

Article

Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018

Update on feasibility research producing income by ethnic group statistics for England and Wales from administrative data.

Contact:
Joanna Harkrader and Michelle
Bellham
admin.based.
characteristics@ons.gov.uk
+44 1329 444974

Release date:
20 February 2023

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [About our transformation research](#)
3. [Population coverage for admin-based income by ethnicity statistics \(ABIES\): tax year ending 2018](#)
4. [Income by ethnic group, England and Wales: tax year ending 2018](#)
5. [Income by ethnic group, subnational data: tax year ending 2018](#)
6. [Developing subnational multivariate income by ethnicity statistics from administrative data: England and Wales, tax year ending 2018: Data](#)
7. [Glossary](#)
8. [Data sources and quality](#)
9. [Future developments](#)
10. [Related links](#)
11. [Cite this article](#)

1 . Main points

- This update follows on from previous feasibility research to derive admin-based income statistics (ABIS) and admin-based ethnicity statistics (ABES), by linking together multiple administrative data sources, and it includes data for Wales for the first time.
- By combining admin-based income and admin-based ethnicity datasets, we established an income and a stated ethnicity for 77.1% of people in England and 82.1% of people in Wales aged 16 years and over, in the Statistical Population Dataset (SPD) V3.0, tax year ending 2018.
- For the first time, we have produced income percentiles for ethnic groups at different levels of geography in England and Wales, including national figures, regional figures, and figures for local authorities and lower layer super output areas (LSOAs); statistical disclosure control (SDC) means that some of the figures have been suppressed.
- We have produced income percentiles for the 15.3% of people in England and 11.5% of people in Wales with income information but no stated ethnicity; these figures indicate the potential bias in the income percentiles for different ethnic groups, which could change, were it possible to allocate all income amounts to a specific ethnic group.
- Future work will include research to improve the measures that are used to produce the admin-based income by ethnicity statistics (ABIES), to continue to explore limitations of the ABIES dataset, explore ABIES data by occupied address, and exploring methods to adjust for missingness.

These research outputs are not official statistics on income by ethnic group. They are published as outputs from research into the feasibility of producing subnational multivariate statistics using administrative data. These outputs should not be used for policy or decision-making.

2 . About our transformation research

This article presents an update on feasibility research on the potential to produce subnational multivariate income by ethnicity statistics, to demonstrate our progress towards producing more frequent subnational multivariate statistics on population characteristics. It follows on from our [first subnational multivariate income by ethnicity statistics from administrative data article](#). In that article, we presented an initial assessment of the coverage for admin-based income by ethnicity statistics (ABIES) for England, tax year ending 2016. In this update, we provide initial research findings on income by ethnic group. For the first time, we present data for Wales as well as England. We use individual net income, also known as disposable income. Our [Income and earnings statistics guide](#) outlines income measures and the national statistics currently being produced. Our [Income estimates for small areas bulletin](#) contains the currently recommended estimates for small area household income.

To produce the ABIES, this research combines the [admin-based income statistics \(ABIS\) dataset](#) with V3.0 of the [admin-based ethnicity statistics \(ABES\) dataset](#). These datasets are derived from multiple administrative data sources, which have been linked together to produce statistics about a single topic. Both datasets use the [Statistical Population Dataset \(SPD\) V3.0](#) as a population base. The data sources are linked together using pseudonymised identifiers. In producing statistics using linked administrative data, particularly for small populations, we apply the same rigour in data security and privacy as with all official statistics. For further information about the security of these linked data, see our [Population and social statistics transformation: 2019 progress update](#).

This research forms part of our [population and social statistics transformation programme](#), which aims to provide the best insights on population, migration, and society, using a range of data sources. The findings will form part of the evidence base for the National Statistician's Recommendation in 2023 on the future of population, migration, and social statistics in England and Wales.

3 . Population coverage for admin-based income by ethnicity statistics (ABIES): tax year ending 2018

The Statistical Population Dataset (SPD) 2018 V3.0 contains around 44.2 million individuals aged 16 years and over in England, and around 2.5 million individuals aged 16 and over in Wales. The coverage of the population base used, which is also under development, will affect the coverage and quality of the ABIES. For further information about the population base, see our [Population and migration statistics system transformation 2020 update](#).

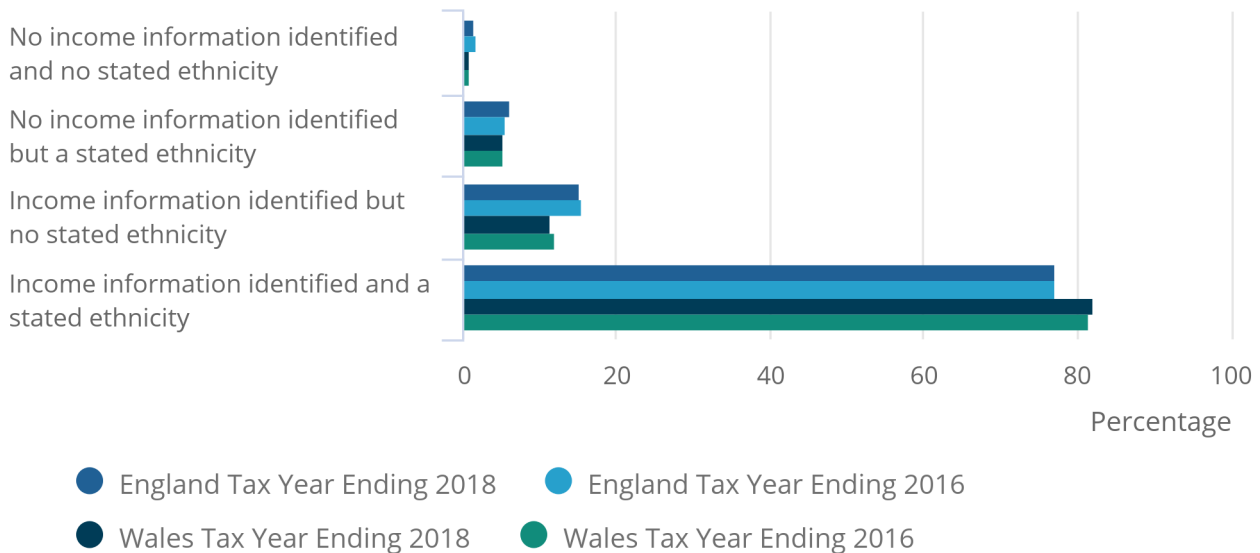
Within the SPD 2018 V3.0 for England and Wales, the admin-based income statistics (ABIS) dataset provides income information for 92.5% of individuals, and the admin-based ethnicity statistics (ABES) dataset provides a stated ethnicity for 83.4%. Figure 1 shows that 77.1% of people in England, and 82.1% of people in Wales, in the ABIES dataset for the tax year ending 2018 had income information identified and a stated ethnicity.

Figure 1: Over three-quarters of people in the SPDs for 2016 and 2018 V3.0 in England and Wales have both a source of income information identified and a stated ethnicity, tax years ending 2016 and 2018

Proportion of the population aged 16 years and over in the SPDs for 2016 and 2018 V3.0 with income information identified, a stated ethnicity, both or neither, England and Wales, tax years ending 2016 and 2018

Figure 1: Over three-quarters of people in the SPDs for 2016 and 2018 V3.0 in England and Wales have both a source of income information identified and a stated ethnicity, tax years ending 2016 and 2018

Proportion of the population aged 16 years and over in the SPDs for 2016 and 2018 V3.0 with income information identified, a stated ethnicity, both or neither, England and Wales, tax years ending 2016 and 2018



Source: Office for National Statistics

Notes:

1. Proportions may not sum to 100.0% because of rounding.
2. 'Stated ethnicity' refers to people with a stated ethnicity and no refusal on their most recent administrative data record in 2016 and 2018.
3. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to [derive an individual's ethnic group in the ABES dataset](#). This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
4. 'Income information identified' refers to people identified with income data (regardless of value) from at least one source in the ABIS dataset. See [Section 7. Glossary](#) for more information about the sources used to derive income in the ABIS dataset.
5. 'No income information identified' refers to those individuals who are in the SPD V3.0 but who have not appeared in any of the sources in the ABIS dataset.
6. Figures for 2016 have been derived using ABES 2016 V3.0 and include both England and Wales; in our [first publication](#) ABES 2016 V2.0 was used and data published for England only.

Coverage of the ABIES dataset varies by age, sex, and geography, though England and Wales show very similar trends. Coverage of the 2018 and 2016 ABIES datasets is consistent. However, there was improved coverage for women aged 65 in the 2018 dataset, compared with age 63 in the 2016 dataset, corresponding with the change in state pension age. These data can be looked at in more detail in [Section 6 . Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018](#) data.

Variation in coverage of English regions is also consistent between the 2018 and 2016 ABIES datasets. The proportion of individuals with income information identified and a stated ethnicity for the tax year ending 2018 ranged from 66.3% in London to 81.7% in the North West. These data can be looked at in more detail in [Section 6. Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018](#) data.

Figure 2 shows that coverage at the local authority (LA) level varies widely, from 50.5% for Westminster to 89.3% for St Helens. Of LAs in England, 83.8% of them (n=259) have income information identified and a stated ethnicity for at least 70% of individuals aged 16 years and over. Of LAs in Wales, 100% of them (n=22) have income information identified and a stated ethnicity for at least 70% of individuals aged 16 years and over.

Figure 2: Coverage of income and ethnicity varies by local authority in England and Wales

Number of local authorities by proportion of individuals aged 16 years and over in the SPD V3.0 with both income information identified and a stated ethnicity, England and Wales, tax year ending 2018.

Notes:

1. There are 309 English and 22 Welsh local authorities in this analysis. The SPD V3.0 uses the 2021 local authority geographies
2. 'Stated ethnicity' refers to people with a stated ethnicity and no refusal on their most recent administrative data record in 2016 and 2018, in line with the methods used to [derive an individual's ethnic group in the ABES dataset](#).
3. 'Income information identified' refers to people identified with income information from at least one source in the ABIS dataset. See [Section 7. Glossary](#) for more information about the sources used to derive income in the ABIS dataset.

Download the data

[.xlsx](#)

Coverage at lower layer super output area (LSOA) level varies more widely than LA level. These data can be looked at in more detail in [Section 6. Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018](#) data.

4 . Income by ethnic group, England and Wales: tax year ending 2018

Background and limitations

The figures in this section should not be used for analysis or policy purposes, owing to important quality considerations and limitations of these data.

While not directly comparable with other income sources, we are aware these findings from the admin-based income by ethnicity statistics (ABIES) do not align with the general trends by ethnic group observed in other analyses of income by ethnic group.

The income by ethnic group data presented in this article are based on the 77.1% of people aged 16 years and over in England and 82.1% in Wales that we have income information identified and a stated ethnicity for in the ABIES dataset for tax year ending 2018. These people and their associated incomes may not be representative of the income percentiles of the underlying population by ethnic group. Additionally, the Statistical Population Dataset (SPD) used as the population base has some over-coverage and under-coverage, which could affect the accuracy of income statistics produced from the ABIES dataset. We have included the income data for those with no stated ethnicity for comparison.

The admin-based income statistics (ABIS) are still under development and do not capture some components of income, such as investment income and income from rental and royalties. These gaps have the potential to affect the accuracy of the income figures by ethnic group. See our [ABIS QMI](#) for further details on current ABIS methodology. The income figures have not been adjusted to account for missingness in the data. We present individual income figures in this publication, whereas most income statistics are presented at the household level.

There may be issues of mis-recording of ethnicity in the administrative data sources used to provide ethnicity data. Guidance for collecting ethnicity data across all sources states that the ethnicity should be self-reported. However, achieving this in practice can be challenging. This means that some people, and their associated income, may be recorded within an ethnic group that is different from the one they identify with. For example, the [Nuffield Trust research on ethnicity coding in English health datasets](#) identified challenges including:

- staff being under pressure
- patients being unwell or lacking capacity
- staff being unaware of the required procedure for capturing ethnicity information
- inconsistent response options across organisations and care settings

Potential sources of error in ethnicity data collection in hospitals are explored in our [Methods and systems used to collect ethnicity information in health administrative data sources, England: 2022 article](#).

The extent of ethnicity mis-recording is unknown. Some insight can be gained through our previous work comparing the ethnicity recorded in the administrative data against the 2011 Census, as discussed in our [Producing admin-based ethnicity statistics for England: changes to data and methods article](#). This showed high levels of agreement, overall, but with differences by ethnic group. However, it is important to note that differences between the recorded ethnicity in the administrative data and the 2011 Census does not necessarily mean that the ethnicity has been mis-recorded in the administrative data. It may reflect that the individual has chosen to give a different response on different occasions. There are several reasons why this could happen, including different questions or response options, an individual identifying with multiple options, and change in ethnic identity over time.

We have not been able to analyse how representative the individuals with income information identified and a stated ethnicity are of the total population by local authority (LA). If the proportion of people for whom we do have income information identified and a stated ethnicity share the same characteristics as those for whom we do not have data, then lower proportions of individuals with income information identified and a stated ethnicity would not automatically equate to less reliable analysis.

As we are not yet able to determine how representative our analysis is of the total population in terms of income and ethnicity, the figures in this feasibility research may not reflect the incomes of the underlying population and should be interpreted with caution.

We feel the ABIES have potential despite these limitations, and this publication provides an update including producing subnational income by ethnic group data for the first time. Ongoing development of the ABIS and admin-based ethnicity statistics (ABES) in the year ahead will aim to assess the representativeness and completeness of our univariate and multivariate measures. We will do this through aggregate and individual level comparisons with data from Census 2021 and other relevant sources. We will also further develop our [Methods for producing multivariate population statistics using administrative and survey sources \(PDF, 353KB\)](#). See also [Section 9. Future developments](#) for further information on our plans.

ABIES income by ethnic group, tax year ending 2018

Figure 3 shows that the median net income for individuals in England with income information identified and a stated ethnicity in the ABIES dataset for the tax year ending 2018 is £15,456. The median income for individuals with income information identified but no stated ethnicity is £16,008.

The 10th percentile for individuals in England with a stated ethnicity and individuals with no stated ethnicity is similar. However, the 90th percentile is higher for individuals in England with no stated ethnicity. This implies there could be more variation in incomes among those with no stated ethnicity than among those with a stated ethnicity in England (15.3% and 77.1% of individuals aged 16 years and over in the SPD 2018 V3.0, respectively).

There are some differences when the findings are reviewed by age group. For individuals aged 25 to 29 years and 35 to 44 years in England, the White ethnic group has the highest median net income. For individuals aged 45 to 84 years, the Black ethnic group has the highest median income.

Figure 3: Income measures in the ABIES dataset for England show variation between the ethnic groups

The 10th percentile, median, and 90th percentile net income for individuals aged 16 years and over in the SPD 2018 V3.0 with income information identified by 5 category ethnic group, England, tax year ending 2018

Notes:

1. 'Stated ethnicity' refers to people with a stated ethnicity and no refusal on their most recent administrative data record in 2018.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to [derive an individual's ethnic group in the ABES dataset](#). This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'All with income information identified' refers to people identified with income data (regardless of the value) from at least one source in the ABIS dataset, and includes individuals with a stated ethnicity and individuals with no stated ethnicity. See [Section 7. Glossary](#) for more information about the sources used to derive income in the ABIS dataset.
4. See [Section 7. Glossary](#) for information about the ethnic groups included in each of the five category ethnic groups represented in the figure.

Download the data

[.xlsx](#)

Figure 4 shows there is considerable variation in the 10th percentile, median, and 90th percentile income within ethnic groups in England.

For the Asian ethnic group, the Asian Bangladeshi and Asian Pakistani ethnic group 90th percentile incomes are considerably lower than the 90th percentile incomes for the other Asian ethnic subgroups. The income percentiles for the Black ethnic group and the Mixed ethnic group are generally similar across all subgroups. Income figures have not been provided for the subgroups within the higher-level White and Other ethnic groups because of differences in response options across data sources. More information about the differences can be found in our [Developing admin-based ethnicity statistics for England and Wales: 2020 article](#). These differences mean that there is considerable bias in who is captured within these subgroups, which means that any income figures derived for these subgroups are unlikely to be representative. It is our ambition to be able to produce income statistics for all lower-level ethnic groups in the future.

Figure 4: Income measures in the ABIES dataset for England show variation within ethnic groups

The 10th percentile, median, and 90th percentile net income for individuals aged 16 years and over in the SPD 2018 V3.0 with income information identified by 14 category ethnic group, England, tax year ending 2018

Notes:

1. 'Stated ethnicity' refers to people with a stated ethnicity and no refusal on their most recent administrative data record in 2018.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to [derive an individual's ethnic group in the ABES dataset](#). This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'All with income information identified' refers to people identified with income data (regardless of the value) from at least one source in the ABIS dataset and include individuals with a stated ethnicity and individuals with no stated ethnicity. See [Section 7. Glossary](#) for more information about the sources used to derive income in the ABIS dataset.
4. Income figures have not been provided for the subgroups within the higher-level White and Other ethnic groups because of differences in response options across data sources. These differences mean that there is considerable bias in who is captured within these subgroups, which means that any income figures derived for these subgroups are unlikely to be representative.
5. See [Section 7. Glossary](#) for information about the ethnic groups included in each of the 14 category ethnic groups represented in the figure.

Download the data

[.xlsx](#)

Figure 5 shows that the median net income for individuals in Wales with income information identified and a stated ethnicity in the ABIES dataset for the tax year ending 2018 is £15,126. The median net income for individuals with no stated ethnicity is £14,681. The 90th percentile is slightly higher for individuals in Wales with a stated ethnicity.

Figure 5: Income measures in the ABIES dataset for Wales show variation between the ethnic groups

The 10th percentile, median, and 90th percentile net income for individuals aged 16 years and over in the SPD 2018 V3.0 with income information identified by five category ethnic group, Wales, tax year ending 2018

Notes:

1. 'Stated ethnicity' refers to people with a stated ethnicity and no refusal on their most recent administrative data record in 2018.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to [derive an individual's ethnic group in the ABES dataset](#). This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'All with income information identified' refers to people identified with income data (regardless of the value) from at least one source in the ABIS dataset and includes individuals with a stated ethnicity and individuals with no stated ethnicity. See [Section 7. Glossary](#) for more information about the sources used to derive income in the ABIS dataset.
4. See [Section 7. Glossary](#) for information about the ethnic groups included in each of the five category ethnic groups represented in the figure.

Download the data

[.xlsx](#)

Figure 6 shows there is variation in the 10th percentile, median, and 90th percentile net income within ethnic groups in Wales. The Asian ethnic subgroups have similar 10th percentile net incomes, but the Asian Indian ethnic group has a higher 90th percentile net income (£48,953) than the other Asian ethnic subgroups. The smaller sizes of some of the subgroups in the 14 category ethnic breakdowns make them more susceptible to outlier income amounts, which could explain the higher 90th percentile for the Asian Indian ethnic group. The income percentiles for the Black ethnic group and the Mixed ethnic group are similar across all subgroups.

When variations by age group are reviewed, the Other ethnic group has the highest median net income for people aged 65 to 74 years; the no stated ethnicity group has the highest median net income for people aged 16 to 24 years, and the Black ethnic group has the highest median net income for people aged 50 to 64 years.

Figure 6: Income measures in the ABIES dataset for Wales show variation within ethnic groups

The 10th percentile, median, and 90th percentile net income for individuals aged 16 years and over in the SPD 2018 V3.0 with income information identified by 14 category ethnic group, Wales, tax year ending 2018

Notes:

1. 'Stated ethnicity' refers to people with a stated ethnicity and no refusal on their most recent administrative data record in 2018.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to [derive an individual's ethnic group in the ABES dataset](#). This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'All with income information identified' refers to people identified with income data (regardless of the value) from at least one source in the ABIS dataset and includes individuals with a stated ethnicity and individuals with no stated ethnicity. See [Section 7. Glossary](#) for more information about the sources used to derive income in the ABIS dataset.
4. Income figures have not been provided for the subgroups within the higher-level White and Other ethnic groups because of differences in response options across data sources. These differences mean that there is considerable bias in who is captured within these subgroups, which means that any income figures derived for these subgroups are unlikely to be representative.
5. See [Section 7. Glossary](#) for information about the ethnic groups included in each of the 14 category ethnic groups represented in the figure.

Download the data

[.xlsx](#)

Income by ethnic group, official statistics

There are several different sources of official statistics related to income, as set out in our [Income and earnings statistics guide](#). In this section, we review trends from the household disposable income and inequality (HDII) national statistics in our [Average household income, UK: financial year ending 2021 bulletin](#), derived from the [Household Finances Survey \(HFS\)](#). Note that disposable income in HDII and net income in ABIES are referring to the same income concept, although several important differences between HDII and ABIES mean direct comparison is not possible. The HDII data below are for England only. Sample sizes mean that HDII data for Wales, broken down by ethnic group, have not been possible to include.

Firstly, disposable income definitions differ slightly between the ABIES and HDII because of some income components not yet being captured in the ABIES. See [Section 8. Data sources and quality](#) for more information about income components captured in each of the measures.

Secondly, ABIES data are currently based on individual disposable income and individual ethnicity, whereas HDII estimates are based on household disposable income. Household disposable income is typically equivalised, to account for the impact of household composition (number of adults and children) on living conditions.

Although there are important differences between ABIES and the HDII official statistics, national trends are analysed to highlight what the ABIES show with the admin data currently available.

The ABIES for England, tax year ending 2018, show that the Asian ethnic group has the lowest median individual net income, and the no stated ethnicity group has the highest individual net income. Figure 7 shows that the HDII data have a different ordering of ethnic groups by income for household disposable income, possibly because of the differences to ABIES methodology. For instance, the Black ethnic group have the lowest median household disposable income, and the White ethnicity group have the highest median household disposable income. Confidence intervals overlap for multiple groups, so caution should be taken when interpreting differences. Please see our [Uncertainty and how we measure it for our surveys article](#).

Equivalised household income broken down by ethnic group is available at each stage of the tax and benefit system for the tax year ending 2021. It is part of our annual [Effects of taxes and benefits on household income dataset](#).

Figure 7: Trends in household disposable income estimates by ethnic group show the variation between the ethnic groups with some overlapping confidence intervals.

Median equivalised household disposable income by ethnic groups, three-year rolling average centres on tax year ending 2018, England.

Notes:

1. For HFS ethnic group classifications, see [Section 7. Glossary](#).
2. Three-year rolling averages are used because of sample-size restrictions. Tax year ending 2018 results include responses from 2017, 2018 and 2019.
3. Confidence intervals for some ethnic groups are wider because of both smaller sample sizes and greater variability in data.

Download the data

[.xlsx](#)

5 . Income by ethnic group, subnational data: tax year ending 2018

For the first time, we have produced admin-based net income by ethnicity statistics for subnational levels of geography. These data can be looked at in more detail in [Section 6. Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018](#) data. Official statistics on net income by ethnic group are not routinely published for subnational levels of geography for single reference years; this research has the potential to provide more granular data than is currently available.

Regional income by ethnic group

Figure 8 shows that income by ethnic group trends vary across English regions in the ABIES dataset. Across all ethnic groups, median net incomes were higher than their respective England medians in London, the South East, and East of England. Across all English regions, the no stated ethnicity group are among those with the highest 90th percentile net incomes, which may influence the trends being observed across the ethnic groups. Regional trends are likely to be affected by the limitations outlined in [Section 4. Income by ethnic group, England and Wales: tax year ending 2018](#).

Figure 8: 10th percentile, median and 90th percentile net income in the ABIES dataset show regional variation between the ethnic groups

The 10th percentile, median and 90th percentile net income for individuals aged 16 years and over in the SPD 2018 V3.0, by ethnic group and English region, tax year ending 2018

Notes:

1. 'Stated ethnicity' refers to people with a stated ethnicity and no refusal on their most recent administrative data record in 2018.
2. 'No stated ethnicity' refers to the ethnicity being recorded as refused or unknown, in line with the methods used to [derive an individual's ethnic group in the ABES dataset](#). This includes individuals who are in the SPD V3.0 but have not been linked to any sources of ethnicity data.
3. 'Income information identified' refers to those identified with income data (regardless of the value) from at least one source in the ABIS dataset. See [Section 7. Glossary](#) for more information about the sources used to derive income in the ABIS dataset.
4. See [Section 7. Glossary](#) for information about the ethnic groups included in each of the five category ethnic groups represented in the figure.

Download the data

[.xlsx](#)

Local authority (LA) and lower layer super output area (LSOA) income by ethnic group

Income percentiles at LA and LSOA level geographies have been produced for the five category ethnicity classification. These data show the variation in incomes across ethnic groups within regions and can be looked at in more detail in [Section 6. Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018](#) data.

ABIES data for LAs and LSOAs are likely to be affected more by coverage issues and accuracy of the administrative data. As with the ABES and the admin-based housing by ethnicity statistics (ABHES), this can be caused by small population sizes in some geographical areas and potential localised factors. Where an ethnic group is small in an LA or LSOA, the value in the accompanying table has been suppressed for statistical disclosure control (SDC). This is in accordance with the approaches currently required by administrative data suppliers, which vary from those used for other statistics (for example, Census 2021). A full explanation of the SDC approach used can be found in the notes of the accompanying datasets.

Under the current SDC approach, we are not yet able to produce complete ABIES for all areas at sub-regional levels of geography. We can produce complete income percentiles for all ethnic groups in 88.3% (n=273) of LAs and nine LSOAs in England, and 45.4% (n=10) of LAs and one LSOA in Wales, using the five category ethnicity classification.

Further work to reduce the missingness in the ABIS and ABES will improve the number of geographical areas income percentiles can be reported for. However, it is still likely that there will be high levels of suppression using the current approach to SDC.

6 . Developing subnational multivariate income by ethnicity statistics from administrative data: England and Wales, tax year ending 2018: Data

[Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018](#) Dataset | Released 20 February 2023 Data for update on feasibility research producing income by ethnicity statistics for England and Wales from administrative data.

7 . Glossary

Equivalisation

Equivalisation is the process of accounting for the fact that households with many members likely need a higher income to achieve the same standard of living as households with fewer members. Equivalisation considers the number of people living in the household and their ages. It acknowledges that while a household with two people in it will need more money to sustain the same living standards as one with a single person, the two-person household is unlikely to need double the income.

See concepts and definitions in our [admin-based income statistics quality and methodology information \(QMI\) report](#) for more detail and other income related definitions.

Ethnic group

The Office for National Statistics (ONS) notes that there is no consensus on [what constitutes an ethnic group](#). For this article, we define ethnic group as the self-reported ethnic group of the individual, according to their own perceived ethnic group and cultural background.

However, when providing their ethnic group in the administrative and census data collections, respondents must choose from the response options presented to them or decide not to respond. Response options differ across the different data collections. For some administrative data collections, a more detailed set of response options was used. We have aggregated these into the ethnic groups used in this article.

A [harmonised standard](#) is available for the collection of data on ethnicity. [The standard is currently under review](#) to ensure that it continues to meet user needs and reflects the diversity of the population.

Ethnic groups defined by admin-based income by ethnicity statistics

The admin-based income by ethnicity statistics (ABIES) data in this article are presented for the five high-level ethnic groups. This is a list of the ethnic groups included in each category:

- Asian ethnic group: Bangladeshi, Chinese, Indian, Pakistani, Asian Other
- Black ethnic group: African, Caribbean, Black Other
- Mixed ethnic group: White and Asian, White and Black African, White and Black Caribbean, Mixed Other
- White ethnic group: British, Gypsy, Roma or Irish Traveller [note 1], Irish [note 2], White Other, White not specified [note 3]
- Other ethnic group: Arab [note 4], Any other ethnic group

ABIES data are also presented for the subgroups within the Asian, Black and Mixed ethnic groups. They are not provided for the subgroups within the higher-level White and Other ethnic groups because of differences in response options across data sources, as noted below.

Notes for ethnic groups defined by admin-based income by ethnicity statistics

1. The Gypsy, Roma and Irish Traveller ethnic groups have been aggregated because of differences in response options across data sources, meaning that it is not possible to separate them. Hospital Episode Statistics (HES) and Improving Access to Psychological Therapies (IAPT) do not include any Gypsy, Roma or Irish Traveller response options.
2. Welsh School Census (WSC) and HESA collections in England and Wales do not have a White Irish response option.
3. The Higher Education Statistics Agency (HESA) data for England and Wales only have categories for White and Gypsy or Traveller within the higher-level White ethnic group. Those with a sub-category ethnicity of White in HESA were re-coded as White not specified.
4. The Arab ethnic group is not a category in the Birth Notifications, HES, ESC, or IAPT data sources.

Ethnic groups defined by Household Finances Survey (HFS)

Ethnicity is self-reported on the Household Finances Survey (HFS) for all individual respondents. Answers are based on eighteen response options from which the five ethnic group categories presented in this article are compiled.

HFS ethnic group classifications:

- White encompasses White British, Irish, Gypsy or Irish Traveller and White Other ethnic group categories
- Mixed encompasses White and Asian, White and Black African, White and Black Caribbean and Other Mixed ethnic group categories
- Asian encompasses Bangladeshi, Chinese, Indian, Pakistani and Asian Other ethnic group categories
- Black encompasses Black African, Black Caribbean and Black Other ethnic group categories
- Other encompasses Arab and Other ethnic group categories

Ethnicity stated and no ethnicity stated

Ethnicity stated refers to the ethnicity being recorded as a specific ethnic group and not refused or unknown on their most recent administrative data record in 2018, in line with the [methods used to derive an individual's ethnic group defined in our Producing admin-based ethnicity statistics for England: changes to data and methods article](#).

No stated ethnicity refers to the ethnicity being recorded as refused or unknown, in line with the methods used to derive an individual's ethnic group defined in our Producing admin-based ethnicity statistics for England: changes to data and methods article. No stated ethnicity also includes individuals who are in the Statistical Population Dataset (SPD) V3.0 but have not been linked to any sources of ethnicity data.

Income information identified

Income identified refers to those identified with income data from at least one source in the admin-based income statistics (ABIS) dataset. For more information about the income measures used in the ABIS and the ABIES, please refer to our [ABIS Quality and Methodology Information \(QMI\) report](#).

Lower level super output area (LSOA)

A geographic hierarchy designed to improve the reporting of small area statistics in England and Wales. They are built from groups of contiguous Output Areas and have been automatically generated to be as consistent in population size as possible, and typically contain from four to six Output Areas. The minimum population is 1,000 and the mean is 1,500.

Percentiles

The percentile is the value at which the percentage of records falls on or below. For example, a 90th income percentile for an ethnic subgroup of £40,000 means 90% of incomes within that subgroup fall on, or below, £40,000.

8 . Data sources and quality

Population base

The 2018 Statistical Population Dataset (SPD) V3.0 was used as the population base for the admin-based ethnicity and admin-based income statistics datasets. It aims to approximate the usually resident population as of 30 June 2018. The quality of the population base, set out in our [Admin-based population estimates and statistical uncertainty: July 2020 article](#), will have an impact on the quality of the admin-based income and ethnicity statistics. More information about the coverage of the population base can be found in our [Population and migration statistics system transformation 2020 update](#).

The SPD was previously called the admin-based population estimates (ABPE), and is one of the inputs to our [Dynamic population model \(DPM\) for England and Wales](#), our statistical modelling approach to produce more timely estimates of the population that are able to better respond to user needs.

Admin-based income statistics (ABIS) dataset

The [Admin-based income statistics \(ABIS\) dataset](#) was produced using the following administrative data sources:

- Department for Work and Pensions' (DWP) National Benefits Database
- DWP's Single Housing Benefit Extract
- DWP's Personal Independence Payment
- DWP's Universal Credit
- HM Revenue and Customs' (HMRC) Pay As You Earn P14 data
- HMRC's Tax Credits
- HMRC's Child Benefit
- HMRC's Self Assessment
- Office for National Statistics (ONS)-derived Winter Fuel Payment and Christmas Bonus

Please see [Section 6. Methods used to produce the admin-based income statistics data of our ABIS QMI](#) for detail on the methods used to produce the ABIS.

Admin-based ethnicity statistics (ABES) dataset

The [Admin-based ethnicity statistics \(ABES\) dataset](#) was produced using the following administrative data sources:

- [English School Census \(ESC\)](#), 2011 to 2018
- [Welsh School Census](#), 2011 to 2018
- [Individualised Learner Record](#), 2008 to 2018
- [Higher Education Statistics Agency \(HESA\)](#), 2010 to 2018
- [Hospital Episode Statistics \(HES\)](#), 2009 to 2018
- [Improving Access to Psychological Therapies \(IAPT\)](#), 2012 to 2018
- [Birth Notifications](#), 2006 to 2018
- [2011 Census](#) data

Ethnicity records from these data sources were linked to the 2018 SPD V3.0 using unique identifiers. A method to select a final ethnicity per person was then implemented, as described in Section 4. Exploring alternative ethnicity selection rules of our [Producing admin-based ethnicity statistics for England: changes to data and methods](#) article.

Creating the income and ethnicity joined data asset

As the admin-based income and ethnicity datasets were both created using the SPD V3.0, they could be joined together using a unique identifier. All records for people aged under 16 years were also dropped, as the population of interest for income statistics is people aged 16 years and over. This left 46.7 million records for the analysis.

Ethnic group breakdowns

Income figures have been provided for the five high-level ethnic groups and for the subgroups within the Asian, Black, and Mixed ethnic groups. Income figures have not been provided for the subgroups within the higher-level White and Other ethnic groups because of differences in response options across data sources. More information about the differences can be found in our [Developing admin-based ethnicity statistics for England and Wales: 2020](#) article. These differences mean that there is considerable bias in who is captured within these subgroups, which means that any income figures derived for these subgroups are unlikely to be representative. It is our ambition to be able to produce income statistics for all lower-level ethnic groups in future.

Statistical disclosure control (SDC)

Our disclosure control approach in this analysis has been to suppress data based on seven or fewer people and round figures to the nearest five. This is in line with the current requirements of agreements we have made with administrative data suppliers. This means that a higher volume of suppression is required for data at lower levels of geography or for some ethnic groups, because of small sizes of the populations of some ethnic groups in some areas. This suppression would be required even if 100% coverage in administrative data was achieved, and is the case for both univariate and multivariate analysis.

Household Finances Surveys (HFS) disposable income

HFS disposable income estimates are made up of:

- original income, including wages and salaries, imputed income from benefits in kind, self-employment income, private pensions or annuities, investment income (including rent received before any income tax deduction) and other income
- direct benefits in cash minus direct national insurance contributions, income tax, student loan repayments, council tax and national insurance rates, minus council tax rebates or rates rebates

9 . Future developments

Further work is required to explore how we can develop the admin-based income by ethnicity statistics (ABIES). Areas for further research include:

- continuing to progress our research to improve the measures that are used to produce the ABIES, such as the admin-based income statistics (ABIS) and the admin-based ethnicity statistics (ABES), including incorporating additional data sources to improve population coverage
- continuing to explore how the limitations of the ABIES dataset may be influencing our findings that do not align with trends by ethnic group in official statistics
- exploring methods for producing multivariate population statistics using administrative and survey sources, building on research set out in the [Methods for producing multivariate population statistics using administrative and survey sources \(PDF, 353KB\)](#) research paper
- exploring ABIES data by occupied address and compare with official statistics, such as household disposable income and inequality (HDII) estimates
- working in collaboration with data suppliers to explore alternative methods of disclosure control which could enable more data for small population groups to be released, while still preventing identification of individuals

Feedback

We welcome feedback on the quality, value, and potential impact of the admin-based income by ethnicity statistics. We also welcome feedback on our planned future developments. Please email your feedback to Admin.Based.Characteristics@ons.gov.uk.

10 . Related links

[Developing subnational multivariate housing by ethnicity statistics from administrative data, England and Wales: 2020](#) Article | 16 February 2023 Feasibility research on producing housing by ethnicity statistics for England and Wales from administrative data.

[Occupied address-level ethnicity measures for multivariate statistics, England and Wales: 2020](#) Article | 16 February 2023 Approaches used to identify occupied address-level ethnicity for feasibility research on producing multivariate statistics for England and Wales from administrative data.

[Admin-based income statistics, England and Wales: tax year ending 2018](#) Article | Released 20 December 2022 Experimental gross and net, individual and occupied address income statistics for small areas using administrative data from Pay As You Earn, Self Assessment and benefit systems.

[Admin-based income statistics QMI](#) Methodology | Released 20 December 2022 Quality and Methodology Information (QMI) for admin-based income statistics in England and Wales, detailing the strengths and limitations of the data, methods used, and data uses and users.

[Developing subnational multivariate income by ethnicity statistics from administrative data, England: tax year ending 2016](#) Article | Released 31 August 2022 Feasibility research on an initial case study producing income by ethnic group statistics for England from administrative data

[Developing admin-based ethnicity statistics for England and Wales: 2020](#) Article | Released 7 February 2023 Research update on producing population statistics by ethnic group for England and Wales from administrative data, with comparisons with Census 2021 estimates.

[Income and earning statistics guide](#) Methodology | Last revised 27 January 2023 Explains the relationship between income and earnings data and outlines the statistics produced by the Office for National Statistics, Department for Work and Pensions and HM Revenue and Customs.

[Methods for producing multivariate population statistics using administrative and survey sources \(PDF, 353KB\)](#) Methodological Assurance Panel Paper | EAP 186 This paper provides an outline for the programme of methodological work to produce multivariate population outputs which are primarily based on administrative data but use survey and other data sources to provide robust outputs that account for missingness and other data problems.

11 . Cite this article

Office for National Statistics (ONS), released 20 February 2023, ONS website, article, [Developing subnational multivariate income by ethnicity statistics from administrative data, England and Wales: tax year ending 2018](#)