Office for National Statistics

Wealth and Assets Survey User Guide Round 7

Table of Contents

Overview

The Wealth and Assets Survey (WAS) is a longitudinal survey that interviews across Great Britain i.e. England, Wales and Scotland (excluding North of the Caledonian Canal and the Isles of Scilly).

Respondents to Wave 1 (July 2006 – June 2008) of the survey were invited to take part in a Wave 2 follow up interview two years later (July 2008 – June 2010). Respondents to Wave 2 were then invited to take part in a Wave 3 follow up interview a further two years later (July 2010 – June 2012). In addition to these re-curing follow up interviews, a new random sample of addresses was also added at Wave 3, Wave 4, Wave 5, Round 6 and Round 7.

The approximate number of achieved households in each wave/round was:

- Wave 1 achieved approximately 30,000 household interviews
- Wave 2 achieved approximately 20,000 household interviews
- Wave 3 achieved approximately 21,000 household interviews
- Wave 4 achieved approximately 20,000 household interviews
- Wave 5 achieved approximately 18,000 household interviews
- Round 6 achieved approximately 18,000 household interviews
- Round 7 achieved approximately 17,500 household interviews

This periodicity, that started in July and ended in June two years later and is referred to as a "Wave", was maintained until Wave 5, which covered the period July 2014 to June 2016.

The survey has now moved to a two-year, financial year-based periodicity (April to March), referred to as a "Round"; Round 6 covered the period April 2016 to March 2018 and Round 7 covers the period April 2018 to March 2020. This move allows WAS to be integrated and analysed with other household financial surveys that report on financial years.

The economic well-being of households is often measured by its income, despite a household's resources being 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 by gathering information on the ownership of assets (financial, physical and property), pensions, savings and debt.

Round 7 of WAS was funded by a consortium of government departments: Department for Work and Pensions; HM Revenues and Customs; Scottish Government and the Office for National Statistics. Fieldwork is undertaken by the Office for National Statistics.

Interviews in all waves were conducted using Computer Assisted Personal Interviewing (CAPI).

Note from WAS team - November 2023:

"The Office for National Statistics has identified a very small number of outlier cases present in the seventh round of the Wealth and Assets Survey covering the period April 2018 to March 2020. Our current approach is to treat cases where we have reasonable evidence to suggest the values provided for specific variables are outliers. This approach did not occur for two individuals for several variables involved in the estimation of their pension wealth. While we estimate any impacts are very small overall and median pension wealth and median total wealth estimates are unaffected, this will affect the accuracy of the breakdowns of the pension wealth within the wealthiest decile, and data derived from them. We are urging caution in the interpretation of more detailed estimates."

Using Wealth and Assets Survey Data

Content of data files

The data are split into two linked files:

- 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
- a person level file consisting of all person level financial wealth, pension wealth and income component variables and DVs

Variable naming conventions

Variable categories

Naming conventions are applied in order to aid understanding of the theme related to a variable.

Variables belonging to the same category often start the same character that relates to the category. For example, pension variables start with the letter ‘P’ e.g. ‘PocTyp1R7’. Most of the variable names are linked to the description of the variable e.g. ‘LnLft1’ – Number of monthly installments left to pay loan one. The majority of derived variables names have the prefix ‘DV’.

Round suffix

Most variables in both datasets are given the suffix ‘R7’, to indicate that they contain values collected in Round 7. Equally, variables within the R6 datasets are given the suffix ‘R6’ to indicate collection within the round 6 period (April 2016 to March 2014). Refer to “Using the “Round” datasets” for further detail.

Wave suffix

A smaller number of variables are given the suffix ‘W5’, ‘W4’ ‘W3’ ‘W2’ or ‘W1’ to indicate they contain values from wave one to wave five. Most of the variables with a previous wave suffix are present to allow the datasets to be matched to those from previous waves of the survey, e.g. HHSERIALW2.

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 R7 as described above, e.g. FSInValR7. The version that contains both observed and imputed values will end with the suffix ‘_i’, e.g. FSInValR7_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. DVFBondVR7_aggr at household level and DVFBondVR7 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. For further information on how weights are created, please see the following Weighting section of this guide.

Table 1 – Cross-sectional weights used across each wave of the WAS

Wave/Round	Cross-sectional Calibration Weight
1	XS_wgtW1
2	XS_calwgtW2
3	W3xswgt
4	W4xshhwgt
5	W5xshhswgt
6	R6xshhswgt
7	R7xshhwgt

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.

Table 2 – Longitudinal calibrations weights used at each wave of the WAS

Wave/Round	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
R2-R5 survivor weight	SurvWgtR2R5
R5-R6 longitudinal weight	R5R6LongWgt
R2-R6 survivor weight	SurvWgtR2R6
R6-R7 longitudinal weight	R6R7LongWgt
R2-R7 survivor weight	SurvWgtR2R7

Interview Outcome codes

The datasets include responding households only. The variable HOutR7 gives an outcome code which provides an indication of the type of interview outcome of the household.

Table 3 – Interview outcome codes for responding households in Round 7

Response Status	HOut Code	Code Description
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
Partially co-operating	211	Full response in person from HRP economic unit. One or more

		other interviews missing or complete
	212	Non-contact with one or more respondents
	213	Refusal by one or more respondents
	220	HRP economic unit not complete (one of two eligible adults missing; 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 **IOut1R7** indicates the interview outcome of individuals.

Table 4 – Individual response outcome codes for Round 7

IOutR7	**Response Status**
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's 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

As a longitudinal survey some users may wish to link data for individual respondents across Waves or Rounds. Due to the change from 'Wave' to 'Round' datasets, longitudinal data linkage can only be done correctly between Round 5, Round 6 and Round 7. For longitudinal linkage between previous Waves 1-5, please see the Wave 5 User guide.

To permit linkage all files include a single variable for linking cases. 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.

For Secure Research Service (SRS)/Government License (GL) Datasets:

- always used the linked file as a base when matching variables across waves
- use HHSerialR6/R7 for household linking
- use PIDNO for person level linking, this remains the same over the survey lifetime 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 R6 variables to the R7 person file, keeping only R7 cases:

- sort R6R7 person level linked files by PIDNO
- sort R6 Person file by PIDNO
- sort R7 Person file by PIDNO
- match R6 and R7 files with the linked file being used as a look up TABLE; use PIDNO to MATCH
- this will add R6 variables to R7 cases and R7 variables to R6 cases
- select the required cases e.g. for R7 cases (including linked R6 cases) use HHSerialR7> 0

To add R6 variables to the R7 household file, keeping all cases:

- sort the linked file by HHSerialR7
- sort R7 household file by HHSerialR7
- match files using HHSerialR7
- sort the new file by HHSerialR6
- sort R6 household file by HHSerialR6
- match files using HHSerialR6
- this will produce a linked R6R7 file with R6 and R7 variables.

Linking within a wave:

To add household variables to the R7 person file:

- sort both files by HHSerialR7 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

Please Note: Person level variables cannot be directly added to the household file unless they are first aggregated to the household level to permit a 1 to 1 match to be made.

Linking End User License (EUL) data:

As the EUL datasets are treated to minimize the risk of disclosure, the variables HHSerial and PIDNO have been anonymised. For household files, the variable CASE should be used when linking, in place of HHSerial. For person files, the variable Person should be used to link in place of PIDNO.

Linking using EUL datasets:

- use CaseR6-R7 for household linking
- use PersonR6-R7 and CaseR6-R7 for person level linking

To add R6 variables to the R7 <u>person</u> file, keeping only R7 cases:

- sort R6 by CaseR6 and PersonR6
- sort R7 by CaseR6 and PersonR6
- match R6 and R7; using PersonR6 and CaseR6 to MATCH
- this will add linkable R6 cases to the R7 file and add R7 variables to R6 cases
- select the required cases e.g. for R7 cases (including linked R6 cases) use CaseR7 > 0

To add R6 variables to the R7 <u>household</u> file, keeping all cases:

- sort R6 household file by CaseR6
- sort R7 household file by CaseR6
- match files using CaseR6
- this will produce a linked R7R6 file with R6 and R7 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.

Table 5 – Indicator for linkage status, shown through variable 'TypeR7'

TypeR7	Variable Label	Linkage Status
1	W5 and R7 linked (not W6)	Regardless of interview eligibility and response status.
2	W6 and R7 linked	Regardless of interview eligibility and response status.
3	R7 HAK Joiner	Individual joined the household when keep-in-touch exercise was conducted.
4	R7 re-entrant – linked to earlier wave	Individual joined survey prior to R6, did not respond at R6, but re-entered survey at R7.
5	R7 New respondent	Individual joined the household when R7 interview was conducted.
6	R7 New Household	Individual is part of a household that responded at R7 for the first time.
7	Individual not present at R7	This person was part of a responding household in R7 but left the household at R7 and did not respond.
8	Household not present at R7	Individual was part of a responding household in R6 but the whole household did not respond at R7.

Table 6 – Sample member status at R7, shown through variable 'P_flag1R7'

P_Flag1R7	Sample Member Status	Description
1	Longitudinal original sample member (LSM)	Individual was a member of a responding household in previous wave and current wave.
2	Entrant original sample member (ESM)	Individual was a member of a responding household in current wave, but household did not respond in previous wave.
3	Secondary sample member (SSM)	Individual was not a member of any household in previous wave but joined a longitudinal household in current wave.
4	Non-responding sample member (NSM)	Individual was a member of a responding household in previous wave but left the sample at current wave.

Table 7 – Wave entrant status at Round 7, shown through variable 'P_flag2R7'

P_Flag2R7	Wave Entrant Status	Description
1	Original Sample Member (OSM) birth entrant	Child entrant (15years or younger) born to OSM household member.
2	SSM birth entrant	Child entrant (15years or younger) born to SSM household member.
3	Other SSM entrant	Adult entrant (16years or older).

Table 8 – Wave eligibility status at Round 7, shown through variable 'P_flag3R7'

P_Flag3R7	Wave Eligibility Status	Description
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.

Table 9 – Household representative Status at Round 7, shown through variable 'P_Flag4R7'

P_Flag4R7	Household Representative (HRP) Status
1	HRP in both previous and current wave.
2	HRP in previous wave, but not in current wave.
3	HRP in current wave, but not previous wave.
4	Never a HRP.

Using the "Round" Datasets

The periodicity of the data Wealth and Assets Survey (WAS) was changed from a two-year July to June periodicity to a two-year April to March periodicity during the course of the Round 6 questionnaire. Information regarding this shift and how to use the Round-based datasets in conjunction with the Wave-based datasets can be found in the Round 6 User Guide. The Round 7 dataset is the first to comprise a fully round-based 24-month data collection period, spanning the period April 2018 to March 2020.

Individual level wealth

Individual level wealth variables are available on the person files from W3 onwards. Information on the person level wealth derived variables including the variable names and how they are derived can be found in the 'Individual level wealth derived variable specification'. The methodology used to create the measure of individual wealth can found within the document "Measuring wealth on an individual level".

ONS use the household weights (see the 'weighting' section for further information on the weight variables) when using the person level wealth variables. This is to ensure that aggregate estimates of wealth produced using the individual level wealth variables are consistent with aggregate estimates of wealth produced using the household level wealth variables.

Published Totals

You may find that the EUL dataset totals don't match the totals published by ONS. This is due to treatments applied to the dataset that permit the release of microdata that complies with disclosure best practice to maintain the anonymity of all respondents. The ONS Secure Research Service (SRS) contains a version of the data unaffected by these treatments. The linked pages provide information on how you can apply to access secure data from home. More information on Assured Organisational Connectivity can be obtained by emailing SRS.Connectivity@ons.gov.uk.

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 the area 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 the 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. This results in 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. Over sampling of wealthy address occurs because wealth has a heavily skewed distribution, with a relatively small number of addresses holding considerable wealth. Wealthy households are often harder to secure responses from, so oversampling is necessary. For year 1 of Wave 1, 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 1 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 shareholdings at a household level.

Sample sizes of each wave

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

Table 10 – Summary of sample sizes in all waves of the WAS.

Wave/Round	Issued addresses	Achieved households	Achieved adults*
1	62,800	30,500	53,300
2[1]	32,200	20,000	34,500
3	37,900	21,450	40,400
4	35,300	20,200	38,300
5	32,700	18,400	35,600
6[2]	32,000	18,000	34,000
7	33,800	17,500	38,900

*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 1 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 1. 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.

A total of 32,200 addresses were issued for Wave 2. In Wave 3, follow-up of the respondents and non-contacts at Wave 1 and Wave 2 was supplemented by the introduction of a new random sample of around 12,000 addresses. In Wave 4, the follow-up of the respondents and non-contacts at Wave 2 and three was again supplemented by the introduction of another new random sample of around 8,300 addresses.

For Wave 5, the follow up of households who responded in either Wave 3 or Wave 4 was supplemented by a random sample of approximately 6,000 addresses. For Round 6, the follow up of households who responded in either Wave 4 or Wave 5 was supplemented by a random sample of approximately 9,000 addresses. For Round 7, the follow up of households who responded in either Wave 5 or Round 6 was supplemented by a random sample of approximately 13,000 addresses

[1] For wave two, the achieved Wave 1 sample was issued, plus all of the non-contacts
[2] Due to the change from a July to an April start of fieldwork, the Round 6 response rates include the last 3 months of Wave 5 (April, May and June 2016).

Wave structure

The following diagram illustrates the longitudinal design of the Wealth and Assets Survey. Wave 1 started in July 2006 with fieldwork being spread over a two-year period. Wave 2, a follow up to Wave 1, was conducted between July 2008 and June 2010. The introduction of a new cohort of addresses in Wave 3 and subsequent Waves/Rounds 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 1 interviews conducted during July 2006 would be repeated for Wave 2 in July 2008. It is important that this gap remains constant so that estimates of change are comparable wave on wave.

Table 11 – Longitudinal survey design of the WAS – Wave Structure

	July-06 June-08	**July-08 June-10**	**July-10 June-12**	**July-12 June-14**	**July-14 June-16**	**July-16 March-18**	**April-18 March-20**
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		
Round 6						Yr 1 Yr 2	
Round 6 new cohort						Yr 1 Yr 2	
Round 7							Yr 1 Yr 2
Round 7 new cohort							Yr 1 Yr 2

Due to the survey periodicity moving from “Waves” (July, ending in June two years later) to “Rounds” (April, ending in March two years later), interviews using the ‘Wave 6’ questionnaire started in July 2016 and were conducted for 21 months, finishing in March 2018. From Round 7, the fieldwork returned to a full 2-year, 24-month data collection period.

Round structure

Data for Round 7 covers the period April 2018 to March 2020. Data for Round 6 covered the period April 2016 to March 2018. This comprises of the last three months of Wave 5 (April to June 2016) and 21 months of Wave 6 (July 2016 to March 2018). Round 8 of WAS commenced in April 2020 and will run for two years (ending in March 2022). A level of integration took place with the other HFS surveys, with a common sample being drawn for all three surveys, and harmonisation of some income questions across the surveys. Detailed information on moving the wealth and assets survey onto a financial year basis was published on the ONS website in July 2019.

2014	Q1		
	Q2	Wave 4	
	Q3		
	Q4		
2015	Q1		
	Q2		
	Q3		
	Q4		
2016	Q1		Round 5
	Q2	Wave 5	
	Q3		
	Q4		
2017	Q1		
	Q2		
	Q3		
	Q4		
2018	Q1	Wave 6	Round 6
	Q2		
	Q3		
	Q4		
2019	Q1		
	Q2		
	Q3		
	Q4		
2020	Q1		Round 7

Round 5 and Round 6 datasets are based on a mixture of original wave-based datasets. Each wave of the survey has a unique questionnaire and therefore each of these Round-based datasets are based on two questionnaires. While there may be some changes in the questionnaires, the derived variables for the key wealth estimates have not changed over this period. The diagram above illustrates the move from Waves to Rounds. From Round 7, the dataset is based on one Round 7 questionnaire which covered the whole 24-month period.

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. The mainstage interview data is what is used for analysis, whereas the KITE is important for management of the respondents' contact information and keeping people engaged with the survey.

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 on-line, 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, by 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 of 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 work on this survey
- how progress is monitored
- how performance is assessed during data collection
- how contact is maintained with 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, and activity-based applications that test the interviewers' 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.

Interviewers also participate in specific training events such as Achieving Cooperation Training (known as ACT). This training package has been reviewed and rolled out to the entire field force. This is managed through training days at ONS HQ 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 telephone operations management, covers all aspects of the KITE interview including recording changes of address, changes in household composition and updating contact details. 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 questionnaire content. They are also sent training cases to complete to set scenarios outlined in the Home Study pack.

Prior to the Covid-19 Pandemic, interviewers attended a face-to-face briefing of up to 12 interviewers.

The briefing session 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 their 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. All new trainees are validated through a mock interview with an experienced member of the field management team which is treated as if a live interview. Only if they pass this validation can they be allocated live work on 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 interviewer guidance on calling patterns designed to maximise contact as efficiently as possible. This strategy is underpinned by a Calling Pattern 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.

Interviewers are required to call at each address in their monthly quota of 13 addresses (new cohort) at least six times with at least two calls after 6pm and a call on a Saturday. Best practice procedures whereby interviewers varied their calling times and days across the field period 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 the case number within the quota (i.e. address 001-013) as the grid row value in the kish grid, if 6 concealed addresses are found at case number 012 then the interviewer would select address 2 for interview (see example below). Any sampled addresses identified by the interviewer as non-private or non-residential are coded as ineligible.

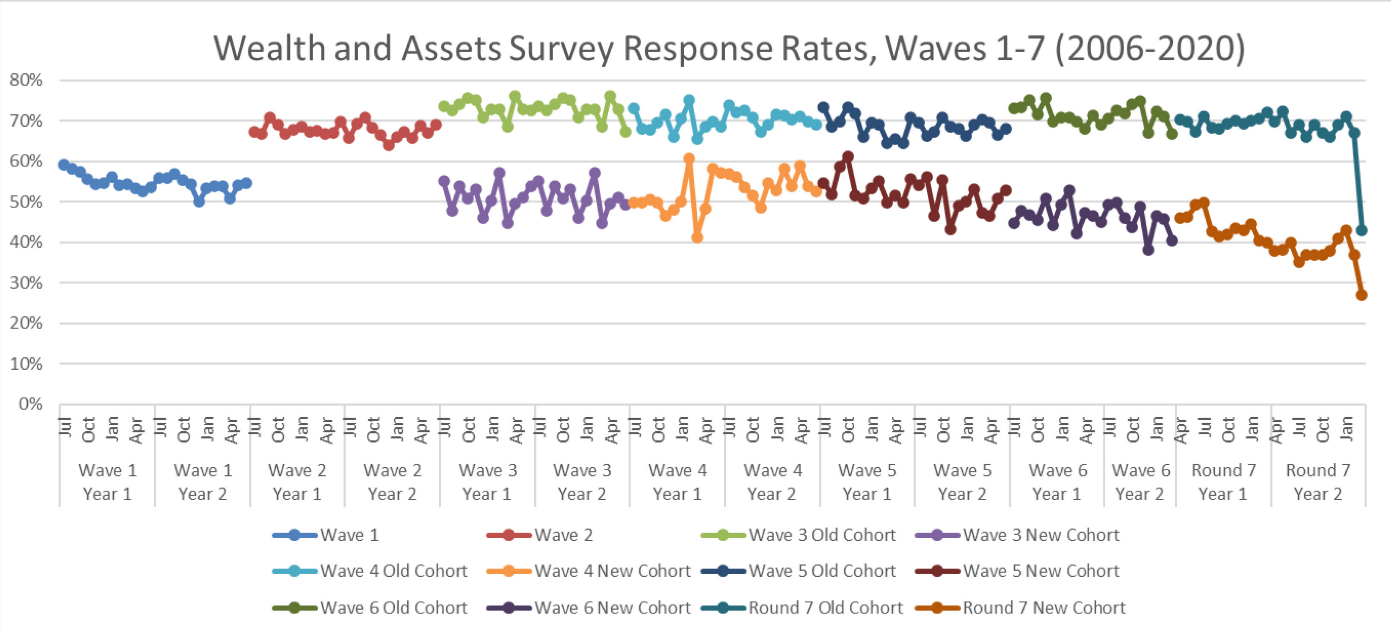
Table 12 – Kish grid used by interviewer to sample concealed multi-household addresses

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

Response rates

The following graph provides household response for Waves 1, 2, 3, 4, 5 and Rounds 6 and 7 by the monthly field periods.

Table 13: Response rates from Wave 1-7



WAS achieved an average response rate of 55 percent for Wave 1, with fieldwork being conducted between July 2006 and June 2008. The achieved sample for Wave 1 was issued for re-interview between July 2008 and June 2010, yielding an improved response of average response rate of 68 percent for Wave 2.

From Wave 3 onwards the original sample and the additional, supplementary sample have had their information collected as separate cohorts. For Wave 3, from July 2010 to June 2012, the average response rate was 65 percent. For Wave 4, from July 2012 to June 2014, the average response rate was 66 percent. For Wave 5, from July 2014 to June 2016, the average response rate was 65 percent. For Round 6, from April 2016 to March 2018, the average response rate was 63 percent.

For Round 7, from April 2018 to March 2020, WAS achieved an average response rate of 58 percent. The “old” cohort had a response rate of 68 percent whilst the “new” cohort had a response rate of 41 percent.

Response rates for these “old” and “new” cohorts in all Waves/Rounds are shown separately.

The following table provides a detailed breakdown of the outcome of cases included in the set sample for Waves 1, 2, 3, 4, 5 and rounds 6 and 7 .

Table 13 Response rates from Wave 1-7 by cohort

Outcome	**Wave 1**	**Wave 2**	**Wave 3 old**	**Wave 3 new**	**Wave all**	**Wave 4 old**	**Wave 4 new**	**Wave 4 all**	**Wave 5 old**	**Wave 5 new**	**Wave 5 all**	**Round 6 old**	**Round 6 new**	**Round 6 all[3]**	**Round 7 old**	**Round 7 new**	**Round 7 all**
Issued cases	61,917	32,195	25,234	12,683	37,917	27,062	8,269	35,331	26,739	6,002	32,741	22,441	9,712	32,153	19,836	12,987	32,823
Eligible cases	55,835	29,584	21,397	11,297	32,694	23,199	7,417	30,616	22,795	5,371	28,166	19,655	8,676	28,331	18,583	11,608	30,461
Co-operating households	30,511	20,009	15,517	5,734	21,251	16,238	3,894	20,132	15,622	2,793	18,415	13,926	4,033	17,959	12,815	4,726	17,541
Non-contacts	3,889	2,717	1,503	988	2,491	1,648	614	2,262	1,783	483	2,266	1,162	864	2,026	1,266	1,136	2,402
Refusal to HQ	3,805	1,268	809	876	1,685	930	561	1,491	782	423	1,205	625	655	1,280	559	944	1,503
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	2,720	2,684	5,404	2,793	4,105	6,898
Total Refusal	19,202	5,795	3,677	4,172	7,849	4,365	2,621	6,986	4,269	1,872	6,141	3,345	3,339	6,684	3,352	5,049	8,401
Other non-response	1,770	1,063	700	403	1,103	948	288	1,236	1,121	223	1,344	1,222	440	1,662	1,420	667	2,087
Response rate	55%	68%	73%	51%	65%	70%	53%	66%	69%	55%	65%	71%	46%	63%	68%	41%	58%
Non-contact	7%	9%	7%	9%	8%	7%	8%	7%	8%	9%	8%	6%	10%	7%	7%	10%	8%
Refusal to HQ	7%	4%	4%	8%	5%	4%	8%	5%	3%	8%	4%	3%	8%	5%	3%	8%	5%
Refusal to interviewer	28%	15%	13%	29%	19%	15%	28%	18%	15%	27%	18%	14%	31%	19%	15%	35%	23%
Other non-response	3%	4%	3%	4%	3%	4%	4%	4%	5%	4%	5%	6%	5%	6%	8%	6%	7%

[3] Due to the change from a July to an April start of fieldwork, the Wave 6 response rates include the last 3 months of Wave 5 (April, May and June 2016).

Keeping in Touch

WAS is a longitudinal survey that follows all adults interviewed in Wave 1 (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.[4]

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[5]. 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.

---

[4] Portanti, M.: "Attrition on Longitudinal Survey – Literature Review", ONS Working Paper, Social Survey Division, November 2009, pg. 2

[5] Plewis, I., 2007. Non-Response in a Birth Cohort Study: The Case of the Millenium Cohort Study. *International Journal of Social Research Methodology*, 10(5), p. 3

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

Questionnaire Content

Overview

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

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

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

The Wave 1 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.

For Round 7 the mean time taken to complete the questionnaire was 80 minutes.

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.

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[6].

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 1 set of hierarchical blocks to another set. This provides valuable flexibility as it, for instance, allows an interviewer to pause an interview with 1 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

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

- 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. This helps ensure the screen seen by the interviewer is as well designed as possible, with effective interviewer prompts. This in turn helps promote interviewer-respondent rapport, thereby contributing to better data quality
- the ability to interact with a 'question by question' (QbyQ) help facility. This provides interviewers with real-time access to guidance on specific questions during the interview. This is an electronic programme that operates in conjunction with Blaise

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

Editing

An extensive range of computer edits are applied to both the household and individual questionnaires during data entry in the field and to the aggregate data file in the office. Round 7 checks followed a similar pattern to those previously completed in Round 6.

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
- relationships between items were within acceptable limits

Edits are 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 are 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 involved 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 was 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.

Due to the longitudinal component of the survey design, part of the achieved sample size in Round 7 is linkable to the previous 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 seventh round to validate or edit outliers. Checks for financial and pension wealth data were exclusively done on individual level because the data is collected for each individual in the household, while checks for the property and physical wealth data are carried out at the household level. The investigation of outliers largely focused on the top and bottom ten per cent of the distribution of each wealth component, although for some variables this proportion was reduced if the number of cases highlighted for investigation was particularly high. When outliers were investigated in the pensions or the financial section, various variables within the same wealth component section or even different sections of the questionnaire were included to establish whether particularly large outliers could be explained by the circumstances of respondents. The majority of investigated cases proved to be genuine and only a small number of cases had to be edited, data was only edited if sufficient information was recorded by interviewers to establish the correct response, and the majority of investigated cases proved to be genuine with only a small number of cases being edited.

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 current and previous waves had to be considered to decide whether the value in either Round 6 or Round 7 was correct. As with the cross-sectional checks, only a small number of longitudinal corrections were made for each wealth component variable where enough information was available.

Imputation

General Methodology

In a similar way 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 survey question. Unit non-response refers to cases where an individual in a responding household refuses to be interviewed or 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 is not known, participants are 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 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. The strategy is 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, & Schltus, 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 avoid 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 & Shoa, 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 needs 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:

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

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

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

For categorical data with no ordinal relationship between categories:-

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

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

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

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

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

Table 16 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 14 - Imputation Class and Matching Variables used for imputing values for Private Pensions[1]

Imputation Class		**Matching Variable**			
Variable	Classification	Variable	ω^1	ω^2	Classification
Banded	1: Less than £2,500	Previous value of private pension	0.5	0.5	N/A
Estimate	2: £2,500 > £4,999	Banded Net Salary (An.)	0.25	0.5	12 bands
	3: £5,000 > £9,999				1: Employee
	4: £10,000 > £19,999	Employment Status	0.05	0.1	2: Self-Employed
	5: £20,000 > £49,999				
	6: £50,000 > £99,999	Single year of age	0.05	0.1	>= 0
	7: £100,000 or more				
		Sex	0.05	0.1	1: Male
					2: Female
		NS-SEC	0.05	0.1	1: Professional
					2: Intermediate
					3: Routine
					4: Never worked
					5: Unclassified
		Employment	0.05	0.1	
		Sector	0.05	0.1	1: Private
					2: Public
					3: Other
		Time period	0.05	0.1	Quarterly from Q3 2018 to Q1 2020

[1] Weights applied to each matching variable when previous pension is observed
[2] Weights applied to each matching variable where previous pension is not observed

To impute missing values for private pensions donors were selected from an imputation class derived from the Banded Estimates. The Banded Estimate was either given by the respondent when they were no able to answer the exact amount, or imputed beforehand where not available[7]. The Banded Estimate provided an important constraint on donor selection based on observed data.

The matching variable set consisted of variables related to the observed data identified through modelling and domain- expert review. Where the previous pension value is observed, this is the highest weighted matching variables, with banded net salary assigned the next highest weight,

[7] Banded Estimates imputations would use the same matching variable set at the continuous variable

while other matching variable weights are given equal weighting. Where previous pension is not observed Annual Net Salary has the highest weight when calculating the distance between the recipient record and the potential donor as the strength of association was stronger for these variables.

Processing Strategy

The Wealth and Assets Survey data were processed in four sections: Income, Property & Physical, Pensions & 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. The simplification of the imputation strategy meant that cross-sectional and longitudinal groups (i.e., those with previous observations and those without) could be processed together. 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. Any relevant longitudinal observations were used in the imputation to help ensure interdependencies and rates of change in the data between previous and current rounds were preserved.

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. 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[8]

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, e.g. longitudinal, analyses. 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 from Wave 3 (W3) onwards to firstly increase the sample size and to secondly update the sample.

This chapter will discuss the survey weighting methods applied to calculate the Round 7 (R7; April 2018 – March 2020) longitudinal and cross-sectional weights. Methods used to account for attrition, non-response and the complex multi-panel design are discussed. The switch from waves (two-year reporting periods starting *July* of even years) to rounds (start in *April)* from Round 6 onwards necessitated some changes to the established weighting methodology for WAS waves; these modifications are also described below. Some properties of the final weights are presented.

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. From Wave 3 (W3) onwards, three types of weights were produced; for Round 7 (R7), these are as follows:

- longitudinal weights for the survivors from Round 2 to Round 7 (R2-R7)
- longitudinal weights from Round 6 to Round 7 (R6-R7)
- cross-sectional weights for Round 7 (R7)

The survivor weights are computed for individuals who responded in all rounds from Round 2 (April 2008 – March 2010) to R7. Note that for WAS waves, survivor weights were calculated for survivors from Wave 1 (July 2006 – June 2008). With the change from Waves to Rounds, it was decided to use Round 2 as the basis for the weights rather than Round 1 as the latter covers only seven quarters of sampling.

The longitudinal weights for the latest two consecutive rounds are calculated for all individuals who responded in both R6 and R7. The cross- sectional weights incorporate all responders to R7.

Computing cross-sectional weights from rounds from those for waves

Weighting a longitudinal survey such as WAS is an iterative process in which the weights computed for a given reporting period are used as input for the subsequent cycle. Weights for WAS waves were calculated this way. However, after the switch to rounds, it would have been

[8] Verma V, Betti G, Ghellini G (2007), *Cross-sectional and longitudinal weighting in a rotational household panel: applications to EU-SILC*, Statistics in Transition New Series 8, 5-50.
Davies R (2016), *The weighting methodology for Wave Four of the Wealth and Assets Survey*, ONS Survey Methodology Bulletin **75**, 42-56.
Zobay O, Law E, Merad S (2020), Changes *to the Wealth and Assets Survey weighting system because of a shift in the reporting period,* ONS methodology paper, https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/changestothewealthandassetssurveyweightingsystembecauseofashiftinthereportingperiod

impractical to systematically re-compute weights for WAS rounds by starting at Round 1 and using only the raw input information. Therefore, a method was devised with which cross-sectional weights for any WAS round could be directly derived from the existing weights of the corresponding WAS waves. This method is described below for the example of obtaining R5 weights from weights for quarter 8 (Q8) of W4 and Q1-Q7 of W5. The R5 weights are then used as input to a "round-based" system to derive weights for R6 and subsequent rounds.

1. Form quarterly datasets from the existing cross-sectional person weights for WAS W4 Q8 and W5 Q1-7
2. Calibrate each quarterly dataset to the midpoint of R5 (i.e., March 2015) using the age/sex and regional calibration groups already established for the weighting of WAS waves
3. Rescale each of the eight quarterly datasets by 1/8 and combine them to form the R5 cross-sectional person weights

Longitudinal R6-R7 Weights

The longitudinal 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 possible with the current sample design and respondent follow-up procedures. A summary of the computation of the R6-R7 weight is given below, followed by a more detailed description of some of the steps.

1. Separate the dataset for the R6 cross-sectional person weights into the original wave-based panels (i.e., W1, W3, W4, W5 Q1-Q7, R6)
2. For each panel, compute and apply attrition adjustments for drop-out from unknown eligibility and non-response between R6 and R7. This yields the longitudinal base weights
3. Combine the panels according to their effective sample sizes using the base weights computed in step 2
4. Form quarterly datasets and calibrate each to the midpoint of R6 (March 2017) using the standard calibration groups (i.e. age-sex groups and Government Office Regions
5. Rescale each of the eight quarterly sets by 1/8 and combine them to form the R6-R7 longitudinal weights

Attrition

As a first step to attrition modelling, every R6 responder is classified according to their R7 response status as either R7 responder (R), ineligible case (IE), non-responder (NR) or case with unknown eligibility (UE). R7 ineligible cases are those who do not fulfil the R7 sample inclusion criteria, for example because of death or emigration. Non-responders are known to be still in the R7 population but were not productive in R7, most often because of refusal. For some cases, R7 eligibility cannot be established, for example, if the current location of the household is unknown.

If there were no NR and UE cases, the original R6 population would be faithfully represented by the R7 R and IE cases, and no reweighting would be necessary. To deal with attrition, a two-step process is applied. First, UE attrition is handled by deriving an adjustment factor from a suitably defined logistic regression model with eligibility as outcome (i.e., unknown versus known eligibility) and fitted using all R6 responders. As all these cases have R6 data, a rich set of person- and household-level predictors is available. The 'unknown eligibility' adjustments are calculated as;

$$w_{6,i}^{0} = \frac{1}{\hat{\phi}_i^{0}}, i \in s_6^0$$

where $\hat{\phi}_i^0$ is the predicted probability that the eligibility status of case i at R7 is known, as estimated from the logistic regression model. s_7^0 is the set of people who have a known eligibility status at R7. These individuals are weighted up to compensate for the loss of cases whose eligibility status has become unknown between R6 and R7.

The second attrition model estimates the probability $\hat{\phi}_i^r$ of response, using a logistic regression fitted on the R7 responders and non-responders and including R6 individual and household level characteristics as predictor variables. The non-response adjustment is given by;

$$w_{6,i}^{nr} = \frac{1}{\hat{\phi}_i^{r}}, i \in s_7^R$$

with s_7^R the set of individuals within a responding household at R7. These cases are weighted up to compensate for loss to follow-up because of non-response.

For R7 responders, the longitudinal base weight $w_{7,i}^{long}$ is the product of the R6 cross-sectional weight $w_{6,i}^{XS}$ and the two attrition adjustments. For IE cases, the base weight is the product of $w_{6,i}^{XS}$ and the unknown-eligibility adjustment.

$$w_{7,i}^{long} = \begin{cases} w_{6,i}^{XS} w_{7i}^{0} w_{7,i}^{nr}, i \in s_7^R \\ w_{6,i}^{XS} w_{7,i}^{0}, i \in s_7^{IE} \end{cases}$$

The base weights are trimmed at the 99th percentile to reduce the effects of outliers and scaled to the R6 population total.

Combining panels.

Regression models are derived for each panel separately. After applying the attrition adjustments, the panels are combined in proportion to their respective effective sample sizes (step 3 above). The effective sample size of panel p is defined as $n_{\text{eff},p} = n_p/(1 + \text{CV}_p^2)$ with n_p the actual sample size and CV_p the coefficient of variation of the panel's weights (i.e., standard deviation divided by mean). When combining by effective sample size, weights in panel p are multiplied by a factor $n_{\text{eff},p}/N_{\text{eff}}$ with N_{eff} the sum of the effective sample sizes of all panels.

Calibration.

The combination of eligible responders and ineligible outflows is, after adjustment for attrition, representative of the population to which they relate back to (i.e., time point at the middle of R6 for the R6-R7 dataset), 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 ($w_{7,i}^{long}$) and the calibrated weight (which we write as an adjustment of the pre-calibration weight, $g_i w_{7,i}^{long}$). While summing to a set of known calibration totals, the g-weight g_i helps to rebalance the sample towards the population values of the variables included in the calibration model. As mentioned in Step 4 above, calibration is done on quarterly datasets after panels have been combined.

R2-R7 survivor weights

After the switch from waves to rounds, R5R6 longitudinal and R6 cross-sectional weights were computed starting from R5 cross-sectional weights, since a complete re-calculation starting at R1 weights was not feasible. The computation of survivor weights, in contrast, is much simpler as complications from multiple panels, joiners to households etc. do not arise. Therefore, the survivor weights could be obtained in a systematic way with an iterative process starting from the R2 cross-sectional weights.

1. *R2 weights*: R2 person-level cross-sectional weights are computed for all R2 responders using the procedure outlined above for R5 weights
2. *R2-R3 weights*: The computation of R2-R3 weights follows along the lines of the R5-R6 calculation but includes only a single panel. The R2 cross-sectional weights of R3 responders and ineligible cases are first adjusted for attrition and then calibrated to the R2 population totals
3. *R2-R4 weights*: As a first step, cases who responded in both R2 and R3 are classified according to their R4 response status. The R2-R3 weights of the R4 responders and ineligible cases are then adjusted for R4 attrition. Their new weights are combined with the unadjusted R2-R3 weights of the R3 ineligible cases and the complete set is calibrated to the R2 population totals
4. *R2-R5 and subsequent weights*: The R2-R4 weights of R5 responders and ineligible cases are adjusted for attrition between R4 and R5 and then combined with the R2-R4 weights of the R2 and R3 ineligible cases before calibration to R2 population totals. The yields the R2-R5 weights. R2-R6 and subsequent weights are computed accordingly

It should be noted that for the survivor weights, the logistic regression models describing attrition were fitted on a wave rather than a round basis. For example, R2-R3 attrition factors for R2Q1 (=W1Q8) cases were computed using the regression model developed earlier for W1-W2 weights. In contrast, attrition factors for R2Q2-8 are based on the W2-W3 model. Alternatively, one could have used a single R2-R3 model for both groups. However, it seems that attrition for R2Q1 cases who had only their second chance of responding in R3 is better modelled by the W1-W2 model in which every participant is in the same situation. In the R2-R3 model, R2Q1 responders would be mixed in with a large majority of cases who already have their third chance of responding in R3. As attrition for W2-W3 is probably lower than for W1-W2, the resulting adjustments for R2Q1 would likely be too small.

While it was feasible to implement this more complex way of attrition modelling for survivor weights, it would have been too cumbersome to maintain a similar approach for cross-sectional and round-to-round longitudinal weighting involving multiple panels (for the cross-sectional weights, attrition modelling is used for re-entrants, see below). Therefore, they make use of simpler round-to-round logistic regressions.

Cross-sectional weights

An R7 pseudo cross-sectional weight has been created; the designation “pseudo” stems from the fact that the set of weights contains R6 respondents from all panels. Therefore, the samples that were selected in, for example, R1 or R4 may not be representative of the R7 time point population.

The calculation of the cross-sectional weights proceeds in two main steps.

1. For each of the W1, W3, W4, W5 (Q1-7), R6 and R7 panels, compute a set of pre-calibration weights
2. Combine the six panels according to effective sample size and calibrate by quarter

Pre-calibration weights for returning panels (W1, W3, W4, W5 Q1-7, R6)

In the process of computing pre-calibration weights for returning panels, one needs to distinguish between different types of respondents:

- continuing cases (R6 respondents)
- joiners to households (first responding in R7 and not part of the original sample)
- births between R6 and R7
- entry original sample members (EOSMs, members of the original sample but first responding in R7)
- re-entrants (cases who last responded before R6)

The calculation proceeds as follows.

- Cases continuing from R6 are assigned their pre-calibration longitudinal R5-R6 weights $w_{7,i}^{long}$
- In a continuing household h with joiners, a weight-sharing method is applied. The weights of all continuing household members i are summed ($w_h = \sum_{i \in h} w_{7,i}^{long}$), and this sum is divided by the number n'_h of people in the household including joiners but excluding any births. Each household member is then assigned this average weight w_h/n'_h.
- The WAS weight share method was constructed following Kalton and Brick (1995) [9].. This standard approach is based on the continuing household members' weights not including the joiners and sharing these weights between all associated R7 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, all joiners are treated as if they were in the population at the time the sample was drawn.
- Any births to a continuing household between R6 and R7 are assigned their mother's weight/
- After applying these steps, the weights of all continuing households in a panel are calibrated to the R7 midpoint population totals (i.e., March 2019).
- EOSMs are assigned a weight which is the product of their original design weight d_i (i.e., the inverse of their initial sampling probability) and an attrition factor $f_{g,s}$. This factor depends on the government office region (GOR) g and the WAS wealth stratum s that the EOSM belongs to. It estimates the inverse probability of a household sampled in the same round as the EOSM to respond in Round 7, conditional on GOR and wealth stratum. It is given by $f_{g,s} = n_{g,s}/r_{g,s}$ with $n_{g,s}$ the number of addresses in the EOSM's panel that were

[9] Kalton G, Brick JM (1995), *Weighting schemes for Household Panel Surveys*, Survey Methodology 21, 33-44, https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X199500114412.

originally sampled in (g, s) and $r_{g,s}$ the number of responding households in R6. Further information on the design weights and the WAS wealth strata is provided in the section on the R7 panel.

- Re-entrants are assigned a weight that is the product of their last cross-sectional weight and an attrition factor. This factor is computed from a round-based regression model which describes the attrition rate, within the re-entrant's panel, between the round that the re-entrant last responded in and the current round (more specifically, the regression estimates the probability of responding in R7 using all cases in a panel who (a) responded in the re-entrant's last round and (b) are classified as either responders or non-responders in R7).
- EOSMs and re-entrants are added into their respective panels p after applying a shrinking factor S_p to the calibrated weights of the continuing households. This factor is given by
- $S_p = \frac{\text{Pop}_{R7} - (\sum_{i \in \text{EOSMs}} w_i + \sum_{i \in \text{Re-entrants}} w_i)}{\sum_{i \in R5-R7} w_i + \sum_{i \in \text{Joiners}} w_i + \sum_{i \in \text{Births}} w_i}$
- with Pop_{R7} the R7 population total and w_i the weights that have been assigned to the cases in the previous steps. As the continuing households have already been calibrated to the population total the shrinkage factor ensures that the weights of the complete panel still sum to Pop_{R7}.

Pre-calibration weights for the R7 panel

The computation of the weights for the new panel (i.e., cases sampled in R7) makes use of a procedure that accounts for differing sample sizes between subpanels.

1. Assign each case their original design weight
2. Form three data sets containing the Quarter 1 (Q1), Q2-Q4 and Q5-Q8 subpanels and rescale the design weights to 1/8, 4/8 and 3/8 of the R6 midpoint population total, respectively. Recombine the datasets into a single dataset. Q1 is treated separately from Q2-Q4 as its sampling scheme was different from that of subsequent quarters.
3. Apply a non-response factor computed from a logistic regression model for response propensity. The model considers all households sampled in R6 and estimates response probability using output area classification, region and the WAS wealth stratum as predictors
4. Calibrate the panel to the R6 midpoint population totals

Design weights.

Up to the end of Wave 6 (i.e., June 2018; first quarter of R7), WAS followed a two-stage stratified cluster design with primary sampling units (PSUs) given by postcode sectors. PSUs were selected by systematic sampling with probability proportional to size. Implicit stratification was achieved by ordering the sampling frame by region, metropolitan borough status, socio-economic status and proportion of households with no car. The probability p_s of selecting a given PSU s is given by $p_s = nN_s/N$ with n the number of sampled PSUs, N_s the size of the PSU (i.e., number of addresses) and N the total number of addresses on the frame.

From R7 Q2 onwards (i.e., July 2018), WAS PSUs are selected from the PSUs of the Survey of Living Conditions (SLC) that were drawn for Wave 1 of the previous year. In this way, WAS households can be visited at the same time as SLC W2 households in a given PSU area which provides some savings on cost. After deciding on the sample size for the new WAS panel, the numbers of PSUs that are needed in each NUTS2 region are determined, based on the sample composition in previous rounds. WAS PSUs are then drawn by simple random sampling from the SLC PSUs in that region. If more PSUs are required in a region than are available from the SLC, PSUs from the Living Cost and Food survey (LCF) are used as a backup. PSUs for SLC and LCF are drawn from an implicit stratification scheme that is similar to the one previously used for WAS. The probability of selecting a WAS PSU can be shown to be given by $p_s = n_{PSU,R} * S_P/S_R$ with $n_{PSU,R}$ the number of sampled PSUs in the NUTS2 region R of the PSU P, S_P the size of PSU (number of addresses on the frame) and S_R the size of the NUTS2 region.

Based on a matching against tax data, HMRC ranked all selected addresses and returned a stratification into the top 1%, 2-10% or remaining 90% presumptive “wealth groups”. To sample households within a PSU, the two higher wealth strata are oversampled by factors of 5 and 3, respectively. The sampling probability $p_h$of a household in wealth stratum m is thus obtained as

$$p_{h,m} = \frac{f_m * n_{\text{PSU}}}{M_{90\%} + 3M_{2\text{-}10\%} + 5M_{1\%}}$$

with n_{PSU} the total number of households sampled in the PSU, $f_m$the oversampling factor of wealth stratum m and $M_m$the total number of addresses in the stratum. The final design weight of a household is given by $d_h = 1/(p_s * p_{h,m})$. Note that the stratification into three groups was only introduced in R6; for previous waves two strata were used with the top 10% wealth stratum oversampled by factor of 3.

Person- and household-level cross-sectional weights

After computing the pre-calibration weights as described above, panels are combined by effective sample size and the resulting dataset is split again by quarter. Person- and household-level cross-sectional weights are then obtained by performing a specific final calibration step on the quarterly datasets. The calibrated *person-level* weights $W_{Pers,i}$ have to fulfil constraints of the type

$$\text{Calibration group total} = \sum_{\text{All individuals } i} I\ (i\ in\ calibration\ group) *\ W_{\text{person},i}$$

in each calibration group. Here, the summation extends over all individual household members, and I is an indicator taking the values 1 or 0 depending on whether the person is part of the calibration group or not. WAS uses age-sex groups and Government Office Regions for calibration.

For the *household-level* weights $W_{\text{HH},h}$, the same calibration groups are used but the constraints are now of the form

$$\text{Calibration group total} = \sum_{\text{All households } h} m_h *\ W_{\text{HH},h}.$$

Here, the summation is over all households in the round, and m_h denotes the number of household members in the calibration group in question. The input (pre-calibration) weights are the means of the person-level pre-calibration weights within households.

Following the integrated weighting method proposed by Lemaître and Dufour (1987)[10], all analyses of WAS data are carried out with the household-level weights, with individuals being assigned household weights in person-level calculations. Person-level weights are only used as starting point for the longitudinal weights.

---

[10] Lemaître G, Dufour J (1987), *An integrated method for weighting persons and families,* Survey Methodology Bulletin 13, 199-207, https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X198700214607.

Descriptive statistics for R7 weights

The table below presents basic descriptive statistics for the various types of R7 weights. For the R7 person-level cross-sectional, R6-R7 longitudinal and R2-R7 survivor weights, the sum of the weights equals the R7, R6 and R2 population totals, respectively, that were used for calibration. The sum of the R7 household-level weights provides an estimate of the total number of households in R7. However, when applied at person level, the household weights again sum to the R7 population total.

For the R6-R7 and R2-R7 weights, ineligible cases need to be retained to ensure consistency in the calibration. After excluding these cases, the sum of the weights estimates the size of population remaining since R6 and R2, respectively.

Table 17 - R7 Weights

Type of weight	N	Sum	Mean	Std Dev	Coefficient of Variation	Minimum	Maximum
R7 household-level cross-sectional	17,541	26,427,863	1506.63	1271.66	88.4%	14.35	14952.3
R7 HH weights at person level	38,930	63,726,785	1636.96	1341.84	82.0%	14.35	14952.3
R7 person-level cross-sectional	38,930	63,726,785	1636.96	1276.47	78.0%	14.07	12721.2
R6-R7 longitudinal	26,009	62,933,121	2419.67	2093.40	86.5%	26.33	15248.3
R6-R7 longitudinal excl. ineligibles	24,816	60,881,841	2453.33	2114.27	86.2%	26.33	15248.3
R2-R7 survivor	11,605	59,094,732	5092.18	5568.76	109.4%	122.86	55740.7
R2-R7 survivor excl. ineligibles	8,701	50,827,860	5841.61	5967.74	102.2%	196.07	55740.7

Cross-sectional weights from WAS waves

The table below summarizes person-level cross-sectional weights from WAS Waves 1 to Round 6. Due to updates, there can be slight differences to data published in previous user guides.

Table 18 Person level weights Waves 1 to Round 6

Type of weight	N	Sum	Mean	Std Dev	Coefficient of Variation	Minimum	Maximum
R6 XS	40,540	62,933,121	1552.37	1184.74	76.3%	18.01	9262.5
W5 XS	42,896	62,134,570	1448.49	1056.47	72.9%	21.35	12387.9
W4 XS	46,451	61,169,463	1316.86	880.69	66.9%	45.79	7618.4
W3 XS	49,445	59,658,894	1206.57	877.26	72.7%	68.90	9999.0
W2 XS	46,342	59,183,270	1277.10	730.94	57.2%	105.82	3700.0
W1 XS	71,268	58,012,026	814.00	283.66	34.8%	132.51	2245.1

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 is 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. The following guidelines were applied 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.

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
- 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:

AR is the age- specific annuity factor at normal pension age, R, based on (single life) annuity rates sourced from www.williamburrows.co.uk assuming average age- specific life-expectancies. The Discount Rate used is the Superannuation Contributions Adjusted for Past Experience (SCAPE) rate, which is set at 3 per cent above Consumer Prices Index (CPI). All rates are specific to the month of interview.

Where:

- Y_i^P is annual pension income, defined as $Y_i^P = \alpha_i n_i s_i$
- α_i is the accrual fraction in the individual's scheme 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 (SCAPE Discount Rate),
- 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 an 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 in Wave 1, up to three current personal pensions from Wave 2 onwards; 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 Wave 1) in which rights have been retained. From Wave 2 onwards individuals could report up to six pensions in which rights have been retained. These could be either DB of DC schemes. The wealth in each DB retained scheme was calculated separately (in much the same way as for current DB schemes described above) and then summed to derive total wealth held as retained rights in defined benefit (DB) occupational schemes.

Wealth in these schemes was defined as:

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

Where:

- A_R is the age-specific annuity factor at retirement age matched by month of interviewR (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 (SCAPE discount rate, see above).
- 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.

Retained rights in group personal, group stakeholder, private personal, private stakeholder, self-invested personal pension or retirement annuity contract

The wealth in each of these schemes types was calculated separately. In these cases, individuals were asked to report what the fund value for their scheme was at the time of interview.

Rights retained in schemes from which individuals are drawing down

At the start of the round 7 questionnaire wording changes were made to questions that relate to the entry into private preserved pension wealth. These changes relate to how preserved pension wealth is divided amongst three main sections, drawdown, defined benefit and defined contribution. Consequently from round 7 onwards preserved pension wealth is separated into either defined benefit or defined contribution and whether the pension pot has been partially accessed, (eg. from taking a lump sum payment, but a portion of the pot remains for access at a later date) or whether the pension pot has not been accessed at all. Preserved pension entitlements in pensions for drawdown are therefore included within either defined benefit or defined contribution where the pot has been partially accessed and not shown as a separate category. Please reference the pension derived variable specifications for details on the derivation of these estimates.

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_R 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

Household income

From Round 6, WAS changed the total net household annual income variable. Before July 2016 'Total net regular household annual income' was used. From July 2016 onwards 'Total net regular household annual income before housing costs' was used. This removes the 'Household Annual Council Tax paid' from 'Total net regular household annual income'. The move to income before housing costs was implemented to bring the definition of income in line with that used on other household financial surveys. It is not possible to create this measure of income for earlier waves as the variables needed are not available. The variables needed were introduced on the Wave 6 questionnaire that began in July 2016.

The income derived variable specifications defines how 'Total net regular household annual income’ was derived in previous waves and how 'Total net regular household annual income before housing costs’ is derived for Wave 6.

The Round 6 datasets cover the period April 2016 to March 2018 and comprise of the last 3 months of Wave 5 and 21 months of Wave 6. As the before housing costs measure cannot be created before July 2016, this measure cannot be created for the last 3 months of Wave 5. Therefore a Round 6 version of the variable has been created on the Round 6 household file (called DVTotInc_NETR6), where the value for Wave 5 respondents has been fed through from ‘Total net regular household annual income’ (DVTotNIRW5) and the value for Wave 6 respondents has been fed through from ‘Total net regular household annual income before housing costs’ (DVTotInc_BHCW6). The Wave 5 and Wave 6 derived variables remain on the file.

For Wave 6, a measure of household income after housing costs was produced called ‘Total net regular household annual income after housing costs’ (DVTotInc_AHCW6). This removes mortgage interest payments and rent from ‘Total net regular household annual income before housing costs’ (DVTotInc_BHCW6). Ideally utility bills, ground rent and service charges would also be removed to bring this measure of income after housing costs in line with that used for other household financial surveys, but these data are not collected in WAS.

The income derived variable specification outlines how this measure has been created. This is only available for the 21 months of Wave 6 and so a Round 6 version of this has not been created. This was introduced to allow testing of the variable. From Round 7 onwards this variable will be available for the full round.

Contact details

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

Phone: 01633 455295

Email: gareth.rusgys@ons.gov.uk