

HBAI 2024/25 Harmonised Dataset: User Guide

This guide provides users with information on the HBAI 2024/25 harmonised dataset and the associated resample datasets for producing confidence intervals.

The 2024/25 HBAI release will take place in two stages - in March 2026 and summer 2026. Please read Section A for more information.

Please download these new datasets each year as they will be revised for latest publication year prices, contain any additional variables as well as any improvements to data processing.

The contents are:

- A. HBAI 2024/25 Summary Information
- B. HBAI Estimates, Datasets and Documentation Available.
- C. Analysis Requirement.
- D. HBAI Income Variable Adjustments
- E. Merging Datasets.
- F. Grossing Factor Selection.
- G. Measuring Uncertainty in HBAI Estimates: Using the HBAI Resamples Datasets.
- H. Quality Assurance and Presenting Analysis.

Annex A: Worked Examples

Annex B: Converting code for the individual-level HBAI dataset

Annex C: Grossing factor for the family (benefit unit) level HBAI dataset

We are seeking your feedback on this edition of the data

Given the breadth of groups covered by the HBAI data, it has not been possible for DWP statisticians to assess or understand every breakdown which is of interest to external researchers and users. Therefore, we are inviting you to let us know of any insights you have relating to data quality or trends when analysing this data for your area of interest. This will help us as we begin to process and quality-assure the 2025/26 dataset. Please send any feedback direct to our HBAI Team Inbox: team.hbai@dwp.gov.uk

A. HBAI 2024/25 Summary Information

The Households Below Average Income (HBAI) dataset provides information on living standards in the United Kingdom based on household income measures for a financial year.

The HBAI dataset is on an individual level and uses the net, weekly income of their household. Breakdowns allow analysis of individual, family (benefit unit) and household characteristics of the individual.

Data is taken from the [Family Resources Survey \(FRS\)](#), whose focus is capturing information on incomes, as well as a lot of contextual information on household and individual's circumstances.

This dataset provides estimates for single financial years, from 1994/95 to 2001/02 for Great Britain and 2002/03 to 2024/25 for the United Kingdom.

The 2024/25 HBAI publication features an improved approach to using administrative data in place of FRS survey responses both for 2024/25 and some back-series years. Information on the major state benefits and tax credits is now based on administrative data (admin-linked) rather than survey responses. Revisions to the back-series will take place in two stages - in March 2026 (back to 2021/22) and summer 2026 (back to 2018/19).

For March 2026, the 2024/25 HBAI harmonised dataset consists of:

- admin-linked data from 2021/22 to 2024/25.
- unlinked data from 1994/95 to 2020/21.

From summer 2026, the 2024/25 HBAI harmonised dataset consists of:

- admin-linked data from 2018/19 to 2024/25.
- unlinked data from 1994/95 to 2017/18.

Section B explains how to use the data in more detail.

Please note:

- The improvement for the 2024/25 HBAI publication means key low-income measures, including number/proportion of people identified as being in low income for all groups and in all years from 2021/22 (in March 2026) and from 2018/19 (in summer 2026) have changed. It also means there will be a break in the HBAI series at these points. We advise users that income data before and after the break are not directly comparable and if comparisons across the break point are required users should follow the advice set out in section 3.2 of the [HBAI Background, Information and Methodology report](#). Some

measures of poverty, such as material deprivation statistics, are unaffected by the improvement of using administrative-linked data.

- Statistics published by Scotland, Wales and Northern Ireland have been classified as Official Statistics in Development. This signals that the changes introduce uncertainty that is acute for estimates below UK level which needs to be reflected in their Official Statistics reporting.
- From the 2024/25 HBAI publication, the fixed reference year for absolute low income has been updated from 2010/11 to 2024/25 for 4 years (2021/22 to 2024/25). Data for years prior to 2021/22 continue to use 2010/11 as the fixed reference year. From summer 2026, the fixed reference year will be 2024/25 for 2018/19 to 2020/21. See the [HBAI Background, Information and Methodology report](#) for more information.

While this dataset displays estimates for single financial years, **please ensure that HBAI estimates for regional, country and ethnicity estimates are calculated using three-year averages**. This method helps smooth erratic series and so give more robust estimates as the FRS sample size and coverage issues mean that single year results broken down below the level of UK region are unlikely to be reliable. For alternative sub-regional sources, please see [Annex 4: Other Relevant Statistics](#) of the HBAI Background, Information and Methodology report.

All rolling three-year averages calculated and published for any period including 2020/21 are based on the remaining two available data points only. Therefore:

- When calculating 2018/19 to 2020/21 estimates, please calculate a two-year average using 2018/19 and 2019/20 estimates.
- For the March 2026 release, estimates cannot be calculated for 2019/20 to 2021/22 as 2020/21 data is not available and 2019/20 and 2021/22 data are based on different methodologies.
- For the summer 2026 release, when calculating 2019/20 to 2021/22 estimates, please calculate a two-year average using 2019/20 and 2021/22 estimates.
- When calculating 2020/21 to 2022/23 estimates, please calculate a two-year average using 2021/22 and 2022/23 estimates.

A complete list of edits for the HBAI 2024/25 harmonised dataset can be found at the end of the HBAI_2425_Harmonised_Dataset_Variables_Guide (which accompanies this guide). Further information on methodological changes for HBAI 2024/25 can be found in Section 2 of the [HBAI Background, Information and Methodology \(BIM\) report](#).

B. HBAI Estimates, Datasets and Documentation Available

The following options are available for users to access HBAI estimates and underlying data:

- [Households below average income: for financial years ending 1995 to 2025](#), with the main report, tables and charts, as well as the HBAI Background, Information and Methodology (BIM) report.
- [HBAI Stat-Xplore Dataset](#), a free online tabulation tool with [user guide](#).
- [UK Data Service](#) with a suite of accompanying user guides
- DWP via '/secure/im/im3/project/hbai/released/gross4' and [SharePoint](#).

One individual level dataset has been published which includes the latest year, 2024/25, as well as the full back-series to 1994/95.

To allow users to analyse variables consistently over the time period, the dataset has been harmonised to provide:

- the same naming and definition of variables for each year (with values set to missing where information is unavailable or not applicable).
- removal of older variables that are no longer used in HBAI analysis to reduce confusion (such as variables based on the previous McClement's equivalisation scale).
- inclusion of additional variables used for the HBAI publication tables.
- all income and amount values presented in the latest, published CPI-based prices so that quicker, consistent comparisons can be made.

For UK Data Archive users, this dataset has been broken down into three-year sub-datasets that will need to be stacked together – this is to avoid download limits being breached (see Table B1 below for more information).

Associated individual level resamples datasets for each HBAI year are available for measuring uncertainty (confidence intervals) around HBAI estimates - see the 'Measuring Uncertainty' section for more information. Please note:

- for the 2024/25 release, as well as the resample dataset for 2024/25, the 2021/22 has been revised and republished – please see Section 2 of the [HBAI Background, Information and Methodology \(BIM\) report](#).
- to support limited analysis of unlinked data, main data back to 1994/95 in 2023/24 prices has been retained under the same study number.

Please download these new datasets each year as they will be revised for latest publication year prices, contain any additional variables as well as any improvements to data processing.

As the 2024/25 release will take place in two stages, users will be required to download data in summer 2026. Please see below for more information.

This guide should provide all the information needed to help produce HBAI estimates but please email the HBAI Team at team.hbai@dwp.gov.uk if you need further support.

Table B1: Summary of New HBAI Harmonised Datasets

DWP Users	UK Data Archive Safe Room Versions	UK Data Archive End User Licence Versions
Main Dataset: allind_2425prices	Main Datasets: i9497s_2425prices_unlinked i9700s_2425prices_unlinked i0003s_2425prices_unlinked i0306s_2425prices_unlinked i0609s_2425prices_unlinked i0912s_2425prices_unlinked i1215s_2425prices_unlinked i1518s_2425prices_unlinked i1821s_2425prices_unlinked i2124s_2425prices_admin_linked i2425s_2425prices_admin_linked In summer 2026: i1821s_2425prices_admin_linked	Main Datasets: i9497e_2425prices_unlinked i9700e_2425prices_unlinked i0003e_2425prices_unlinked i0306e_2425prices_unlinked i0609e_2425prices_unlinked i0912e_2425prices_unlinked i1215e_2425prices_unlinked i1518e_2425prices_unlinked i1821e_2425prices_unlinked i2124e_2425prices_admin_linked i2425e_2425prices_admin_linked In summer 2026: i1821s_2425prices_admin_linked
Resamples Datasets: ind9495_res to ind2425_res	Resamples Datasets: i9495s_res to i2425s_res	Resamples Datasets: i9495e_res to i2425e_res

Please note that the resample datasets, i2122s_res and i2122e_res, have been revised.

The HBAI datasets are available in two versions on the [UK Data Service](#).

- **End User (or Anonymised) Datasets:**

These datasets contain no names, addresses, telephone numbers, bank account details, NINOs or any personal details that can be considered as being disclosive under the terms of the '[ONS Disclosure Control](#)'.

Additional changes to the datasets are as follows:

- Remove households containing 10 or more individuals.
- Reset individuals aged over 80 to aged 80 and change ageband=16 (aged 85 and over) to ageband=15 (80 to 84).
- Round income variables to nearest £1.
- Remove variables for council tax liability (ctliab), pensioner flags for the individual (penflag), the head (penflaghd) and the spouse (penflagsp), inner/outer London flag (london) and the number of food parcels received from food bank in past 12 months (foodbk12) and past 30 days (foodbk30).

Safe Room Datasets: all of the HBAI data is available without the above restrictions (aside from removing the number of food parcels received from food bank in past 12 months (foodbk12) and past 30 days (foodbk30)).

To avoid download limit issues, the Data Archive datasets are published in three-year sub-datasets (the latest subset may contain less than three years) that will need to be merged together to build the main dataset. Table B2 and B3 provide examples of how to do this using SAS Eguide code for the March 2026 and summer 2026 versions.

Table B2: March 2026 All Years Individual Level Dataset

UK Data Archive Safe Room Version	UK Data Archive End User Licence Version
data allind_2425prices; set i9497s_2425prices_unlinked i9700s_2425prices_unlinked i0003s_2425prices_unlinked i0306s_2425prices_unlinked i0609s_2425prices_unlinked i0912s_2425prices_unlinked i1215s_2425prices_unlinked i1518s_2425prices_unlinked i1821s_2425prices_unlinked i2124s_2425prices_admin_linked i2425s_2425prices_admin_linked; run;	data allind_2425prices; set i9497e_2425prices_unlinked i9700e_2425prices_unlinked i0003e_2425prices_unlinked i0306e_2425prices_unlinked i0609e_2425prices_unlinked i0912e_2425prices_unlinked i1215e_2425prices_unlinked i1518e_2425prices_unlinked i1821e_2425prices_unlinked i2124e_2425prices_admin_linked i2425e_2425prices_admin_linked; run;

As presented in table B2, the March 2026 HBAI harmonised dataset should contain unlinked data from 1994/95 to 2020/21 and admin-linked data from 2021/22 to 2024/25. The use of the unlinked 2018/19 to 2020/21 data is highlighted in blue.

Table B3: Summer 2026 All Years Individual Level Dataset

UK Data Archive Safe Room Version	UK Data Archive End User Licence Version
data allind_2425prices; set i9497s_2425prices_unlinked i9700s_2425prices_unlinked i0003s_2425prices_unlinked i0306s_2425prices_unlinked i0609s_2425prices_unlinked i0912s_2425prices_unlinked i1215s_2425prices_unlinked i1518s_2425prices_unlinked i1821s_2425prices_admin_linked i2124s_2425prices_admin_linked i2425s_2425prices_admin_linked; run;	data allind_2425prices; set i9497e_2425prices_unlinked i9700e_2425prices_unlinked i0003e_2425prices_unlinked i0306e_2425prices_unlinked i0609e_2425prices_unlinked i0912e_2425prices_unlinked i1215e_2425prices_unlinked i1518e_2425prices_unlinked i1821e_2425prices_admin_linked i2124e_2425prices_admin_linked i2425e_2425prices_admin_linked; run;

As presented in table B3, the summer 2026 HBAI harmonised dataset should contain unlinked data from 1994/95 to 2017/18 and admin-linked data from 2018/19 to 2024/25. The use of the admin-linked 2018/19 to 2020/21 data is highlighted in blue.

Please note:

- The improvement for the 2024/25 HBAI publication means key low-income measures, including number/proportion of people identified as being in low income for all groups and in all years from 2021/22 (in March 2026) and from 2018/19 (in summer 2026) have changed. It also means there will be a break in the HBAI series at these points. **We advise users that income data before (unlinked) and after (admin-linked) the break are not directly comparable and if comparisons across the break point are required users should follow the advice set out in section 3.2 of the on the [HBAI Background, Information and Methodology \(BIM\) report](#).** Some measures of poverty, such as material deprivation statistics, are unaffected by the improvement of using administrative-linked data.

For DWP users, the all years' individual level dataset, allind_2425prices, and separate HBAI year individual level resamples datasets are available at: [/secure/im/im3/project/hbai/released/gross4](#)

A comprehensive variable guide is available along with a record of the variables that have been removed, added or altered compared to the previous HBAI year datasets stored at and on [SharePoint](#) in Edge (DWP users only). More detailed variable derivation documentation has also been provided.

Please contact the HBAI team at team.hbai@dwp.gov.uk if you have any feedback or suggestions for the documentation.

C. Analysis Requirement

Users are encouraged to consider the following questions when assessing the analysis required and quality-assurance of results:

Table C1: Analysis Requirement Questions

Analysis already available?	- Is the analysis already available in the published report and tables found on Households below average income: for financial years ending 1995 to 2025 published tables or the HBAI Stat-Xplore Dataset online tool? - Is there code already that can be used or amended for the current analysis? - Can the publication tables or Stat-Xplore database be used to check totals?
What is exactly required?	- Is analysis needed on a different basis to published figures? - What will the data be used for? This will help provide the most appropriate figures and advice. - Is analysis required Before Housing Costs (BHC), After Housing Costs (AHC) or both? - Should the income be presented equivalised or unequivalised? Net or gross? - Should low income be presented as 2024/25 absolute or relative low income? - Is the analysis assessing individuals, families (benefit units) or households? Is analysis restricted to a particular group, such as children, or to a subset, e.g. only families with children as opposed to all families? - Is a time series required? Should latest, publication prices be presented or should values be uprated/deflated to another year?
Have other variables been correctly changed for amendments made?	When making amendments to the income distribution, for example by stripping out disability benefits in order to see the effects of this on low income, the mean and median income measures need to be recalculated to be consistent with this benefit removal.

D. HBAI Income Variable Adjustments

The HBAI Variable Guide provides details of which adjustments have been made to income variables or flags derived from income variables:

- SPI Adjustment
- Equivalisation
- Deflation for average of survey year prices
- Deflation for latest prices

It also provides the CPI deflators used and how to convert income values back from 'in latest prices' to 'survey year prices'.

This guide is available on the [UK Data Service](#) or for DWP users, on [SharePoint](#) in Edge.

E. Merging Datasets

It is important that the **correct level** is used when merging two datasets. The main choices of levels are:

- Individual level.
- Family (benefit unit) level.
- Household level.

One way of checking that datasets have been merged to the correct levels is to examine the number of observations in each input datasets and compare this to the merged dataset. In SAS, this can be viewed in the log.

When merging two datasets that are on the same level, the 'by' variables must be to the level of the datasets. So, for example, two datasets that are both on a benefit unit level must be merged "by sernum benunit" in SAS Eguide code and "a.sernum=b.sernum and a.benunit=b.benunit" for proc sql code.

When merging two datasets that are on different levels, the 'by' variables must go to the level of the highest dataset. So, for example, to merge an adult level dataset with a benefit unit level dataset (and so create an adult level dataset), the higher level dataset is the benefit unit level dataset, so "by sernum benunit" would be used. The variable 'person' from the adult dataset cannot be used as a 'by' variable as this information isn't on the benefit unit level dataset, and so SAS won't be able to compute a match.

An easier way of thinking about this is to match by those variables that are common to both datasets – see Table E1 below.

Table E1: By Variables for Merging to Datasets at Different Levels

Dataset A: Household Level (sernum)	Dataset A: Household Level (sernum)	Dataset A: Household Level (sernum)
Dataset B: Household Level (sernum)	Dataset B: Family Level (sernum benunit)	Dataset B: Individual Level (sernum benunit person)
Merge: By sernum	Merge: By sernum	Merge: By sernum
Dataset A: Family Level (sernum benunit)	Dataset A: Family Level (sernum benunit)	Dataset A: Family Level (sernum benunit)
Dataset B: Household Level (sernum)	Dataset B: Family Level (sernum benunit)	Dataset B: Individual Level (sernum benunit person)
Merge: By sernum	Merge: By sernum benunit	Merge: By sernum benunit
Dataset A: Individual Level (sernum benunit person)	Dataset A: Individual Level (sernum benunit person)	Dataset A: Individual Level (sernum benunit person)
Dataset B: Household Level (sernum)	Dataset B: Family Level (sernum benunit)	Dataset B: Individual Level (sernum benunit person)
Merge: By sernum	Merge: By sernum benunit	Merge: By sernum benunit person

When **merging HBAI datasets with Family Resources Survey (FRS) datasets**, only cases that appear on both datasets must be merged, and therefore not include FRS cases that are excluded due to missing spouses.

To do this, after the HBAI dataset name in the set statement, write **(in=a)**. Then, when merging, after the by statement, add a new line that says if a:

```
data child2425;
set hbaig4.allind_2425prices;
where year=31 and type=1;
run;

data child2425;
merge child2425 (in=a)
      frs2425.child(keep=sernum benunit person CDISD01);
by sernum benunit person;
if a;
run;
```

This does the following: it creates a temporary variable 'a' that doesn't stay in the final dataset. It is 1 if there was a record so the "if a;" statement acts as a gatekeeper only allowing records where a=1 through the rest of the dataset.

Please note that some FRS years have the same sernums.

To ensure the correct FRS cases are merged onto the correct HBAI year:

a) isolate to the HBAI year

```
data ind2425;
set hbaig4.allind_2425prices;
where year=31;
run;
```

b) merge on FRS variables

```
data ind2425;
merge ind2425 (in=a)
      frs2425.househol(keep=sernum bedroom);
by sernum;
if a;
run;
```

c) repeat for all other years i.e. create ind2324 back to ind9495.

d) rebuild the all years HBAI dataset. e.g.

```
data allind;
set ind9495 ind9596 ind9697 ind9798 ind9899 ind9900 ind0001 ind0102
ind0203 ind0304 ind0405 ind0506 ind0607 ind0708 ind0809 ind0910
ind1011 ind1112 ind1213 ind1314 ind1415 ind1516 ind1617 ind1718
ind1819 ind1920 ind2021 ind2122 ind2223 ind2324 ind2425;
run;
```

When using a **retain statement** (which adds up cases), the correct ‘by’ values must be used – down to the level of x that “if first.x” and “if last.x” statements go down to i.e. whatever “x” is, that is the level that the new dataset will be on.

For example, if using a family (benefit unit) level dataset, and summing up to a family (benefit unit) level, “by sernum benunit” is used as well as using “if first.benunit” and “if last.benunit”.

The code below creates a new variable, tot60plus, that adds up the number of adults in the benefit unit aged 60 or over. To do this, it uses ‘agehd’ for the age of the head of the benefit unit and ‘agesp’ for the spouse of the benefit unit.

```
data totpen;
set hbaig4.allind_2425prices (where=(year=31));
run;

data totpen;
set totpen;
by sernum benunit;
retain tot60plus;
if first.benunit then do;
    tot60plus=0;
end;
if agehd>59 then tot60plus=tot60plus+1;
if agesp ne . and agesp>59 then tot60plus=tot60plus+1;
if last.benunit;
run;
```

Some variables are useful only at **certain levels** of dataset. For example, for a household level dataset, depchldh (dependent children in the household) should be used for the number of children, not depchldb (dependant children in the benefit unit). Variables have been clearly labelled in the accompanying HBAI variable guide to present at what ‘level’ it is for.

Variables can be built by starting at an individual level and ending up at a household level. An example is given below, which adds up the benefit amounts (benamt) of Personal Independence Payment – Daily Living (benefit 96) and

Personal Independence Payment – Mobility (benefit 97) at an individual level (bentot), a family level (bentotbu) and household level (bentothh).

Worked Example: Creating a Variable At An Individual Level, Family (Benefit Unit) Level and Household Level.

*/*1. Individual Level*/*

*/*Isolate the HBAI harmonised dataset to just 2024/25*/*

```
data ind2425;
set hbaig4.allind_2425prices;
where year=31;
run;
```

*/*Merge on the benefit and benamt variables from the FRS dataset BENEFITS which is also at a person level*/*

```
data ind2425;
merge ind2425 (keep=sernum benunit person in=a)
         frs2425.benefits (keep=sernum benunit person benefit benamt);
by sernum benunit person;
if a;
run;
```

*/*Create ben96 which is the benefit amount for benefit=96 (Personal Independence Payment – Daily Living) and ben97 which is the benefit amount for benefit=97 (Personal Independence Payment – Mobility). Then create the variable bentot which adds up the PIP benefit amount for ben96 and ben97 for each person*/*

```
data ind2425 (keep=sernum benunit person bentot);
set ind2425;
by sernum benunit person;
retain ben96 ben97;
if first.person then do;
    ben96=0;
    ben97=0;
end;
if benefit=96 then do;
    ben96=benamt;
end;
if benefit=97 then do;
    ben97=benamt;
end;
if last.person;
    bentot=sum(0,ben96,ben97);
run;
```

*/*2. Family (benefit unit) level*/*

*/*Create the bentotbu, which is the total PIP benefit income in each benefit unit and adds up PIP benefit income for each person in the benefit unit. As this is a family-level dataset, remove the person level variables person and bentot*/*

```
data bu2425 (drop=person bentot);
set ind2425;
by sernum benunit;
retain bentotbu;
if first.benunit then do;
    bentotbu=0;
end;
bentotbu = sum(0,bentotbu,bentot);
if last.benunit;
run;
```

*/*3. Household Level: */*

*/*Create the bentothh, which is the total PIP benefit income for the household. As this is a household-level dataset, remove the family level variables benunit and bentotbu*/*

```
data hh2425 (drop=benunit bentotbu);
set bu2425;
by sernum;
retain bentothh;
if first.sernum then do;
    bentothh=0;
end;
bentothh = sum(0,bentothh,bentotbu);
if last.sernum;
run;
```

F. Grossing Factor Selection

Grossing factors are used to scale up sample cases to population levels. The HBAI datasets contain several grossing factors and must be selected appropriately for the analysis required.

The variables in the harmonised individual level HBAI dataset present breakdowns and income values for the individual themselves, their family (benefit unit) or their household. Therefore, it may not be necessary to create a family (benefit unit) level dataset or household level dataset from it.

Table F1 below summarises the grossing factor to be used for various types of individuals, family level and household level estimates:

Table F1: Grossing Factors to Use

Estimate	Grossing Factor	Notes
Number of children	gs_indch	Value=0 for adults
Number of adults	gs_indad	Value=0 for children
Number of working-age adults	gs_indwa	Value=0 for children and pensioners
Number of pensioners	gs_indpn	Value=0 for children and working-age adults
Total population	gs_indpp	
Number of families (benefit units)	gs_indbu	Value= 0 unless the first person in the family (benefit unit) i.e. where uperson=1
Number of households	gs_indhh	Value=0 unless the household reference person in the household i.e. where hrp1=1

For the grossing factor selection for the limited use of the family (benefit unit-level) HBAI data, please see Annex C.

G. Measuring Uncertainty in HBAI Estimates - Using the HBAI Resample Datasets

As the FRS is a sample from the UK population, any HBAI statistics derived from it are only estimates of the true numbers for the overall population.

Full details on the methodology used to calculate the HBAI resamples datasets and uncertainty in HBAI estimates can be found in Annex 2 of the [HBAI Background, Information and Methodology \(BIM\) report](#).

A separate guide, HBAI 2024/25 Harmonised Dataset: Measuring Uncertainty in HBAI Estimates, is available alongside this user guide on the UK Data Service (or in the DWP Data Documentation Hub) and provides detailed guidance on how to produce confidence intervals around a HBAI estimate and between estimates.

H. Quality Assurance and Presenting Analysis

The published HBAI tables found on the [Households below average income: for financial years ending 1995 to 2025](#) and the [HBAI Stat-Xplore Dataset](#) can help check the analysis produced is consistent. For example, by removing any restrictions on the dataset created, such as England only, so that it represents the whole UK population again, checks can be made against UK totals in published results.

Some top tips for checking analysis are:

- Ensuring that comparisons are being made on a consistent basis to reduce confusion – categories may be ordered differently in datasets compared to published results.
- Making sure results are sensible results:
 - i. For a time-series, do the figures look sensible and exist for the correct years? Do they tell a story that appears sensible?
 - ii. For compositional figures, do they sum to 100?
 - iii. Are BHC and AHC figures different? Generally, they will be, so if they are exactly the same, check the links and check that the final sheet is linking to the correct output.
 - iv. If code has been re-run, has new output been created? If a proc freq and a **pivot table** has been used and code has been re-run, ensure the contents of the pivot table have been refreshed.
- If using Excel to analyse results, it is worth setting up **links** to a final summary table as opposed to copying and pasting values into a final table. This way, if the table needs tweaking slightly, any new output will automatically feed into the summary table.

Sample Sizes

It is important to ensure that sample sizes are sufficient for analysis and reporting.

Table H1 provides some examples of how to do this, using datasets based on the family (benefit unit) level dataset and the individual level dataset:

Table H1: Worked Example for Checking Sample Sizes in Analysis**Option 1****Grossed:**

```
data pen60bhc;
set hbaig4.allind_2425prices;
where year=31;
run;
```

```
proc freq data=pen60bhc;
table low60bhc;
where type=3;
weight gs_indpp;
run;
```

Option 2**Grossed:**

```
data pen60bhc;
set hbaig4.allind_2425prices;
where year=31;
run;
```

```
proc freq data=pen60bhc;
table low60bhc;
weight gs_indpn;
run;
```

Sample:

```
data pen60bhc;
set hbaig4.allind_2425prices;
where year=31;
run;
```

/*For the individual level dataset, remove the 'weight gs_indpp' line from the code (but keep the where statement) to get sample sizes */

```
proc freq data=pen60bhc;
table low60bhc;
where type=3;
run;
```

Users must use the following sample size rules:

Table H2: HBAI Sample Size Rules

Estimate	Sample size rule:	Action to take if rule not met:
Percentages	The denominator (i.e. the row/column total) must be at least 100.	Do not present entire row/column - suppress entire row or column with (..)
Numbers	At least 50 cases.	Do not present cell - suppress cell with (..)

Ethnicity or Regional Analysis

Ethnicity or regional analysis should be presented as three-year averages as single year estimates are considered too volatile.

For example:

Year 1 estimate = 120,000

Year 2 estimate = 145,000

Year 3 estimate = 110,000

Three-year average estimate

= (Year 1 estimate + Year 2 estimate + Year 3 estimate)/3

= (120,000 + 145,000 + 110,000)/3

= 125,000

Presenting three-year averages may also be appropriate for smaller groups.

All estimates we provide excludes the 2020/21 data from the calculation of the rolling three-year average, and we would advise users of the data to adopt the same approach. This follows our decision to not publish breakdowns of the headline 2020/21 estimates. All rolling three-year averages calculated and published for any period including 2020/21 are based on the remaining two available data points only.

Therefore:

- When calculating 2018/19 to 2020/21 estimates, please calculate a two-year average using 2018/19 and 2019/20 estimates.
- For the March 2026 release, estimates cannot be calculated for 2019/20 to 2021/22 as 2020/21 data is not available and 2019/20 and 2021/22 data are based on different methodologies.

- For the summer 2026 release, when calculating 2019/20 to 2021/22 estimates, please calculate a two-year average using 2019/20 and 2021/22 estimates.
- When calculating 2020/21 to 2022/23 estimates, please calculate a two-year average using 2021/22 and 2022/23 estimates.

See the Overview of the Statistics in [HBAI Background, Information and Methodology \(BIM\) report](#) for more information.

Rounding Rules for Final Estimates

In presenting any HBAI analysis, users must use the following rounding conventions:

- Percentages must be rounded to the nearest 1 per cent.
- Numbers must be rounded to the nearest 0.1 million.
- Annual amounts must be rounded to the nearest £100.
- Weekly amounts must be rounded to the nearest £1.

Annex A: Worked Examples

Example 1: How many pensioners live in relative low income households AHC in 2024/25?

```
proc freq data=hbaig4.allind_2425prices(where=(year=31));
table low60ahc;
weight gs_indpn;
run;
```

*/*Remove the 'weight gs_indpn' line from the code to get sample sizes but to ensure that only pensioners are included, add 'where type=3'*/*

```
proc freq data=hbaig4.allind_2425prices(where=(year=31));
table low60ahc;
where type=3;
run;
```

Output Table A1a: Pensioners living in relative low income households AHC, 2024/25

	LOW60AHC	Count	Percent
Not in low income	0	10472578	86.13
In low income	1	1685946	13.87

The variable LOW60AHC = 1 indicates pensioners living in relative low income households after housing costs (AHC). This analysis shows that **1.7 million** or **14 per cent** of pensioners were in relative low income households AHC in 2024/25.

Output Table A1b: Sample Number of Pensioners living in relative low income households AHC in the UK, 2024/25

	LOW60AHC	Count	Percent
Not in low income	0	7522	86.74
In low income	1	1150	13.26

To ensure that these estimates are robust, check the sample sizes:

- Sample number needs to be over 50 sample cases: there are **1150** samples cases for pensioners in low income, so the number (count) estimate is OK.
- Denominator for percentages needs to be over 100: denominator is 8672 (7522+1150) sample cases so the percentage (percent) estimate is OK.

Example 2: How many households are in relative low income in 2024/25?

```
proc freq data=hbaig4.allind_2425prices(where=(year=31));
table low60ahc;
weight gs_indhh;
run;
```

Output Table A2: Households in relative low income in the UK, 2024/25

	LOW60AHC	Count	Percent
Not in low income	0	23412320	80.81
In low income	1	5558555	19.19

The variable LOW60AHC = 1 indicates those living in relative low income households after housing costs (AHC). This analysis shows that **5.6 million** or **19 per cent** of households were in relative low income households AHC in 2024/25.

Example 3: Removing an item from income and recalculating the relative low income line threshold for all individuals in 2024/25.

```
/* Create a temporary dataset using the FRS Benefits dataset keeping only the
details of the benefits required: 1 (DLA self-care), 2 (DLA mobility) and 12
(Attendance Allowance) */
```

```
data disben;
set frs2425.benefits (keep=sernum benunit person benefit benamt);
where benefit in (1,2,12);
run;
```

```
/* Create a benam amount variable at a household level for the HBAI dataset. */
```

```
data disben (keep=sernum benam);
set disben;
by sernum;
retain benam;
if first.sernum then benam=0;
benam=benam+benamt;
if last.sernum;
run;
```

```
/* Merge the HBAI dataset onto the temporary dataset using the variable in the
BY statement, keeping only those records that are held in the HBAI dataset using
the (IN=B) and IF B statements. Also using the IF NOT A statement to enter a
zero value for BENAM if the HBAI record does not have a benefit amount. */
```

```
data hbaind;
set hbaig4.allind_2425prices(where=(year=31));
run;
```

```

data hbaiind;
merge hbaiind (in=b)
        disben (in=a);
by sernum;
if b;
if not a then benam=0;

```

/* Creating a new deflated equivalised household income (S_OE_BHC2) by deducting the total of the benefits from the existing deflated household income. Note: Before any amounts derived from the FRS can be added/deducted from the HBAI income they must be deflated (e.g. BHCDEF) and equivalised (e.g. EQOBHCHH). */

```

s_oe_bhc2=s_oe_bhc-((benam*bhcdef)/eqobhchh);
s_oe_ahc2=s_oe_ahc-((benam*ahcdef)/eqoahchh);

```

```

run;

```

/* Calculating medians for the new variables created above and saving these in a temporary dataset called STATOIND. As dataset is at a benunit level and want the median for the whole population, use the population grossing factor.*/

```

proc univariate data=hbaiind noprint;
var s_oe_ahc2 s_oe_bhc2;
output out=statoind median=mdoeahc2 mdoebhc2;
freq gs_indpp;
run;

```

/* The new medians need to be added to the dataset. The IF _N_ statement tells SAS to put the medians against every record in the dataset */

```

data hbaiind;
if _n_=1 then set statoind;
set hbaiind;
run;

```

/* Create a second temporary dataset using the first, then construct a new low income line using the new income (S_OE_BHC2) and median (MDOEBHC2) variables to indicate if the benunit unit is in low income (=1) or not (=0).*/

```

data newhbaiind;
set hbaiind;
/* Recalculated low income line */
low60bhc2=0;
low60ahc2=0;
if s_oe_bhc2 lt 0.6*mdoebhc2 then low60bhc2=1;
if s_oe_ahc2 lt 0.6*mdoeahc2 then low60ahc2=1;
run;

```

/* Calculating the number of individuals living in low income households. First on the original basis using the low income flag on the HBAI dataset and then on the recalculated basis using the new low income line created above. As dataset is at a benunit level, use the population grossing factor.*/

```
proc freq data=newhbaiind;
table low60bhc;
weight gs_indpp;
run;
```

```
proc freq data=newhbaiind;
table low60bhc2;
weight gs_indpp;
run;
```

Table A3a: Individuals in relative low income households using the original low income threshold in the UK, 2024/25

	LOW60BHC	Count	Percent
Not in low income	0	57388846	84.09
In low income	1	10859363	15.91

The number of individuals living in low income households, before housing costs, is **10.9 million** or **16 per cent** in 2024/25.

Table A3b: Individuals in relative low income households using the new low income threshold (excluding certain disability benefits) in the UK, 2024/25

	LOW60BHC	Count	Percent
Not in low income	0	57192147	83.80
In low income	1	11056062	16.20

The number of individuals living in low income households, before housing costs, using the new low income line is **11.1 million** or **16 per cent** in 2024/25.

Therefore, the effect of removing certain disability benefits from the household income results in an increase in the number of individuals living in low income households.

Annex B: Converting code for the individual-level HBAI data

Example 1:

In this example, there are two simple changes to make:

1. The HBAI dataset needs to be isolated to just 2018/19.
2. The grossing factor needs to be changed to `gs_indpp`.

Old Code	New Code
<pre>/*Create a copy of the HBAI dataset, with a flag for low income mortgage holders*/ data hbai1819; set hbaig4.hbai1819_g4; if (low60ahc = 1 and ptentyp2 = 6) then lowincomemortgageholder = 1; else lowincomemortgageholder = 0; run; /*Calculate number of individuals in low income AHC*/ proc freq data= hbai1819; table low60ahc*ptentyp2 / nopercnt nocol norow; weight gs_newpp; title "Individuals with low income AHC, by tenure"; run;</pre>	<pre>/*Create a copy of the HBAI dataset, with a flag for low income mortgage holders*/ data hbai1819; set hbaig4.allind_2425prices(where=(year=25)); if (low60ahc = 1 and ptentyp2 = 6) then lowincomemortgageholder = 1; else lowincomemortgageholder = 0; run; /*Calculate number of individuals in low income AHC*/ proc freq data= hbai1819; table low60ahc*ptentyp2 / nopercnt nocol norow; weight gs_indpp; title "Individuals with low income AHC, by tenure"; run;</pre>

Example 2:

In this example, the two changes to make are:

1. The HBAI dataset needs to be isolated to just 2018/19.
2. The grossing factor needs to be changed to gs_indhh and the clause to isolate to just the first benunit removed (gs_indhh only has a value for the first person in the household and all other individuals in the household has gs_indhh set to 0. This ensures that a household is only counted once).

Old Code	New Code
<pre>/*Create a copy of the HBAI dataset, with a flag for low income mortgage holders*/ data hbai1819; set hbaig4.hbai1819_g4; if (low60ahc = 1 and ptentyp2 = 6) then lowincomemortgageholder = 1; else lowincomemortgageholder = 0; run; /*Calculate number of households in low income AHC*/ proc freq data= hbai1819; table low60ahc*ptentyp2 / nopercnt nocol norow; weight gs_newhh; where benunit=1; /*Isolate so only first is counted*/ title "Individuals with low income AHC, by tenure"; run;</pre>	<pre>/*Create a copy of the HBAI dataset, with a flag for low income mortgage holders*/ data hbai1819; set hbaig4.allind_2425prices(where=(year=25)); if (low60ahc = 1 and ptentyp2 = 6) then lowincomemortgageholder = 1; else lowincomemortgageholder = 0; run; /*Calculate number of households in low income AHC*/ proc freq data= hbai1819; table low60ahc*ptentyp2 / nopercnt nocol norow; weight gs_indhh; title "Individuals with low income AHC, by tenure"; run;</pre>

Annex C: Grossing factors for the family (benefit unit) level HBAI dataset

This annex aims to support users who continue to use the benefit unit level HBAI datasets via a data sharing agreement or internally at DWP. Users are encouraged to convert code to use the individual-level dataset using the instructions in this guide.

Table AC1 below summarises the grossing factor to be used for the level of the dataset created from or using the family (benefit unit) level and individual level datasets. Note that grossing factors have been created for the individual level dataset to distinguish them from the family (benefit unit) level dataset and reduce confusion.

For the family (benefit unit) level dataset:

- i. If presenting gross unequivalised income averages, restrict to the household level where benunit=1 and use gs_newbu (as gross income is household-level income).
- ii. If using equivalised net income, use the benunit level file and use gs_newpp. Therefore, the total average net income will be the same as in the HBAI publication.
- iii. Values for gs_newbu and gs_newhh are the same, but the dataset needs to be restricted to benunit=1 to produce the same population totals.
- iv. If calculating any average of a variable that is on a household basis, use gs_newhh.

The variables in the harmonised individual level HBAI dataset present breakdowns and income values for the individual themselves, their family (benefit unit) or their household. Therefore, it may not be necessary to create a family (benefit unit) level dataset or household level dataset from it.

Both the family (benefit unit) level dataset and the individual level dataset will produce the same results, if used appropriately. The user should use whichever dataset is best for them. Annex A provides some examples using both datasets to derive results.

Table AC1: Grossing factors to use for the HBAI benefit unit-level and individual datasets

	Using the Family (Benefit Unit) Dataset:			Using the Individual Dataset:
	Final dataset level			Final dataset level
Total	Individual	Family (Benefit Unit)	Household	Individual
Number of children	gs_newbu	gs_newch	can't be done	gs_indch (value=0 for non-children)
Number of adults	gs_newbu	gs_newad	can't be done	gs_indad (value=0 for non-adults)
Number of working-age adults	gs_newbu	gs_newwa	can't be done	gs_indwa (value=0 for non-working-age)
Number of pensioners	gs_newbu	gs_newpn	can't be done	gs_indpn (value=0 for non-pensioners)
Total population	gs_newbu	gs_newpp	can't be done	gs_indpp
Number of families (benefit units)	can't be done	gs_newbu	can't be done	gs_indbu (value= 0 unless the first person in the family (benefit unit) i.e uperson=1)
Number of households	can't be done	can't be done	gs_newhh or gs_newbu*	gs_indhh (value=0 unless the first person in the household i.e. person=1)

* Dataset is restricted to where benunit = 1.