Office for National Statistics

Article

Exploring the use of admin data to derive small area income estimates, England and Wales

Comparison of admin-based small area income (official statistics in development) statistics (ABIS) with official accredited modelled estimates and how they meets the Canberra Handbook criteria.

Contact:
Financial Well-being team
Economic.Wellbeing@ons.gov.
uk
+44 1329 447767

Release date:
30 April 2024

Next release:
To be announced

Table of contents

1 . Main points

- Income information in 2018 was able to be identified from at least one data source for 92.8% of people aged 16 years and over and 98.6% of occupied addresses in England and Wales on our Statistical Population Database (SPD).

- Coverage varied across geographic and demographic characteristics; however, higher proportions of occupied addresses in London compared with other regions and of young people compared with other age groups did not have income information identified.

- Mean gross household incomes, as derived by admin-based income statistics (ABIS) and small area income estimates (SAIE) for Middle-layer Super Output Areas (MSOAs) are similar, with ABIS incomes slightly higher at the upper end of the distribution.

- The main components of gross and net household income for three of the four SAIE metrics are currently available from administrative sources; however, additional data on housing costs are required to generate income statistics after housing costs.

2 . Overview

This article provides an update on our research on the development of income statistics for small areas. It forms part of our wider research towards a future population and migration statistics system, detailed in our dashboard on the topic. The UK Statistics Authority (UKSA) will be publishing a recommendation to government regarding our proposals later this year. As set out in the Census 2021 white paper (PDF, 967KB), we have committed to developing "census type income data" in response to the need of our users for income data produced for local geographies and population sub-groups. It was found in our article, 2007 Census Test (PDF, 449KB), that including an income question on the census had an unacceptable impact on response rates and data accuracy. Since then, we instead have explored approaches to our articles, Imputing an administrative-based income variable to the 2011 Census data, and Developing methods to produce Census 2021 income data. The summary of our article, Population and migration statistics transformation in England and Wales, research overview: 2023, includes further information on the proofs of concept delivered for a number of topic areas, including income.

For a number of years, we, at the Office for National Statistics (ONS), have been publishing admin-based income statistics (ABIS) articles that demonstrate the potential of using administrative data to estimate income for individuals and occupied addresses (for instance, unique property reference numbers (UPRN)). These are official statistics in development, and are published separately to our accredited official statistics, as set out in our Income estimates for small areas (SAIE), England and Wales bulletin. SAIE are statistically modelled to Middle-layer Super Output Area (MSOA)-level from survey data and using the same administrative data to support the predictions.

To progress the development of admin-based statistics, this article explores the two approaches alongside international guidelines for the compilation of income estimates and considers the next steps towards developing more granular breakdowns by subgroups of the population and their characteristics. This would enable Lower-layer Super Output Area (LSOA)-level income estimates of both mean and median income, as well as estimates of the proportion of households which fall below income thresholds within each small area; all of these are statistics commonly requested by users.

After summarising the methodological approaches to the statistics, this analysis assesses the extent to which administrative data is available for all the components of income required in a comprehensive measure (as set out in the Canberra handbook) and how these compare with those used in SAIE. The third section explores the level of demographic coverage for each of the ABIS components of income as well as the overall SPD administrative data frame. Finally, the article compares MSOA-level ABIS incomes against those published for SAIE for the financial year ending 2018.

3 . Comparison of methodological approaches

We currently produce accredited official statistics, small area income estimates (SAIE) for Middle-layer Super Output Areas (MSOAs). MSOAs have a mean population of 7,200 and a minimum population of 5,000 and consist of between 2,000 and 6,000 households. Estimates are produced through modelling of the Department for Work and Pensions's (DWP) Family Resources Survey (FRS) using census and other administrative information available at local levels. As accredited official statistics, these have been confirmed to comply with the standards of trustworthiness, quality and value in the Code of Practice for Statistics.

The model-based SAIE are underpinned by the FRS, which captures a wider range of income information than the admin-based income statistics (ABIS) (as presented in Section 4: Comparison of income components); however, they are based on a sample of households. Producing estimates of income for small areas requires the sample to be sufficiently large to represent each area of interest. The FRS is underpinned by robust and advanced sampling and weighting techniques to make it as representative of the national population as possible.

The current modelling methodology is used to produce estimates of the mean income for each MSOA and cannot be used to produce distributional estimates (such as medians) or support measures for smaller areas.

We recognise the user need for improved geographic detail and the availability of median and threshold levels for distributional analyses; for example, the proportion of households within a small area whose income is below 60% of national median income. Such user needs could be met by development of the modelling methodology, or in the future by ABIS.

ABIS are produced by bringing together individual-level records within administrative datasets (using pseudonymised identifiers) to form a combined measure of income per person. These administrative data are sourced from operational systems and delivered to us for use in statistics production and statistical research. The combined income dataset is linked to a population base using a pseudonymised identifier. Using this method and subject to disclosure controls, estimates of means, medians, and other distributional metrics may be produced for very small geographical areas.

Statistics on income for occupied addresses can be produced by linking records according to their unique property reference number (UPRN). It should be noted that UPRNs do not always refer to single specific households. For multi-occupancy properties, a UPRN may contain several households. In MSOAs with a high proportion of such properties, mean occupied address incomes may be inflated above true mean household incomes.

This method relies on the availability of administrative data and could be further developed by incorporating new data sources (as discussed in Section 4: Comparison of income components).

4 . Comparison of income components

The Canberra Group Handbook on Household Income Statistics created by the United Nations Economic Commission for Europe is widely agreed as the international standard for household income statistics. It provides detailed guidance on how components of income should be defined and measured. Comprehensively describing each standard income component, it serves as a useful checklist to verify the availability and completeness of all components to derive admin-based income statistics (ABIS).

Small area income estimates (SAIE) are currently produced for gross income (before tax and national insurance), net income (after tax and national insurance) and net income after housing costs.

Gross income

Gross income is the income received before the deduction of Income Tax and National Insurance contributions. Gross income, as captured in ABIS, consists of the following components and data sources. The annotations, such as "1a", refer to the category set out in page 11 of the Canberra Group Handbook on Household Income Statistics:

- wages and salaries (1a) – Pay As You Earn (including capturing employed and occupational private pensions income)

- income from self-employment (1b) – Self Assessment

- benefits and cash transfers (4a) – including Attendance Allowance, Disability Living Allowance

- pensions and other insurance benefits (4b) – Pay As You Earn (capturing private pension income)

- Social Assistance benefits (4c) – Department for Work and Pension's (DWP) Benefit Claimant Counts

Gross income as captured in ABIS does not currently include:

- property income (2a, 2b, and 2c) – not currently incorporated; available from Self Assessment but still to be added to the measure

- current transfers from non-profit institutions (4d)

- current transfers from other households (4e)

The modelling approach to SAIE and use of the administrative data alongside survey data means all components listed above are captured in SAIE.

The net value of owner-occupied housing services is not currently captured by SAIE or ABIS measures.

Net income

Net income, as captured in ABIS, is calculated by deducting direct taxes (8a) (Income Tax) and social insurance contributions (8d) (National Insurance contributions). These are derived from PAYE data using HM Revenue and Customs (HMRC) guidelines for the respective tax year. SAIE also captures both components.

Unlike SAIE, ABIS does not currently include current transfers to non-profit institutions (8e), a deduction needed for Net Income.

Compulsory fees and fines (8b) and inter-household transfers paid (8c) are captured by neither ABIS nor SAIE.

Net income after housing costs

Net income after housing costs is calculated by deducting rent, council tax, water, and service charge from net income. These components are captured in SAIE. However, the current ABIS methodology does not include admin data relating to housing costs. Additional admin data could enable estimates of net income after housing costs to be made.

The missing income components may result in ABIS being lower than true income levels. For example, national statistics on income, based on data collected through the Household Finances Survey, show that in the financial year ending (FYE) 2018, it was estimated that 3.4% of gross income constituted "investment income", which includes property income. For more information, see our Effect of taxes and benefits on household income dataset.

If the distribution of those missing components doesn't align to the geographic and demographic distribution of ABIS, conclusions on the distribution of income may be incomplete. This may lead to an understatement of the income for households which most rely on the components of income which are missing from ABIS.

The timeliness of the administrative datasets is an important consideration. Where possible, all datasets used should reflect the same and most recent period. This is particularly important for capturing Universal Credit and not duplicating the recorded benefits by also capturing the benefits replaced by Universal Credit (such as child tax credit and housing benefit) received in previous years.

Table 6 of the accompanying dataset includes further details of the Canberra Handbook components of income and their inclusion in ABIS and SAIE. For further information on the differences between ABIS and SAIE from a coherence viewpoint, see our Income statistics: Coherence and comparison information (PDF, 325KB).

5 . Comparison of population coverage

The admin-based income statistics (ABIS) and the income estimates for small areas (SAIE) are both based on the "usually resident" population of England and Wales and derived using different population datasets. The sampling frame for the Family Resources Survey (FRS), which underpins SAIE measure (for Great Britain) uses Royal Mail's Postcode Address File (PAF) to draw the survey sample from. Census 2011 and mid-year population estimates are used in the FRS grossing regime to produce control totals.

ABIS however, uses the Statistical Population Dataset (SPD), SPD used in our 2018 ABIS being the latest iteration available at the time of production (V3.0). Admin-based population estimates (ABPEs) based on the SPD are consistently higher (by up to 2.0%) than the mid-year estimates (MYE) over 2016 to 2020. Differences between the ABPEs and the mid-year estimates are a net effect of overcoverage and undercoverage of different population groups in the ABPE, and greater uncertainty in the MYEs as we get further away from the 2011 Census. Further information regarding our latest SPD data and methodology can be found in our Developing admin-based population estimates, England and Wales: 2016 to 2020 article. The SPD is under development and forms part of our population and migration statistics transformation programme, which aims to provide users with consistently high-quality population statistics every year.

There will be a minority of individuals who appear on the census but are not included in any of the administrative datasets brought together to create the SPD.

Income is often analysed at the household level. As discussed in Section 3: Comparison of methodological approaches, statistics on income are produced for occupied addresses by linking records. As not all members of the household may have income information identified, it is possible that an occupied address has only partial income information identified, resulting in a lower income statistic than the true income.

Table 1: Count of individuals and households within the statistical population database (SPD) and with specific income components identified
England and Wales, financial year ending 2018

England and Wales	Individuals (number)	Individuals (proportion)	Occupied addresses (number)
SPD v3	54,257,000	100.0	22,425,000
PAYE	33,303,000	61.4	19,452,000
Benefits	21,941,000	40.4	15,466,000
Self Assessment	2,615,000	4.8	2,459,000
Any income component	41,078,000	75.7	22,104,000

Source: Office for National Statistics

Table 1 shows the level of coverage of different income components used to produce ABIS. Income information was able to be identified from at least one data source for 92.8% of individuals aged 16 years and over in England and Wales (on the SPD).

ABIS almost completely covers the SPD (98.6%) in terms of representation of households. ABIS may also include people within non-household settings (such as care homes, halls of residence) which SAIE does not. Allowing for this, the true rate of coverage may be slightly lower than this figure.

At least one source of income information is available for 75.7% of individuals of all ages in England and Wales in the population, which is somewhat lower as many households contain individuals not receiving income. Around 7.2% of individuals aged 16 years and over in the SPD have not had any income information identified. This may reflect that individuals have a genuine zero income for the period, or that their income has not been captured by the current methodology. This could be because of the nature of the income, for example, cash in hand income. Alternatively, Section 4: Comparison of income components discussed a number of additional income components that may be included for a more complete measure. There may also be individuals who do receive income from the components captured, but their income has not been matched to the SPD because of missing or unavailable pseudonymised identifiers in the source data.

According to our [Housing in England and Wales bulletin](#), the majority (22.1 million occupied addresses out of an England and Wales Census 2021 population of 24.8 million households) in England and Wales have at least one source of income information identified for them. This equates to 89.2%, slightly lower than the 98.6% coverage of the SPD.

The number of individuals of all ages with income information represents 69.5% of the 2018 MYE of the population of England and Wales, according to our [Estimates of the population for the UK, England, Wales, Scotland, and Northern Ireland dataset](#).

This section continues by exploring the coverage of income information captured by ABIS by various demographic and geographic breakdowns. Figure 1 shows how the coverage of the SPD and of the ABIS income information varies by region.

Figure 1: Proportion of occupied addresses in the statistical population database (SPD) without identified income information by region

England and Wales financial year ending 2018

Figure 1: Proportion of occupied addresses in the statistical population database (SPD) without identified income information by region

England and Wales financial year ending 2018
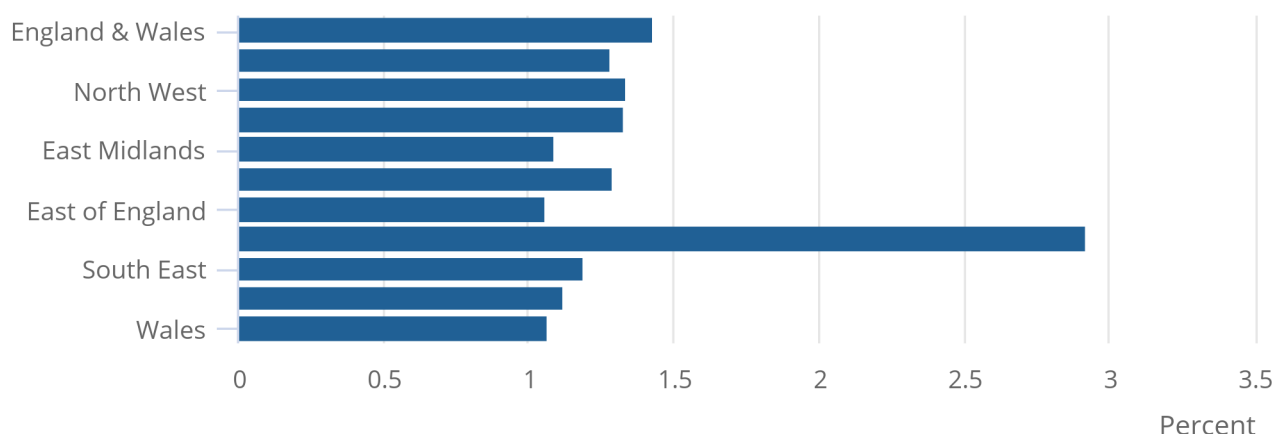
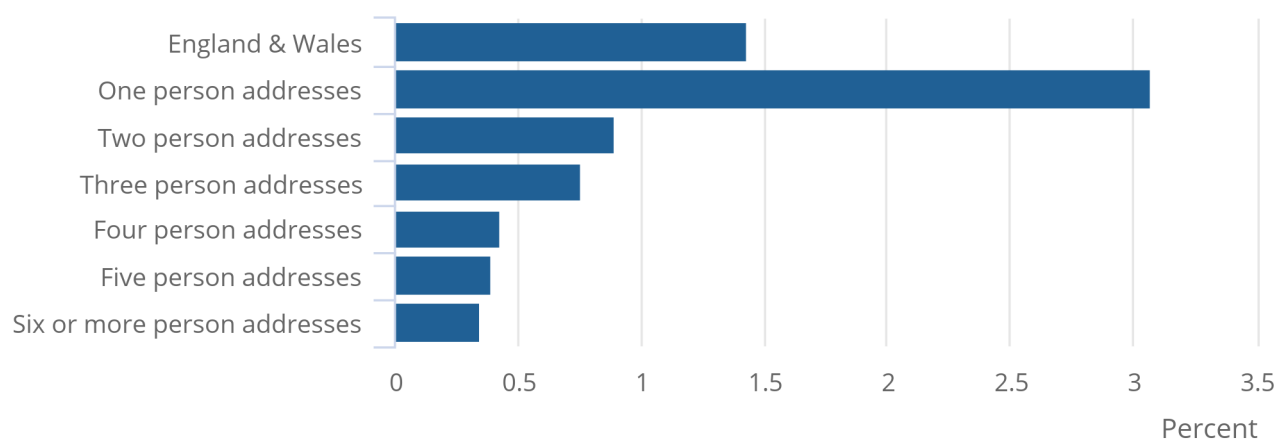**Source: Office for National Statistics**

There are around 1.4% of occupied addresses in our SPD population base where we have not been able to identify any income information for in the ABIS data sources. Figure 1 shows that for all regions, this falls between 1.0% and 1.5%, the exception being London, which is much greater at 2.9%. Figure 2 shows the proportion of occupied addresses in the SPD with no income information by household size.

Figure 2: Proportion of occupied addresses in the statistical population database (SPD) without identified income information by household size

England and Wales financial year ending 2018

Figure 2: Proportion of occupied addresses in the statistical population database (SPD) without identified income information by household size

England and Wales financial year ending 2018

Source: Office for National Statistics

Over 3.0% of single-person occupied addresses (on the SPD) has no income information identified. This is higher than occupied addresses with greater numbers of individuals. However, it should be noted that the chart presents any income identified for the occupied address, and not only occupied addresses where every member has income information identified. There could be components of income that have not been able to be matched to the occupied address within the SPD.
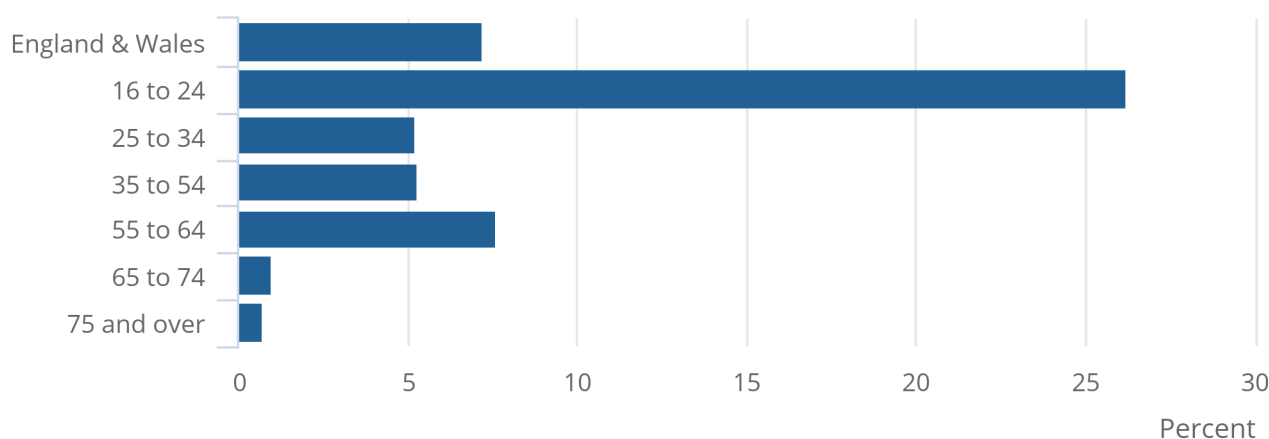
Figure 3 shows ABIS coverage by age.

Figure 3: Proportion of individuals in the statistical population database (SPD) without identified income information by age group

England and Wales financial year ending 2018

Source: Office for National Statistics

The proportion of individuals without identified income information is highest for the age group 16 to 24 years. Regarding individuals, ABIS shows lower coverage for younger ages partly because of the higher proportions of young people in full-time education. Our article describing Feasibility research into admin-based labour market status (ABLMS) for the tax year ending 2016 supports this explanation. It finds that the proportions of ABLMS "Unallocated" and "Economically Inactive: Student" were high for 16- to 17-year-olds, decreasing through the 18 to 24 years and 25 to 34 years age groups.

Coverage remains stable during ages that individuals are typically in the labour market, and rises slightly around the state pension age reflecting the inclusion of state pension admin data as a data source. Figure 4 shows the proportion of the SPD aged 16 years and over with income information identified and the proportion of earnings identified in the total ABIS, by age group.

Figure 4: Proportion of individuals in the statistical population database (SPD) and total income information by age

England and Wales financial year ending 2018

Figure 4: Proportion of individuals in the statistical population database (SPD) and total income information by age

England and Wales financial year ending 2018
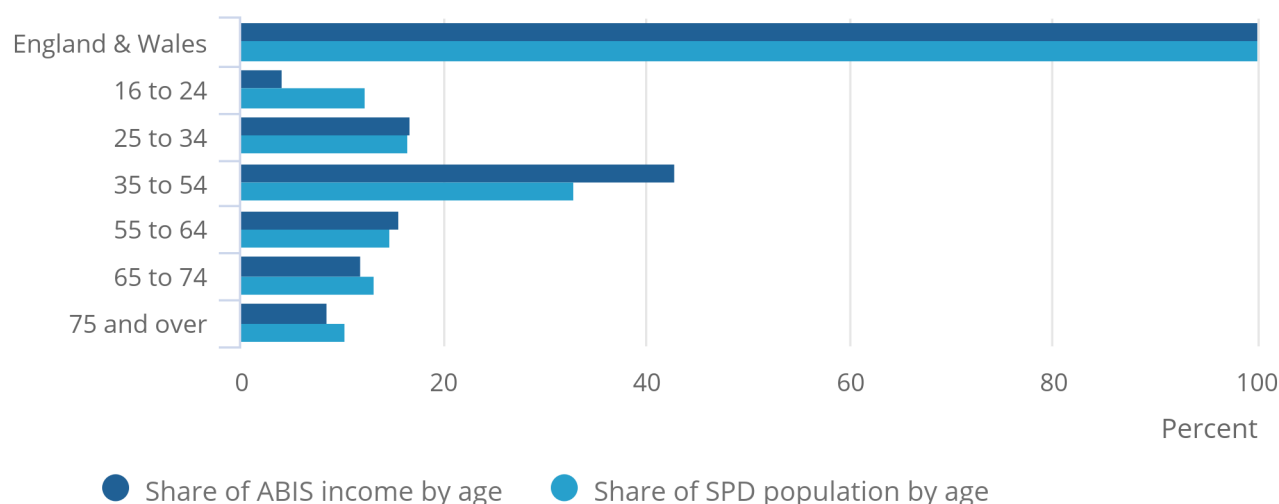
Source: Office for National Statistics

Figure 4 shows that 4.3% of ABIS income is attributed to people aged 16 to 24 years, a group which represents 12.4% of the population. Older age groups show a higher share of the ABIS income relative to their population. Although males and females are split equally in the population and in the ABIS and SPD datasets, 58.0% of total income earned are attributed to males on ABIS.

6 . Comparison of income measures

This section compares published admin-based income statistics (ABIS) with model-based small area income estimates (SAIE) to explore the extent to which the components identified in Section 4: Comparison of income components as missing and population coverage differences in Section 5: Comparison of population coverage lead to differences in the resulting income estimates. Although the most recent release of SAIE relates to financial year ending (FYE) 2020, at the time of carrying out this comparison, the latest available ABIS estimates are for FYE 2018 (because not all of the admin data components needed to obtain equivalent ABIS estimates were available for FYE 2020). Therefore, the comparison is between estimates for the FYE 2018.

As discussed in earlier sections, there are several income components without identified data sources and are missing in the ABIS. However, SAIE is underpinned by survey data with known limitations around sample size and survey related biases. A demonstration of agreement between the two measures provides evidence of the potential for admin-based direct estimates to replace model-based estimates in the future.

Figure 5: Household income by Middle-level Super Output Areas (MSOA) according to admin-based income statistics (ABIS) and small area income estimate (SAIE) approaches

Gross household income, England and Wales, financial year ending 2018

Download the data

Figure 5 shows that the two sets of income estimates are broadly similar at the Middle-layer Super Output Area (MSOA) level. However, ABIS incomes show a long, positively skewed tail, unlike SAIE incomes where ABIS estimates are higher than SAIE. This difference reflects Pay As You Earn (PAYE) incomes, and to a lesser extent self-employed income, which are also positively skewed, rather than the benefits component, which is more normally distributed. There are several possible reasons for the differences, including whether the modelling process for SAIE (particularly the constraint to the mean for England and Wales) is artificially reducing estimates of high incomes, whether income is understated at the highest levels within the survey, and whether low earners in high-income MSOAs are represented on ABIS, which in turn, may increase the mean ABIS incomes in these areas.

To compare the distribution of incomes according to each measure, each MSOA can be classified into a decile according to the gross income estimated by each measure. Table 2 shows the number of MSOAs allocated to each ABIS decile and SAIE decile.

Table 2: Number of Middle-level Super Output Areas (MSOAs) in each admin-based income statistics and small area income estimates decile
England and Wales, financial year ending 2018

ABIS Decile	1 - least prosperous	2	3	4	5	6	7	8	9	10 - most prosperous	Total
1 - least prosperous	483	185	36	7	4	1	2	2	0	0	720
2	186	301	174	37	17	3	1	0	1	0	720
3	55	169	273	154	43	22	4	0	0	0	720
4	13	46	183	260	147	40	24	5	1	1	720
5	2	10	43	174	257	162	57	8	6	1	720
6	1	3	18	52	170	257	168	41	10	0	720
7	1	1	0	15	69	160	233	166	65	10	720
8	0	0	1	2	10	61	163	253	183	47	720
9	0	1	1	1	0	12	62	196	281	166	720
10 - most prosperous	0	0	0	1	2	5	10	47	164	492	721
Total	741	716	729	703	719	723	724	718	711	717	7201

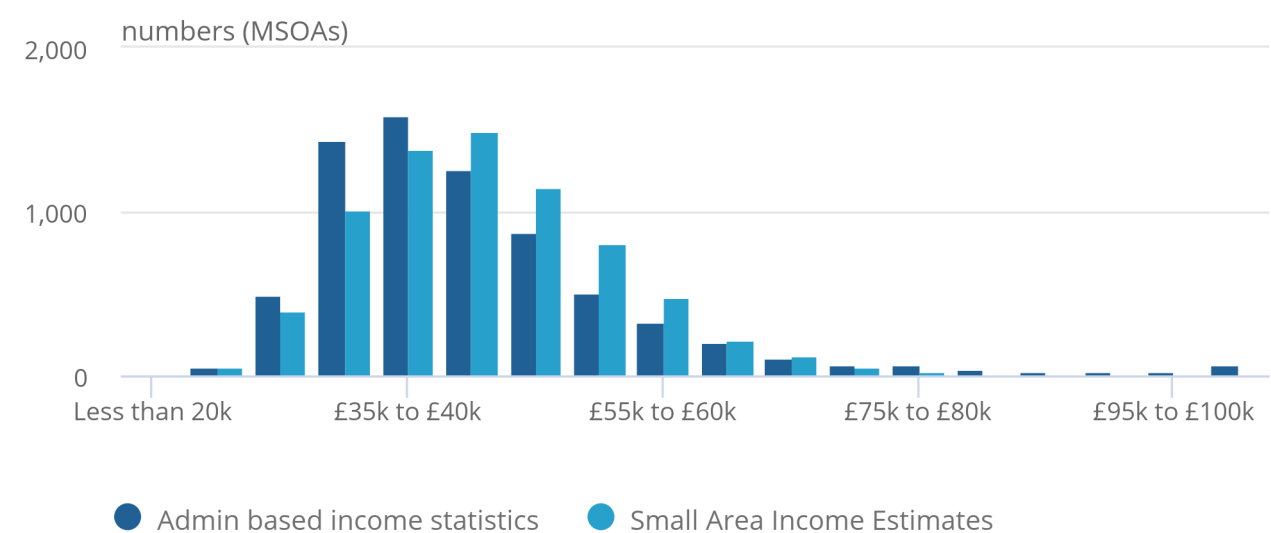Source: Office for National Statistics, Department for Work and Pensions

Table 2 shows that most households classified by ABIS as being in the bottom decile are also classified by SAIE as being in the bottom decile. However, there are several MSOAs which are classified in different deciles across the two measures. This trend is seen in each decile. Around 43% of all MSOAs fall within the same ABIS decile as SAIE, with a further 43% being one decile higher or lower. The total ABIS income across all MSOAs in England and Wales is within 1% of that from SAIE and hence from the Department for Work and Pensions (DWP) Family Resources Survey (FRS).

Figure 6: Frequency of Middle Super Output Areas by admin-based income statistics-derived and Small Area Income Estimates-derived household gross income band

England and Wales financial year ending 2018

Figure 6: Frequency of Middle Super Output Areas by admin-based income statistics-derived and Small Area Income Estimates-derived household gross income band

England and Wales financial year ending 2018

Source: Office for National Statistics

Figure 6 compares the distribution of the gross household incomes across MSOAs for ABIS against SAIE.

There are several MSOAs (255, 3.5%) which have a mean household gross income measured by ABIS exceeding £75,000, compared with much fewer from SAIE (32, 0.4%).

It shows that relative to SAIE, ABIS tends to show a greater number of higher incomes. It is possible that the differences seen here may be amplified because of the fact that the SAIE modelling approach may understate the highest incomes by compressing the estimates to England and Wales mean.

ABIS estimates show more low-income MSOAs (for example, those with less than £25,000) than SAIE, likely reflecting households containing individuals whose incomes are not identified in the current ABIS data sources, including cash in hand income below the tax threshold. These incomes may not be captured in PAYE, bringing overall household ABIS incomes down.

7 . Admin data to derive small area income estimates data

Exploring the use of admin data to derive small area income estimates
Dataset | Released 30 April 2024
Update on research exploring how well admin-based small area income statistics compare with accredited official statistics and align to international best practice.

8 . Glossary

Canberra Handbook

The [Canberra Group Handbook on Household Income Statistics, Second Edition (2011)](#), provides a consolidated reference for those involved in producing, disseminating or analysing income distribution statistics. It reflects the current international standards, recommendations, and best practice in household income measurement. Comprehensively describing each possible income component, it serves as a useful checklist to verification of the availability and completeness of all components to derive admin-based income statistics (ABIS).

Occupied addresses

The addresses of individuals included in the Statistical Population Dataset (SPD) have been referenced to a unique property reference number (UPRN). Individuals with the same UPRN are grouped to form "households" on the basis that they are resident at the same address.

Statistical Population Dataset

The SPD, is the frame to which the various ABIS data sources are linked. It aims to approximate the usually resident population down to small areas with admin data. The most recent version available is in our article [Developing SPD, England and Wales, V4.0 for 2021](#), which is broadly in line with Census 2021 estimates, [set out in our Understanding quality of the SPD article](#). SPD V3.0, described in our [Developing our approach for producing admin-based population estimates: 2011 and 2016](#), has been used for the analyses in this article.

9 . Data sources and quality

Admin-based income statistics dataset

Our [2018 Admin-based income statistics (ABIS) article](#)was produced using the following administrative data sources:

- Department for Work and Pensions's (DWP) National Benefits Database

- DWP's Single Housing Benefit Extract

- DWP's Personal Independence Payment

- DWP's Universal Credit

- HM Revenue and Customs's (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

More information can be found in[Section 6 of our ABIS Quality and Methodology Information (QMI) report](#),for detail on the methods used to produce the ABIS.

Income estimates for small areas dataset

Our [2018 Income estimates for small areas (SAIE) dataset](#) was produced through a model-based method, using a combination of the following survey and administrative data sources:

- Family Resource Survey (FRS)

- 2011 Census

- Department for Work and Pension's (DWP) Benefit Claimant Counts

- Valuation Office Agency (VOA) Council Tax Bandings

- ONS, house price statistics for small areas

- Department of Energy and Climate Change, energy consumption data

- HM Revenue and Customs (HMRC), Pay as You Earn (PAYE) data

- regional or country identification variable

More information can be found in Section 2 of our Income estimates for small areas in England and Wales, technical report: financial year ending 2018.

10 . Future developments

While analysis has shown that income information has been identified for 92.8% of persons aged 16 years and over on the Statistical Population Dataset (SPD), there are still a number of missing components, and groups within the population. Available admin-based income statistics (ABIS) data sources will go a long way towards producing small area estimates of gross and new incomes before housing cost deductions, but there are still significant gaps to fill regarding housing costs, especially if they take time to source in production. There is additional work needed to capture the missing components by including additional data sources in the methodology.

Further work should also be undertaken to consider the ONS Statistical Population Dataset (SPD) population relative to the census population. Consequently, households and individuals who are more likely to be excluded may not have similar earning patterns to those who are on ABIS. Grossing up existing data through these gaps will lead to biases.

Longer term, we have identified stakeholder needs for local area income statistics to go beyond the existing measure of Middle-layer Super Output Area (MSOA) level means and would include medians and specific percentiles of income, along with the proportion of households which are below certain thresholds, such as below the national 60% median. A similar set for lower level geographics (for example, Lower-layer Super Output Areas (LSOAs)) would be developed.

Being able to extend the new process to include incomes net of housing costs will be critical, as one of the main metrics requested. If ABIS-sourced housing cost items are not available, then a statistical model may be developed to cover this. Where non-coverage of households still exist, appropriate weighting and calibration methods will be explored to optimise its projection the true census population.

Work may be needed to improve the linkage between the increasing number of income components being added, whether it is possible to resolve the difference between unique property reference number (UPRN) and "true households" and to extend this to include Scotland and Northern Ireland.

11 . Related links

[Developing methods to produce Census 2021 income data, England and Wales: June 2023](#)
Article | Released 28 June 2023
An update on progress towards developing methods to produce Census 2021 income data for England and Wales.

[Population and migration statistics transformation in England and Wales, research overview:2023](#)
Article | Released 26 June 2023
A summary of our research on the future of population and migration statistics in England and Wales, underpinning our consultation on the proposed new system.

[Developing our approach for producing admin-based population estimates, England and Wales: 2011 and 2016](#)
Article | Released 21 June 2019
Research into developing a new methodology to create population estimates from administrative data.

[Understanding quality of the Statistical Population Dataset in England and Wales using the 2021 Census - Demographic Index linkage](#)
Article | Released 28 February 2023
Analysis of Statistical Population Dataset version 4.0 2021 using a linkage between Census 2021 and the Demographic Index.

[Income statistics: Coherence and comparison information (PDF, 324KB)](#)
Report | Released April 2024
Guidance providing coherence and comparison information on important criteria for income statistics produced by the Office for National Statistics (ONS), Department for Work and Pensions (DWP), and HM Revenue and Customs (HMRC). This information helps users to understand and assess the differences between these statistics.

12 . Cite this article

Office for National Statistics (ONS), released 30 April 2024, ONS website, article, [Exploring the use of admin data to derive small area income estimates, England and Wales](#)