Chapter 8: Wealth and Assets Technical Details, 2012-14

- Coverage: **GB**
- Date: **July 2012 to June 2014**
- Geographical Area: **Region**
- Theme: **Economy**

Introduction

This chapter aims to assist readers in interpreting and using estimates from the Wealth and Assets Survey (WAS) by describing technical aspects relating to the survey. Much of the technical material regarding the survey has already been reported in Chapter 10 of the report 'Wealth in Great Britain'[1] [insert link to this report] published in December 2009 and the wave 2 User Guide[2]. Readers should consult these documents for more general technical detail.

The WAS is a longitudinal survey of private households and individuals in Great Britain (excluding the Isles of Scilly and Scotland north of the Caledonian Canal). The survey is conducted using face-to-face interviews, administered by ONS interviewers. The first wave of interviews was carried out between July 2006 and June 2008; the second from July 2008 until June 2010; the third wave between July 2010 and June 2012; and the fourth July 2012 to June 2014. The results reported on in this release describe the level of Wealth in Great Britain in July 2012 to June 2014. as well as how the level and distribution of Wealth in Great Britain has changed since wave 1 of the survey.

A summary of some of the key points has been provided below:

The Survey

- The WAS draws its sample from the population of private households in Great Britain.
- The first wave of the survey commenced in July 2006 and lasted for two years, ending in June 2008. This comprised 30,595 responding households.
- The second wave of the survey commenced in July 2008 and ran until the end of June 2010. This comprised 20,170 responding households.
- The third wave of the survey commenced in July 2010 and ran until the end of June 2012. This comprised of 21,541 responding households. It returned to responding households who gave their permission to be re-interviewed. Households who were eligible but who could not be contacted in the previous wave were approached again at wave 3. A new cohort was also introduced (12,000 issued addresses) with the aim to maintain an achieved sample size of around 20,000 responding households.

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

[2] http://www.ons.gov.uk/ons/guide-method/method-quality/specific/economy/wealth-and-assets-survey/wealth-and-assets-survey---user-guidance/volume-1--the-basics-of-was-2008-10.pdf

- The fourth wave of the survey commenced in July 2012 and ran until the end of June 2014. This comprised 20,247 responding households. It returned to responding households who gave their permission to be re-interviewed. Households who were eligible but who could not be contacted in the previous wave were approached again at wave 4. A new cohort was introduced (8,000 issued addresses) with the aim to maintain an achieved sample size of around 20,000 responding households.
- Data were collected in the field by Computer Assisted Personal Interviewing (CAPI).
- The WAS questionnaire is divided into two parts: a household questionnaire completed by one person in each household; and an individual questionnaire addressed to all adults aged 16 and over (excluding those aged 16 to 18 currently in full-time education or those aged 19 and in a government training scheme).

Survey processing

- The longitudinal editing introduced with wave 2 data (using information gathered at wave 1 to validate wave 2 data, but also looking at the wave 1 data in the light of the data given at wave 2) has again been applied at wave 3 and wave 4. However longitudinal editing is only done between wave 3 and wave 4. Data for waves 1 and wave 2 have not been re-edited.
- In any sample survey there will always be missing values for individual questions. However, when constructing estimates of wealth it is necessary that valid responses have been given for all component estimates. Therefore, any missing values are imputed. The imputation methodology has remained consistent with that used in wave 3.

Sampling

Details of the sampling design, sampling frame, sample structure and field sampling procedures underlying wave 1 of the survey are provided in the wave 1 report (link to wave 1 report). Responding households, as well as non-contacts and 'soft' refusals were included in the sample for the next wave. Any 'hard' refusals were not approached again in subsequent waves. The WAS aims to follow individuals rather than households. In the case that a household splits, with individuals living at different addresses, WAS will interview all of the original sample members (OSMs); as well as any people living with the OSMs in the next wave of the survey. The new people in the sample are referred to as secondary sample members (SSM). OSMs remain eligible for interview until they leave Great Britain, enter an institution (such as a nursing home), or die. SSMs are eligible for interview as long as they live at the same address as an OSM.

At waves 2, 3 and 4, interviews were sought from those who had been interviewed previously and those who were previously ineligible (i.e. those aged 16 or under or 16-18 and in full time education or those aged 19 and in a government training scheme) and had become eligible at the follow up wave.

The original sample approached in wave 1 was approximately 63,000 households. However, given refusals to the survey, and changes in eligibility etc, the number of households with whom contact was attempted in wave 2 was approximately 35,000. Of the 35,000 addresses attempted for wave 2, 25,000 addresses were attempted for wave 3 and 27,000 in wave 4.

Given the declining sample of eligible addresses over the life of WAS, it was decided to introduce a new panel of respondents to the survey in wave 3. A sample of new addresses was issued to supplement the existing panel. The approach to selecting these new addresses was the same as for wave 1 of the survey.

Data collection

WAS interviews take place two years after the previous wave, and generally within the same calendar month[3]. Interviewers were given an allocation of addresses on a monthly basis and were instructed to make contact and gain an interview at all of these addresses using best practice in terms of varying calling times and days. Where it was not possible to attempt contact within the month, addresses were carried forward for reissue in the following month.

Where information was unlikely to have changed, or earlier responses were likely to provide a useful *aide memoire*, answers from the previous wave were rolled forward and made available, in the computer assisted interviewing programme to the interviewer during the interviewing process. For instance, the type of tenure of the household's accommodation from wave 1 would be available to the interviewer at wave 2. However, value information, such as the value of the property, was not rolled forward.

The wave 2 questionnaire covered the same topics as wave 1, however as a result of the longitudinal nature of the survey and specifically the experience gained during wave 1, it was slightly longer[4]. The flow of questions was also improved, the types and nomenclature of some assets and debts were changed, and certain new requirements of stakeholders[5] were included. The content of the wave 4 questionnaire was broadly comparable with wave 3. Improvements were made to the conditional routing of some questions, but generally questions were unchanged so as to preserve consistency in data collection over time.

Questionnaire changes made between waves were tested both cognitively and via a quantative pilot. This ensured the new questions were both likely to be understood by respondents and were suitable for collecting the information needed.

The mean interview length varied for each wave of the survey. The wave 1 mean interview length was 79 minutes; wave 2 was 85 minutes, wave 3 was 82 minutes and wave 4 72 minutes.

Table 1 shows response rates for completed waves of WAS. An initial sample of 62,800 addresses were selected and sampled at wave 1. Of these, 30,511 took part in the survey, or 55% of the eligible sample. Approximately 10% of sampled addresses were found to be ineligible, and were therefore not interviewed at e.g. non-residential addresses. For wave 2, the cooperating wave 1 households, along with non-contacts and circumstantial refusals from wave 1 were issued for a wave 2 follow up

[3] There is some latitude in this gap in that interviewers were given sufficient time to follow up their monthly quotas so that the actual wave 1 interview could have been 23 months, rather than 24 months, earlier.
[4] Partly due to the fact that in wave 2 the physical wealth questions covered the whole sample, rather than approximately half of the sample, which was the case in wave 1, but also other changes made to the questionnaire.
[5] Stakeholders involved in the development of the questionnaire comprise of the group of funding departments which is currently: ONS, Department for Work and Pensions, HM Revenue and Customs, The Financial Conduct Authority, and the Scottish Government.

interview. The eligible sample for wave 2 of the survey was 29,341 households and of these 19,925 either fully or partially responded, giving a household response rate of 68%. This figure is not comparable with the household response rate of 55% achieved in wave 1 since the wave 2 figure is calculated as a proportion of the sample brought forward from wave 1. As a proportion of the original wave 1 sample, the response rate is 36%, which illustrates both the scale of non-response at wave 1 and subsequent attrition between waves 1 and 2.

Table 1
Household response rates for WAS wave 1 to 4

Great Britain

	Wave 4		Wave 3		Wave 2		Wave 1	
	Count	%	Count	%	Count	%	Count	%
Sample	35,446		37,881		34,737		62,800	
Ineligible	4,715		5,142		5,396		6,965	
Total eligible sample	30,731	100	32,659	100	29,341	100	55,835	100
cooperating households	20,247	66	21,065	64	19,925	68	30,511	55
non contact	2,262	7	2,517	8	2,553	9	3,889	7
refusal to office	1,491	5	1,692	5	1,262	4	3,805	7
refusal to interviewer	5,495	18	6,233	19	4,500	15	15,937	29
other non response	1,236	4	1,152	4	1,101	4	1,770	3

Source: Wealth and Assets survey

Response reflects issued sample; figures do not include second households resulting from movers identified at a follow up interview

Thus, of the eligible households in wave 2, an interview was achieved with over two-thirds while no interview took place with just under one-third. The non-contact rate at wave 2 (9%) was slightly above that observed at wave 1 (7%). However, the refusal rate was considerably higher in wave 1 than in wave 2, in part because hard refusals from wave 1 were not followed up for wave 2.

For wave 3, cooperating households, non-contacts and circumstantial refusals from wave 2 were followed up. In addition, a new panel of households was selected for wave 3 in order to achieve a target of at least 20,000 household interviews. These new panel cases are included in the total figures for wave 3 in Table 1. The wave 3 response rate was 64%; 51% for the new cohort and 72% for the old cohort.

For wave 4, cooperating households, non-contacts and circumstantial refusals from wave 3 were followed up. In addition, a new panel of households was selected for wave 4 in order to achieve a target of at least 20,000 household interviews. These new panel cases are included in the total figures for wave 4 in Table 1. The wave 4 response rate was 66%; 53% for the new cohort and 70% for the old cohort.

The cross-sectional results presented in the Wealth in Great Britain report are based on all those households which responded in the particular wave in question, while the longitudinal analyses will

use only those individuals which responded in all waves, or at least one previous wave. The extent of longitudinal analysis undertaken is clarified in each case.

Data editing and validation

Cross-sectional editing and validation processes for waves 2, 3 and 4 were similar to those used for wave 1: more details are provided in section 10.4 of the wave 1 report[6]. However, collecting data from the same households over time provides an opportunity to conduct longitudinal edit checks. For example, if the recorded property value was similar in waves 2 and 4, but recorded as a very different figure in wave 3, perhaps due to a data entry error. In such circumstances, the wave 3 property value has been retrospectively edited to be more consistent with values recorded in the other waves. Generally, only values in waves 3 were edited.

Before any longitudinal checks could be carried out on the data, the longitudinally-linked records were checked for accuracy. The handling of adding new household members to households that responded in the previous wave, Original Sample Members (OSM) who left a household to be interviewed at their new address, or whole households who moved between waves, added complications to the linking exercise that deserved particular attention when the linkage checks were carried out.

To account for changes of circumstances within households that may impact on the observed wealth, indicator variables were produced to highlight circumstances such as a change of the Household Reference Person (HRP), additional household members, split households, and movers between waves.

Through this process changes between waves were observed that required further investigation. Thorough checking highlighted that the large majority of observed changes were genuine and could be explained through changes of circumstances for some or all individuals in the household, or where there was no evidence to indicate that collected data would be incorrect. However, these longitudinal checks also identified inconsistencies in the longitudinal data which were explained by errors occurring during the interview. These errors were amended where it was possible to establish the correct values.

Outliers exist in WAS data. These reflect the highly skewed nature of wealth. All outliers were checked for supporting evidence from interviewers. Where appropriate, edits were made to 'correct' outliers. In many cases, interviewer notes supported the validity of outliers and these remain in the WAS datasets. Given the skewed nature of wealth data, and the impact that outliers can have on parametric estimates, Wealth in Great Britain July 2012 to June 2014 does not report on any mean values. Mean values, particularly when exploring change across waves, can lead to the reporting of spurious change with the inclusion of extreme outliers. For this reason, all wealth estimates are reported on using median and/or deciles for Wealth in Great Britain July 2012 to June 2014.

[6] *Wealth in Great Britain: Main Results from the Wealth and Assets Survey 2006–08*. Published 10 December 2009.

Imputation: A rolling four wave strategy

General Aims and Methodology

In a way similar to all social surveys, the Wealth and Assets (WAS) survey data contained missing values. Typically, missing values are associated with non-response. Non-response can occur at household level, person level, and item level. The WAS imputation strategy was concerned primarily with item non-response. Item non-response relates to an event where a respondent does not know or refuses to answer a particular survey question. This can impact on estimates derived from WAS data in two ways:

- the missing data can lead to a reduction in the precision of the estimates
- if the characteristics of non-respondents differ from respondents, estimates may be biased

The general aim of the WAS imputation strategy was to counter these risks by estimating accurately the statistical properties of the missing data.

To meet this aim, missing values in the WAS data were imputed using Nearest-Neighbour/Minimum-Change methodology implemented in CANCEIS. CANCEIS is a widely recognised software platform containing a range of integrated imputation techniques (Bankier, Lachance, Poirier 1999; Canceis, 2009). The CANCEIS imputation algorithm employs a donor-based strategy designed to identify and replace missing values with observed values drawn from another record. The donor is selected from a small pool of potential donors with similar characteristics as the record currently being imputed. Similarity is measured by the sum of statistical distances between record and donor across a set of key demographic and other matching variables (MVs). The distance for each individual MV is weighted according to how well it might serve in predicting a valid and plausible range of imputable values in relation to the characteristics of the record currently being imputed. The MVs and associated weights for each WAS variable were identified through statistical modelling and expert review.

The general methodological WAS imputation strategy has several advantages:

- as a non-parametric approach, it avoids the distributional assumptions associated with other methods, facilitating preservation of important properties of the data such as skew and discrete steps in observed distribution functions
- the donor pool also serves as an implicit distributional model of the plausible range of values for each individual imputable record rendering the probability of selecting a particular value proportional to that distribution

These advantages serve to improve precision and reduce bias in point and variance estimates based on the WAS data, contributing to the accuracy of published statistical outputs (Durrent, 2005).

Tuning the General Methodology to the Analytical Aims of the Survey

While the general methodological WAS imputation strategy serves to improve the accuracy of estimates based on WAS data, tuning this strategy to the analytical aims of the survey further improves performance. As a panel survey in its fourth wave, the overarching analytical aims of the survey were fourfold. To provide:

- revised cross-sectional estimates based on the wave 3 data
- cross-sectional estimates based on the wave 4 data
- longitudinal estimates of change over time between waves 3 and 4
- longitudinal estimates of change over time for entire survey duration – here waves 1 to 4

To facilitate these aims, the imputation strategy was divided into three imputation groups (iGroups). Figure 1 outlines the fundamental structure of a variable's data within iGroup.

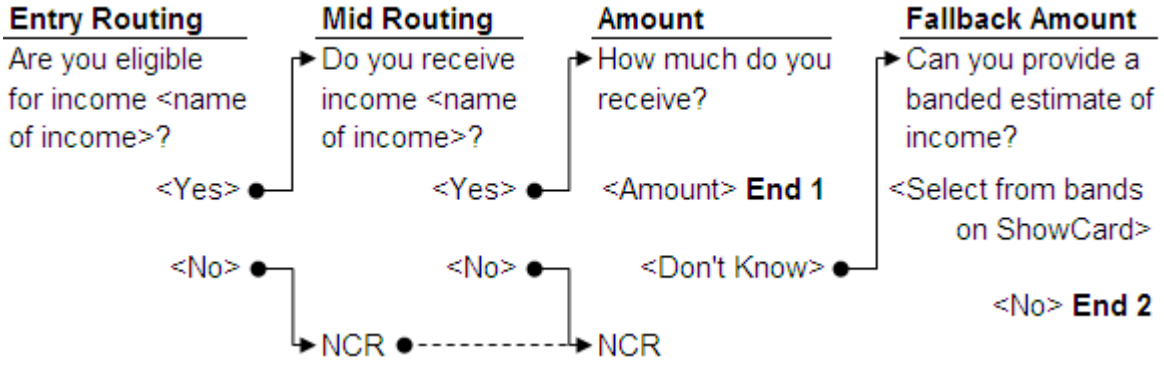
Figure 1: Discrete variable data structure within imputation group

Wave 3	Wave 4	iCode	iGroup
Not in Survey or Observed but NCR*	Observed Missing	n,O n,X	Wave 4 cross sectional
Observed Previously Imputed Observed Previously Imputed	Observed Missing Missing Observed	O,O X(i),X O,X X(i),O	Wave 3 & 4 longitudinal
Observed Previously Imputed	Not in Survey or Observed but NCR*	O,n X(i),n	Wave 3 cross sectional

*NCR: In survey but No Code Required for the current question.

A more specific analytical aim of the WAS survey is to provide estimates of wealth across five key topic areas: Property, Physical, Pensions, Financial, and Income. To facilitate this aim, the imputation strategy was also aligned with the routing structure of the WAS questionnaire. Although there are many variations, Figure 2 represents a schematic overview of the typical structure underlying a question group designed to elicit information about a particular facet of one of the topic areas.

Figure 2: Schematic of typical underlying structure of a WAS question group*



*Note that the WAS routing was often far more complex than outlined here, with many more layers and pathways

Wave 3 and Wave 4 cross sectional imputation

In general, the processing strategy for the Wave 3 and Wave 4 cross sectional data was the same. For every record in the data, entry and mid level routing variables were imputed sequentially employing the donor-based strategy outlined in the General Aims and Methodology Section. At any point throughout the routing where an imputed value indicated subsequent variables did not require a response, a NCR indicator was rolled forward through to the end of the question group. This maintained the integrity of the routing, excluded records from further processing, and ensured ultimately, that amounts related to the variable in question were only imputed for an appropriately sized sub-population.

Two additional constraints were imposed on the imputation of amounts at the end of the question group. If a banded estimate was observed, an imputed amount had to fall within that band. Extreme outliers were also excluded from the donor pool. Extreme outliers were generally defined as values greater than two times the threshold of the highest band for the variable in question. Typically, between 0.5% and 2.5% of observed data values were outliers. These additional constraints ensured that imputed values were selected based on all available information and that estimates based on the WAS data were not inappropriately biased by a few relatively unique observations.

To facilitate the general aim of improving precision and reducing bias in point and variance estimates based on the WAS data, the pool of potential donors was maximised by including all observed data associated with a particular wave. For example, when imputing missing values in the cross-sectional imputation group for Wave 4, the potential donor pool included valid observations from both the cross sectional and longitudinal Wave 4 imputation groups.

Although revision of the WAS Wave 2 data was out of scope, the longitudinal aspect of the imputation strategy was extended by using Wave 2 data as auxiliary information on which to further condition the imputation where is was available.

Wave 3 and Wave 4 longitudinal imputation

Fundamentally, imputation of the Wave 3 and Wave 4 longitudinal data followed the same sequential processing and roll forward NCR strategy as the cross sectional data. However, there were several distinct differences reflecting the change of emphasis in the primary analytical aims of the WAS survey from the provision of cross-sectional estimates to the provision of longitudinal estimates of change over time.

To facilitate the longitudinal aims of the survey the Wave 3 and Wave 4 data were imputed simultaneously. Prior to imputation, in cases where new observed information was available in Wave 4, data previously imputed in Wave 3 was reset to missing as the new Wave 4 information would be used to revise imputed data with improved precision. While imputing entry and mid level routing variables, in cases where Wave 3 was missing and Wave 4 was observed, or visa versa, the observed variable was included in the MV set in addition to any Wave 2 data available and the rest of the MV set. Typically, this variable was given a much higher weight than other MVs ensuring that donor selection was constrained more by observed longitudinal data than other MVs. If routing

variables were missing in both waves, imputed values were drawn from a single donor as this implicitly maintains appropriate longitudinal relationships in the data.

When imputing Wave 3 and Wave 4 amounts, both waves were constrained by the same principles as in the cross sectional strategy; values had to be imputed within observed banded estimates and extreme values were excluded from the donor pool. For the longitudinal data, control over extreme outliers had to be extended. In cases where an extreme value was observed in one wave but missing in the other, the record was excluded and edited manually through expert review. As the number of potential donors would always be extremely limited in this situation, this strategy removed any possibility of imputing an unjustified financial collapse or windfall for an individual respondent in the domain currently being imputed.

To maintain focus on the longitudinal aspects of the data, in cases where amounts were missing in both waves, values were again drawn from a single donor in order to maintain appropriate relationships in the data between waves. To achieve a similar aim in cases where a value was missing in one wave but observed in the other, the standard imputation strategy of drawing values directly from a donor was replaced with a more appropriate ratio imputation. For all potential donors, the ratio between observed amounts in Wave 3 and Wave 4 was calculated and set up as a derived variable and this variable was drawn from the donor instead of the value. The imputed ratio was used to either adjust and roll forward an observed value in Wave 3 or adjust and roll backward an observed value in Wave 4.

Significant variations and extensions

End to end, more than 1,000 variables were treated through the WAS imputation strategy. While the current report provides an overview of the primary design and reasoning behind the principal aspects of that design, it is important to note that almost every variable required some unique adjustment to the micro-parameters of the system. The following list outlines just a few of the most significant but necessary macro-variations and extensions to the base-line strategy:

Multi-tick routing

In many of the WAS entry level routing questions respondents were asked if they held one or more of a number of different but interdependent assets displayed on a show-card. To account for this interdependence and avoid the imputation of unreasonable relationships between asset types, multi-tick routing was imputed simultaneously as a binary string set. Potential donors were identified through a user defined distance matrix.

Multiple amounts & diminishing subpopulations

For some assets, mid level routing asked respondents to specify how many iterations of a particular asset they owned, for example, how many mortgages or private pension schemes? This can be problematic because the subpopulation of respondents becomes smaller as the number of iterations increase, leading to extremely impoverished donor pools and rendering imputation inappropriate. In such cases, observed values from earlier iterations were included in the potential donor pool and the position in the sequence of iterations was included as a weighted MV.

Gross and Net

Unlike most assets in the WAS survey, for a few, such as income from earnings, respondents were asked for a Gross and a Net. The imputation of Gross and Net can be quite complex, particularly when applied to longitudinal data which harbours within wave relationships nested inside relationships between waves. To account for these relationships, the imputation strategy for Gross and Net differed somewhat from the standard cross sectional and longitudinal strategies outlined previously.

In the event that the observed data did not provide any information about longitudinal relationship, the ratio-based roll forward/roll backward strategy outlined in the Wave 3 and Wave 4 Longitudinal Imputation Section was implemented in the first instance. The strategy was applied to impute either the Net to Net or the Gross to Gross relationship between waves depending on which had at least one observation in Wave 3 or Wave 4 to work with. When both Net and Gross were available to facilitate the longitudinal component of the imputation, Net was selected due to higher response rates for this variable.

Once relationships between waves were resolved, Net to Gross relationships within Wave 3 and Wave4 were imputed independently, also using the ratio-based roll forward/roll backward strategy outlined in the Wave 3 and Wave 4 Longitudinal Imputation Section.

Asset and Debt

In general, a ratio-based imputation strategy similar to that applied to Gross and Net was also applied to other corresponding asset & debt variables in the data. This served to ensure that the relationship between these variables was maintained appropriately.

Changes in the questionnaire or the structure of a question group

Changes to the questionnaire or structure of a particular question group can represent difficulties for a longitudinal imputation leading potentially to an inappropriate MV set that can bias results. Changes to the WAS questionnaire were addressed according to a list of strategic priorities: Where possible, Wave 3 variables were harmonised with those in Wave 4 and imputed according to the base-line strategy. If a Wave 3 variable structure was similar to that in Wave 4 but could not be harmonised, previously imputed Wave 3 data was not revised but where relevant, it was included in the Wave 3 MV set. Where Wave 3 variable structures were completely incompatible with Wave 4, a cross sectional imputation strategy was applied to the Wave 3 missing data only.

Market (time) sensitive assets

Through previous research and expert review it had been recognised that some of the assets addressed by the WAS questionnaire such as personal pensions, can be extremely sensitive to ongoing temporal changes in market forces. For these variables, the month the interview was conducted was included in the weighted MV set.

Quality Assurance and Evaluation

Quality assurance and evaluation of the WAS imputation strategy was a three-stage process conducted at different times throughout processing. Typically, assessment was based on analytical results derived through custom software designed in SPSS or SAS and on expert review from domain and topic experts. To ensure thoroughness, three teams were involved in the quality assurance process: Survey Methodology; Collection and Production; and Analysis and Dissemination.

Stage 1: As the efficacy of any imputation method depends of the quality of the input data, prior to imputation the WAS data was examined against a well defined set of imputation specifications. The specifications included a detailed data dictionary, a comprehensive outline of all routing architecture, approved MV sets, and additional notes on expected exceptions and outliers.

Stage 2: On a variable by variable basis throughout processing, the statistical properties of the imputed data were evaluated and compared to those of the observed data. This served to ensure that the imputation process itself did not introduce unwarranted bias into the cross sectional and longitudinal properties of the variable currently being imputed.

Stage 3: Following imputation, further analyses and review evaluated the impact of the imputed data in the calculation and derivation of estimates based on the WAS data. This served to ensure that the imputation strategy did not introduce unwarranted bias or have unnecessary impact on those estimates and thus, on published outputs.

Weighting

Overview

From wave 3 onwards, three sets of weights were created for use with the datasets from each wave: (i) a longitudinal weight for survivors (Wave 1 – Wave T), (ii) a longitudinal weight for the last two consecutive waves (Wave T-1 – Wave T) and (iii) a pseudo cross-sectional weight (Wave T). It's important to ensure that each set of weights is used for analysis of the relevant subsample of respondents. The weights incorporate adjustments for non-response and differential sampling probabilities (Daffin et al., 2009) and also adjust for loss to follow-up (LTFU) at following waves.

Calculation of the Weights

The wave 1 weights were constructed in three stages: first as the reciprocal of the selection probability; secondly adjusted for non-response; and finally calibrated to population totals using an age by sex and regional breakdown (Daffin et al., 2009). 'Integrative calibration' was used which ensures that each person in the household has the same weight; this is also the household weight. At each wave T, the Wave T-1 weight is brought forward to use as the basis of the Wave T base weight. The base weight tracks the progress through the survey of all people enumerated in the household, i.e. includes children and young adults who are deliberately not interviewed for the survey. WAS weights are calculated for all people enumerated in the household.

LTFU occurs through two processes. One process is where eligible people from Wave T-1 cannot be traced for their Wave T interview and, therefore, their eligibility status for the Wave T interview is

unknown. The second process is to adjust for those participants who decide not to take part in the survey between waves.

The cases with unknown eligibility will, in reality, have included both eligible and ineligible cases. A weight is constructed to adjust for unknown eligibility using a weighted binomial regression of known/unknown eligibility status onto a suite of socio-demographic characteristics measured at Wave T-1. The reciprocal of the propensity for known eligibility was used to adjust the Wave T-1 weight by multiplying the Wave T-1 weight through by the eligibility adjustment weight (w_{Tk}^{e}). The resulting weight was then used in a binomial regression of response/non-response status onto a suite of characteristics to adjust for the second stage of LTFU (response attrition). The reciprocal of the response propensity (w_{Tk}^{nr}) was used to adjust further the previous weight.

In summary, the Wave T longitudinal pre-calibrated weight (w_{Tk}^{long}) can be written as (1) below for respondents:

$$w_{Tk}^{long} = w_{(T-1)k}^{cal} w_{Tk}^{e} w_{Tk}^{nr},\ k \in s_T^r \qquad (1)$$

The weight is the product of three quantities, i.e. the Wave T-1 weight ($w_{(T-1)k}^{cal}$) adjusted for those cases moving into unknown eligibility (w_{Tk}^{e}) and non-response (w_{Tk}^{nr}) at Wave T. This weight is defined over the set of (s_T^r) longitudinal respondents at Wave T.

A second group of people included in the construction of the base weight are those people who became ineligible at Wave 2 (s_T^{ie}), described in (2). Typically, this group predominantly comprises those people who have left the population through death, migration or institutionalisation.

$$w_{Tk}^{long} = w_{(T-1)k}^{cal} w_{Tk}^{e},\ k \in s_T^{ie} \qquad (2)$$

Taking the two sets s_T^r and s_T^{ie} together should recover the population prior to LTFU, assuming complete correction for the LTFU processes.

A longitudinal calibration weight ($w_T^{long\_cal}$) was constructed from a trimmed version of the longitudinal pre-calibrated weight by calibrating the combined sub-sets of cases (s_T^r and s_T^{ie}) to the relevant population totals. For the weights of the survivors of all waves the relevant calibration population total are Wave 1; for the (T-1) to T longitudinal weights the relevant population totals are from Wave T-1.

$$w_T^{long\_cal} = w_T^{long} w_T^{g} \qquad (3)$$

The g-weight (w_T^g) ensures that the sums of the calibration control variables (age by sex and region) match those of the relevant population.

A pseudo-cross-sectional weight at Wave t is constructed differently for each subgroup in the sample. Firstly consider the terminology used to describe the subgroups:

- OSM – an Original Sample Member which refers to an individual who responded in the same wave that they were sampled.
- EOSM – an Entrant Original Sample Member which refers to an individual who lives at an address which was sampled but the household did not respond until a later wave.
- SSM – a Secondary Sample Member which refers to an individual who joined a previously responding household.

There are also new panels added from wave 3 onwards, as well as different combinations of response and non-response of sample members over waves to consider when calculating the cross sectional weights.

Any responder who has been in a previous wave will have their wave t longitudinal weight as a base weight. The first challenge for the cross-sectional weight is to assign a weight to people entering the sample. SSMs and births receive a cross sectional weight through a process of weight sharing the base weight of the OSMs. Rather than attempt to work out selection probabilities directly, it is common to use a weight share method to approximate these probabilities (e.g. Huang 1984, Ernst 1989, Kalton & Brick 1995).

A standard approach is to assign weight shares based on Wave T-1 household members to people in target Wave T households. A variety of weight share algorithms exist (e.g. Rendtel & Harms 2009). Following Kalton & Brick (1995), the weight w_i at time T for household i can be defined as the sum of the product of the initial weights w'_{jk} and a constant α_{ijk} summed over the k individuals in households j at time T-1:

$$w_i = \sum_j \sum_k \alpha_{ijk} w'_{jk} \qquad (4)$$

The constant (α_{ijk}) is defined in terms of the number of people in household i at time T who were in the population at time T-1.

$$\alpha_{ijk} = \begin{cases} 1/N_i & \text{if individual } k \text{ lives in household i at T and was in population T-1} \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

Finally, the weight w_i is assigned to all *k* household members of household *i*. In this scheme a population entrant at wave T is ascribed a zero contribution to α_{ijk} and a zero initial weight (w'_{ijk}). However, a sample entrant who was not in the population at wave T-1 but only in the sample at wave T contributes to α_{ijk} but has a zero initial weight. Consequently, sample entrants in the population do not increase the sum of the weights; whereas population entrants do increase the sum of the population weights. This is the fair share method of Huang (1984) and also the weight

share method of Ernst (1989). Unfortunately, with the WAS data it is not possible to determine if an entrant was in the Wave T-1 population or not, except for births, therefore all entrants other than births are treated as former Wave T-1 population members.

The EOSMs weights are their original design weights, constructed as the inverse of the selection probabilities. These are then rescaled to account for the proportion of the original responding sample that they represent, multiplied by the relevant population total. Additionally a small adjustment is made to OSMs and SSMs to counteract previous non-response adjustments made for these cases.

Wave T pre-calibration cross-sectional weights were obtained by averaging the person level pseudo-cross-sectional weights within each household.

The new panel weights were constructed firstly as the reciprocal of the selection probability, followed by a non response adjustment as with the original panel sample in wave 1.

The pseudo-cross-sectional Wave T weights are created through integrative calibration of the pre-calibration weight to the Wave T population totals (6). This is carried out for each panel separately to allow for analysis of each panel if required.

Wave T non-proxy pseudo-cross-sectional weights were computed at person level by calibrating the person level pseudo-cross-sectional weights of non-proxy cases.

$$W_{Tik}^{xs\_cal} = {}^{pre-cal}W_{Ti}^{xs} W_{Ti}^{g} \qquad (6)$$

The final stage is to combine the different panels together; the chosen method combines the panels in proportion to the effective sample size (as proposed by Chu et al 1999, Korn and Graubard 1999). This accounts for the variance within each panel and combines the weights such that the overall variance is minimised. As a result, the newer panel(s) weights will be scaled up whilst the older panel(s) will be scaled down.

Calibration Control Groups

Age x Sex Weighting Classes	
0-9 Male	0-9 Female
10-15 Male	10-15 Female
16-24 Male	16-24 Female
25-34 Male	25-34 Female
35-39 Male	35-39 Female
40-44 Male	40-44 Female
45-49 Male	45-49 Female
50-54 Male	50-54 Female
55-59 Male	55-59 Female
60-64 Male	60-64 Female
65-74 Male	65-74 Female
75+ Male	75+ Female

Region Government Office Region
North East
North West
Yorkshire & Humberside
East Midlands
West Midlands
East
London
South East
South West
Wales
Scotland

Standard errors and estimates of precision

One measure of sampling variability is the standard error. Standard errors are one of the key measures of survey quality, showing 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. WAS has a complex design that 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.

Typically, PSUs tend to be characterised by a positive intra-class correlation coefficient, that is people within a PSU are more alike to each other than they are to people in the rest of the sample. This acts to increase the standard error of an estimate relative to simple random sampling. Conversely, stratification can act to decrease the standard error if people within a stratum are

relatively homogeneous and there is consequently a greater degree of heterogeneity between strata. Both these elements of the design should be accounted for when calculating standard errors.

An identifier of the PSU is included on the WAS dataset. Selection of the PSUs was done by ordering the frame. The first ordering principle was geographic (region x district); whereas the second was socio-demographic, that is within each of the 26 regional districts further ordering was done on the basis of the socio-demographic characteristics of the PSU populace. This ordering fulfils two purposes. Firstly it spreads out the sample in terms of socio-demographic characteristics ensuring people from higher and lower ends of the socio-demographic dimensions were included in the sample. Secondly, it enables stratification. The primary stratification variable, the 26 regional districts, was identified on the dataset but because of the way the sample was selected from the ordered frame it can be regarded as a design selecting a single PSU per stratum. Consequently, it was possible to incorporate a much finer stratification procedure using a 'collapsed stratum' approach.

Finally, the calibration to population totals needs to be taken into account. This will have a beneficial effect, both in terms of adjusting for residual bias after non-response weighting and in reducing the variance of estimates. The extent to which the variance was reduced was related to the extent to which the survey variables were related to the variables in the calibration. The calibration variables were household counts of people within each age group by sex and regional category, so it was to be expected that, for example, the total wealth of a household will be associated with these variables.

The method for taking account of the calibration when calculating standard errors is described in the report 'Variance estimation for Labour Force Survey Estimates of Level and Change', GSS Methodology Series no. 21, Holmes and Skinner.

To enable the reader to gain an appreciation of the variability of the results presented in this report, estimates of the standard errors of some key variables have been produced.

The estimates in this report are based on information obtained from a sample of the population and are therefore subject to sampling variability. 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.

Table 2
Standard errors for totals: Great Britain, July 2006 to June 2014

Wave 4: July 2012 to June 2014

	Estimate of Total (£Billion)	Standard Error (£Billion)
Private pension wealth	4,459	64
Financial Wealth	1,596	125
Physical Wealth	1,169	11

Property Wealth	3,927	66
Total Wealth (excluding private pension wealth)	6,676	156
Total Wealth (including private pension wealth)	11,134	194

Wave 3: July 2010 to June 2012

	Estimate of Total (£Billion)	Standard Error (£Billion)
Private pension wealth	3,586	54
Financial Wealth	1,299	133
Physical Wealth	1,102	10
Property Wealth	3,528	48
Total Wealth (excluding private pension wealth)	5,929	168
Total Wealth (including private pension wealth)	9,515	191

Wave 2: July 2008 to June 2010

	Estimate of Total (£Billion)	Standard Error (£Billion)
Private pension wealth	3,459	81
Financial Wealth	1,085	26
Physical Wealth	1,016	10
Property Wealth	3,379	43
Total Wealth (excluding private pension wealth)	5,485	71
Total Wealth (including private pension wealth)	8,944	129

Wave 1: July 2006 to June 2008

	Estimate of Total (£Billion)	Standard Error (£Billion)
Private pension wealth	2,886	55
Financial Wealth	1,043	21
Physical Wealth	961	9
Property Wealth	3,537	42
Total Wealth (excluding private pension wealth)	5,569	82
Total Wealth (including private pension wealth)	8,462	117

1. Due to approximation methods in the calculation of the standard errors there are some differences between these estimates and those published.

Source: Wealth and Assets Survey, Office for National Statistics.

Table 3

Standard errors for medians: Great Britain, July 2006 to June 2014

Wave 4: July 2012 to June 2014

	Estimate of Median(£)	Standard Error(£)
Private pension wealth (excluding zeros)	97,300	1,800
Financial Wealth	5,900	200

Property Wealth (excluding zeros)	153,800	**2,600**
Total Wealth (including private pension wealth)	225,100	3,700
Total Wealth (excluding private pension wealth)	141,400	2,700

Wave 3: July 2010 to June 2012

	Estimate of Median(£)	Standard Error(£)
Private pension wealth (excluding zeros)	82,300	1,500
Financial Wealth	5,900	200
Property Wealth (excluding zeros)	150,000	100
Total Wealth (including private pension wealth)	218,400	3,600
Total Wealth (excluding private pension wealth)	146,200	2,500

Wave 2: July 2008 to June 2010

	Estimate of Median(£)	Standard Error(£)
Private pension wealth (excluding zeros)	71,600	1,400
Financial Wealth	6,400	200
Property Wealth (excluding zeros)	147,900	1,700
Total Wealth (including private pension wealth)	204,300	3,100
Total Wealth (excluding private pension wealth)	144,400	2,300

Wave 1: July 2006 to June 2008

	Estimate of Median(£)	Standard Error(£)
Private pension wealth (excluding zeros)	60,000	1,000
Financial Wealth	5,700	200
Property Wealth (excluding zeros)	150,000	600
Total Wealth (including private pension wealth)	196,700	3,000
Total Wealth (excluding private pension wealth)	146,600	2,200

1. Due to approximation methods in the calculation of the standard errors there are some differences between these estimates and those published.

Source: Wealth and Assets Survey, Office for National Statistics.

Table 4

Standard errors for change in totals: Great Britain, July 2006 to June 2014

Wave 3 to 4

	Wave 3 Estimate (£Billion)	Wave 4 Estimate (£Billion)	Estimate of Change (£Billion)	Standard Error (£Billion)
Private pension wealth	3,586	4,459	873	55
Financial Wealth	1,299	1,596	298	124
Physical Wealth	1,102	1,169	67	10
Total Wealth (excluding private pension wealth)	5,929	6,676	747	155

Total Wealth (including private pension wealth)	9,515	11,134	1,620	178
Property Wealth	3,528	3,927	399	48

Wave 2 to 4

	Wave 2 Estimate (£Billion)	Wave 4 Estimate (£Billion)	Estimate of Change (£Billion)	Standard Error (£Billion)
Private pension wealth	3,459	4,459	1,000	83
Financial Wealth	1,085	1,596	511	107
Physical Wealth	1,016	1,169	154	11
Total Wealth (excluding private pension wealth)	5,485	6,676	1,191	142
Total Wealth (including private pension wealth)	8,944	11,134	2,191	184
Property Wealth	3,379	3,927	549	52

Wave 1 to 4

	Wave 1 Estimate (£Billion)	Wave 4 Estimate (£Billion)	Estimate of Change (£Billion)	Standard Error (£Billion)
Private pension wealth	2,886	4,459	1,573	68
Financial Wealth	1,043	1,596	553	107
Physical Wealth	961	1,169	209	11
Total Wealth (excluding private pension wealth)	5,569	6,676	1,106	158
Total Wealth (including private pension wealth)	8,462	11,134	2,672	191
Property Wealth	3,537	3,927	391	52

1. Due to approximation methods in the calculation of the standard errors there are some differences between these estimates and those published.

Source: Wealth and Assets Survey, Office for National Statistics.

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 any collection whether it's 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 supporting documentation and processes were thoroughly tested before being finalised for use in the survey.

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. ONS uses training programs, the provision of detailed supporting documentation and regular supervision and checks of interviewers' work to achieve 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 every effort was made to reduce the level of non-response bias through careful survey design and compensation during the weighting process. To further reduce the level and impact of item non-response resulting from missing values for key items in the questionnaire, ONS undertook imputation prior to the release of the datasets for analysis.

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

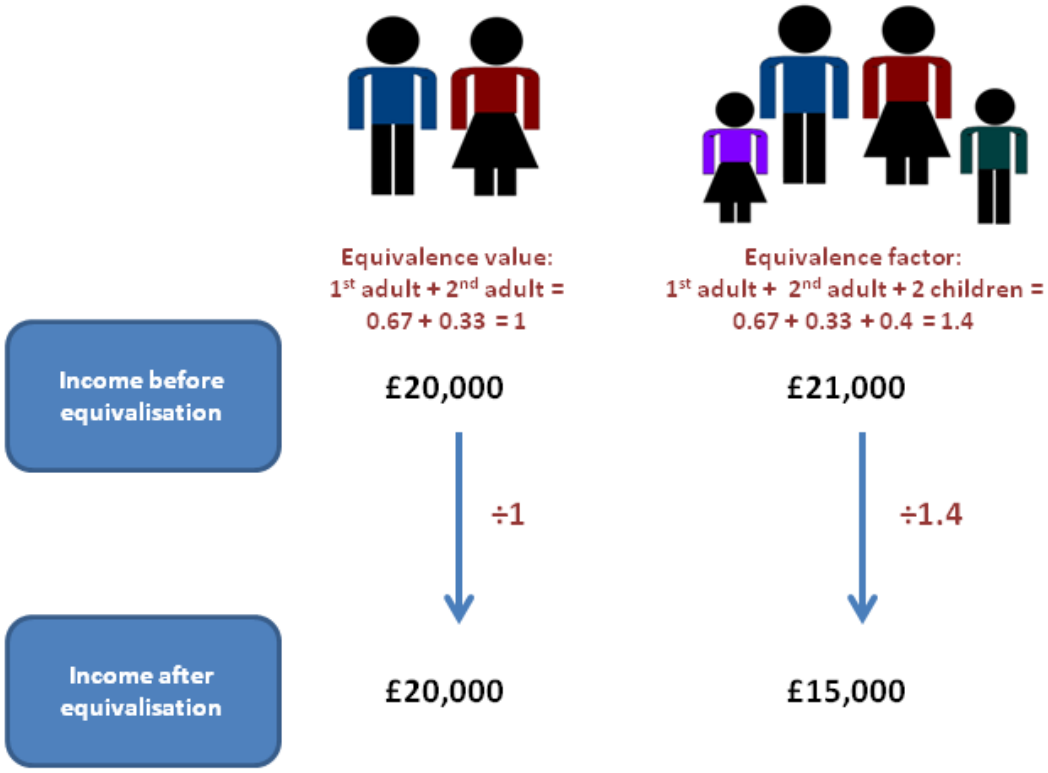
course of deriving variables, manipulating data or producing the weights. To minimise the likelihood of these errors occurring a number of quality assurance processes were employed.

Introduction to equivalisation

It is standard practice, in ONS and elsewhere, when presenting income measures at a household level to use an equivalised income estimate.

What is equivalised income?

Equivalisation is a process that adjusts households' incomes to take into account the number of people living in the household and their ages. For example a household with two people in it will need more money to sustain the same living standards as one with a single person; however the two person household is unlikely to need double the income. The modified-OECD equivalisation scale is used widely in the UK and elsewhere in the world and is the scale proposed to be used for income estimates from WAS.



What measure of income has been presented?

In the Wealth and Income article published using wave 3 data, only gross annual estimates of regular income are presented. Net estimates, which account for the effects of taxation, were not used and no equivalisation was applied.

The reason for this is that WAS shouldn't be considered as the prime source of income data. Income is only collected on WAS so that it can be used as a characteristic variable alongside estimates such as region and household structure. In addition income estimates could not be produced with the same definition as used elsewhere (e.g. net income after housing costs).

In Wealth in Great Britain 2012 to 2014 equivalised net income deciles by each of the wealth components is presented. Net income was used to be more representative off take home pay.

Should wealth be equivalised?

Wealth is a stock of assets that is accumulated over time. In many cases wealth is a resource considered to be available to support consumption in the future; especially during retirement (i.e. wealth is often built up during a person's working life and then run down during retirement). Therefore, when comparing households' wealth as an indicator of economic well-being in terms of potential future consumption, consideration needs to be given to which household members are likely to benefit from that wealth. Of particular interest are households containing children. The children are likely to leave the household before the wealth of the household is used to support household consumption during retirement. Therefore, for this type of analysis, it does not seem relevant to equivalise wealth in the same way as income.

However, when considering wealth as an economic resource that may be used to support current consumption the argument is different. If, for example, the analysis is using wealth to indicate whether a household is at risk of immediate economic hardship (some households have very low income but are not at risk of economic hardship because they can draw on their wealth to support current consumption) it is appropriate to equivalise wealth with the same equivalence scales used to standardise household income.

Whilst this may be the case for total wealth measures, when considering the components of wealth, this is definitely not the case. For example, the value of a property is exactly the same irrespective of the number of people living in that property. The value of a private pension can only be realised at retirement and therefore cannot be seen as a resource available to reduce the risk of economic hardship prior to this. The OECD (2013), therefore, does not recommend the equivalisation of the component measures of wealth.

Since Wealth in Great Britain considers in detail the components of wealth, the estimates contained in this report have not been equivalised.

The effect of inflation on wealth

All estimates within the Wealth in Great Britain 2012 to 2014 report are presented as current values (i.e. the value at time of interview) and have not been adjusted for inflation. Like equivalisation, deflating wealth estimates is not as straight forward as for other economic estimates.

Inflation means that a given amount of money at one time is worth less at another point of time in the future. One way in which inflation is measured is the Consumer Prices Index (CPI) which indexes the relative cost of consumer goods over time. In considering changing wealth over time, it might be expected that the effects of inflation are taken into account such that it can be assumed wealth at one time frame, can purchase the same at another. However, wealth as a stock concept in the way it is presented in this report, presents some challenges in adjusting to consider it on equal terms from one time period to the next.

For physical wealth, household goods are valued as the cost of new replacement goods at the time of interview. It would therefore be expected to reflect changes in cost of goods due to inflation.

Comparing property values over time on an equivalent basis would be best adjusted using an index of property values, such as the House Prices Index (HPI). Gross property wealth is valued based on how much the individual believes their property to be worth. This may not be an accurate reflection of the actual likely sale cost, for example if the owner does not keep track of sale prices of similar properties in their area. In such instances deflating using the House Prices Index would underestimate the value of property and give an unrepresentative view of change over time.

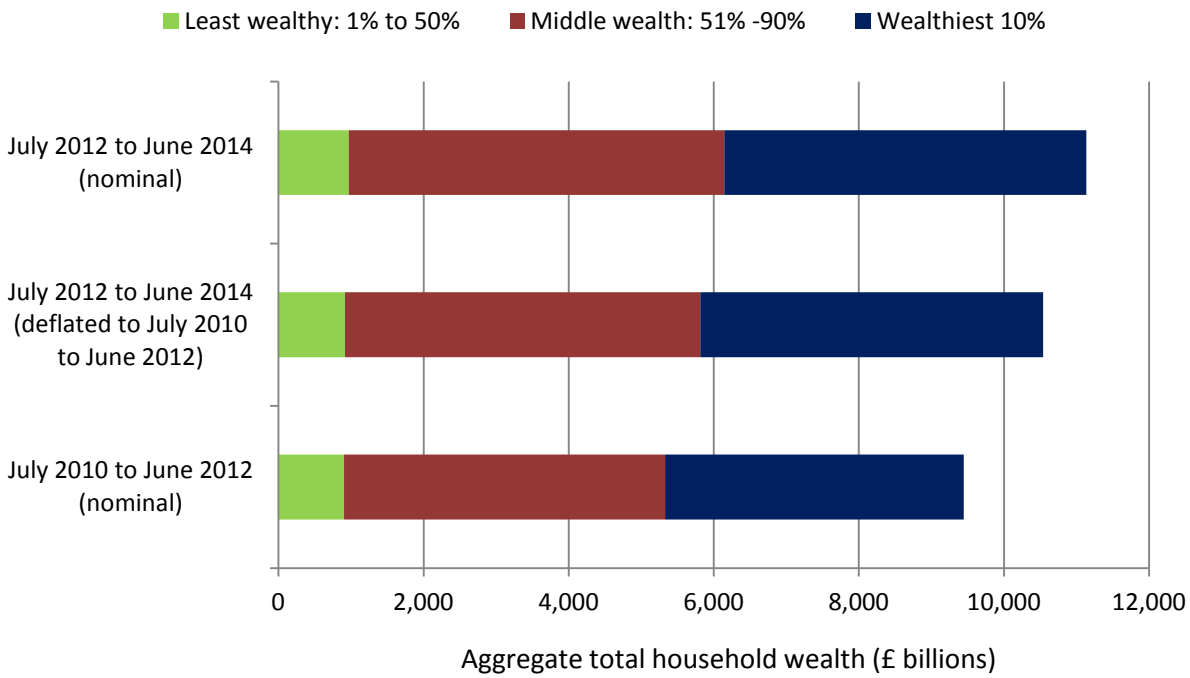
Defined Benefit Private Pension schemes take account of inflation in the calculation when converting expected income to a stock figure comparable with Defined Contribution Private Pensions. The value in a pension is in the income or lump sum it produces at the time of retirement. The eventual income or lump sum is dependent on future inflation up until the point of retirement which will differ for each individual dependent on their current age and actual future retirement age.

For financial wealth, deflation would tend to increase the value of financial liabilities and decrease the value of financial assets. No recommended deflator exists to apply to financial wealth.

Assessing general impact on aggregate wealth using Consumer Prices Index

To assess the relative impact of inflation on wealth between July 2010 to June 2012 and July 2012 to June 2014, aggregate total wealth figures for Wave 4 were deflated to their Wave 3 equivalent using the Consumer Prices Index (CPI). A mean index figure was calculated for each, between July 2010 to June 2012 and July 2012 to June 2014, based on monthly CPI figures. Figures were adjusted using these index figures.

Nominal (unadjusted) figures for July 2012 to June 2014, showed the least wealthy owned £969 billion of aggregate total wealth. Deflated using CPI to equivalent figures for July 2010 to June 2012 this becomes £917 billion. Much closer to the nominal (unadjusted) July 2010 to June 2012 figure of £904 billion. Note that as adjustments are based on aggregate figures, this does not take account of the effect of inflation on increasing financial liabilities. For the wealthiest 10% of households in July 2012 to June 2014 aggregate total household wealth of £4,990 billion is £4,722 when adjusted using CPI, compared with the figure for July 2010 to June 2012 of £4,115 billion.


Least wealthy: 1% to 50%
Middle wealth: 51% -90%
Wealthiest 10%
July 2012 to June 2014 (nominal)
July 2012 to June 2014 (deflated to July 2010 to June 2012)
July 2010 to June 2012 (nominal)
0
2,000
4,000
6,000
8,000
10,000
12,000
Aggregate total household wealth (£ billions)

References

Bankier, M., Lachance, M., & Poirer, P. (1999) A generic implementation of the new imputation methodology. Accessed online: [07/12/1011]
http://www.ssc.ca/survey/documents/SSC2000_M_Bankier.pdf

Canceis (2009). Users Guide V4.5. CANCEIS Development Team. Social Survey Methods Division, Statistics Canada

Felligi, I.P., & Holt, D. (1976) A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association,* Volume 71, No 353, 17-35

Daffin, C. (2009) *Wealth in Great Britain*. Newport: Office for National Statistics.

Durrent, G.B. (2005) Imputation methods for handling item non-response in the social sciences: A methodological review. ESRC Centre for Research Methods and Southampton Statistical Research Institute, NCRM Methods Review.

Ernst L (1989). Weighting issues for longitudinal household and family estimates. p135-159 in Kasprzyk D, Duncan G, Kalton G, Singh MP (eds.) *Panel Surveys*. John Wiley & Sons, New York.

Holmes, D. & Skinner, C (2000) *Variance Estimation for Labour Force Survey Estimates of Level and Change*. GSS Methodology Series no. 21

Huang H (1984). Obtaining cross-sectional estimates from a longitudinal survey: experiences of the Income Survey Development program. *Survey research Section of the Proceedings of the American Statistical Association,* 1984: p670-675.

Kalton G, Brick JM (1995). Weighting schemes for household panel surveys. *Survey Methodology* 21: p33-44.

OECD (2013), OECD Guidelines for Micro Statistics on Household Wealth, OECD Publishing. http://dx.doi.org/10.1787/9789264194878-en

Rendtel U, Harms T (2009). Weighting and calibration for household panels. p265-286 in Lynn, P. (ed.) *Methodology of longitudinal surveys*. Wiley, Chichester.