is also the household weight. The population totals to which the weights were calibrated were based on interpolations of ONS' mid-year estimates taken from the midpoint of the W5 fieldwork period (June 2015).

### **Cross-sectional weights for the W4 panel**

The weighting strategy for the W4 panel followed a similar approach to the W1 panel strategy. The main difference is that there are no re-entrants in the W4 panel and we therefore only have to calculate weights for:

- W4-W5 longitudinal responders
- joiners and births to W4 responding households
- Wave four panel first time entrants

The W4-W5 longitudinal responders were allocated the W4-W5 longitudinal weight as their base weight. Joiners and births were allocated weights using the same methods we described above. The W4 Panel first time entrants were allocated their design weight as their starting weight and adjusted for non-response to give their base weight, which then leads to the W4-W5 responders' weights being scaled down.

Finally, the W5 cross sectional weights of the W4 panel are calibrated to the W5 time point population totals using an integrative calibration approach as described above.

The cross-sectional weights for the W3 panel are calculated in a similar way to the W1 panel.

### **Cross-sectional weights for the W5 panel**

Design weights are computed to the responders in the new panel. As there is oversampling of wealthier households, this has been incorporated into the design weight construction.

The primary sampling units (PSUs) were sampled using a standard probability proportional to size method, where the size is measured as the number of addresses<sup>8</sup> per PSU. Within each PSU, those addresses flagged as being from the predicted high wealth stratum were sampled at 3 times the rate of other addresses in the predicted low wealth stratum.

---

<sup>8</sup> The term address is taken here to refer to the delivery point as listed on the Postcode Address File sampling frame.

The address selection probabilities for addresses from the  $i$ th PSU are thus:

$$P(\text{address sampled}) = P(\text{PSU sampled}).P(\text{address sampled} \mid \text{PSU sampled}) \\ 5. \quad = \frac{nN_i}{N}.P(\text{address sampled} \mid \text{PSU sampled})$$

where  $n$  is the number of sampled PSUs,  $N_i$  is the number of addresses in the this PSU and  $N$  the total number of addresses included on the sampling frame in Great Britain.

For an address in the predicted high or low wealth stratum, respectively, the selection probabilities are:

$$6. \quad P(\text{high stratum address}) = \frac{nN_i}{N} \cdot \frac{3 \times n_{PSU}}{M_i^{lo} + 3 \times M_i^{hi}} \quad ()$$

$$7. \quad P(\text{low stratum address}) = \frac{nN_i}{N} \cdot \frac{n_{PSU}}{M_i^{lo} + 3 \times M_i^{hi}} \quad ()$$

where  $n_{PSU}$  is the number of addresses selected from each PSU  $M_i^{lo}$  and  $M_i^{hi}$  are the number of addresses in the low and high stratum, in the  $i$ th PSU. The design weights for the sampled addresses are then the reciprocal of the appropriate address selection probability.

A non response adjustment was then applied to the design weights of the responders. Non response can bias the estimates if not accounted for; therefore, responders are weighted up to represent the non-respondents from the new panel sample. This was carried out in a similar way to the attrition modelling in the longitudinal weights. A logistic regression model was used to produce the response propensity for each case denoted by  $\hat{\phi}_i^r$  (see formula 1). The adjusted weights for the new panel were calculated as:

$$8. \quad w_i^{nr} = w_i \times \frac{1}{\hat{\phi}_i^r}$$

where  $\hat{\phi}_i^r$  represents the predicted probability of response from the new panel using a logit model with region, output area classification and the wealth indicator as predictor variables and  $w_i$  is the design weight.

The result at this stage is three sets of cross-sectional weights, each calibrated to the W5 time point population totals; one set for the original panel, one for the W4 panel, one for the W4 panel and one for the W5 panel. The panels were joined together with respect to their effective sample sizes.

### Descriptive statistics for the cross-sectional weights

W4 xs	n	Mean	Standard deviation	Coefficient of variation	Minimum	Maximum
<b>W5 xs</b>	<b>42,896</b>	<b>1,462</b>	<b>1,070</b>	<b>73.1</b>	<b>21</b>	<b>12,323</b>
<b>W4 xs</b>	46,455	1,341	889	66.3	36	4,847
<b>W3 xs</b>	49,447	1,207	877	72.7	69	9,999

The mean weight for W4 is quite similar to that of W3; the slight increase is due mostly to the decrease in the sample size from W3 to W4. Similarly the sample size for W5 is smaller than that for W4, and the mean weight for W5 is slightly higher than that for W4.

WAS weights are (model-assisted) design-based weights. Users can re-scale and normalise, if they so wish, in order to get the weights to sum to the sample size. However, the majority of popular statistical packages available on the market now account more accurately for the weights, so we recommend using survey-based procedures, where possible, when using weights in the analysis of survey data.

## 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.

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  $Y_i^P = \alpha_i n_i s_i$

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

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

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

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

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

$R$  is the normal pension age in the pension scheme

$a$  is the individual's age at interview

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

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

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

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

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

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

#### **Definition of current personal pension wealth**

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

Individuals were asked to report the value of their fund at the time of the interview and were encouraged to consult recent statements where available. Current personal pension

wealth is, therefore, simply defined as the fund value reported by the respondent at the time of the interview.

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

Individuals could report up to three pensions in which rights have been retained. These could be either DB or 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 Craig Orchard:

Phone: 01633 455755

Email: [Craig.Orchard@ons.gov.uk](mailto:Craig.Orchard@ons.gov.uk)