Research on producing admin-based estimates of the number and size of households for local authorities in England and Wales

Date 13th November 2025

1. Key Messages of Paper

Purpose

This paper provides a summary of research to date on producing household statistics from administrative data. It sets out the proposed methodology for producing the admin-based household estimates for the number and size of households in local authorities in England and Wales. It also outlines proposed future research on this topic.

Key Asks of MARP

- Do the panel have any key methodological concerns that must be addressed before we consider publishing admin-based household estimates (ABHE) for 2021 in December (number but not size estimates)?
- Do the panel have any feedback on the proposed actions based on the prior review of this work?

2. Executive summary

Work to date on producing household statistics from administrative data primarily focused on the production of a record-level dataset that aims to assign people to addresses. This method groups people based on their Unique Property Reference Number (UPRNs) in their administrative data. Rather than the Census definition of a household, outputs from this dataset have occupied addresses as the statistical units. Outputs have been produced on the number and size of occupied address for local authorities (LAs) in England and Wales. These outputs were raw counts produced directly from the record-level data.

The work presented in this paper is to produce admin-based household estimates. The methodology aims to adjust for coverage issues, linkage error and UPRN mis-recording. It also attempts to align our statistics with the established definition of a household. These estimates are coherent with the admin-based population estimates. It is our intention to publish these estimates as official statistics in development when appropriate. Further work will be needed to develop these estimates to ensure that they are able to fully meet users' needs.

Scope

The scope of the work in this paper is statistics on the number and size (number of usual residents) of households for LAs in England and Wales. Estimates of

household composition, e.g. age or sex of residents, and families[1] are currently out of scope but will need to be addressed in the future to fully meet the user needs expressed below.

People living in communal establishments (CEs) need to be considered as part of the work to produce statistics on households. However, the production of estimates of the CE population is also out of scope for this paper.

Statistics about the characteristics of people within households, also expressed as part of the user needs, is out of scope here. However, they will be considered as part of the wider ONS population statistics research priorities.

Background

The Census is currently the main data source for local-level statistics on the number, size, and composition of households. During intercensal periods, statistics on households are produced using the Labour Force Survey (LFS). However, there are restrictions on the geographic breakdowns that can be produced due to sample sizes. This work explores the feasibility of producing statistics on households at an LA level by using administrative data supported by survey data.

User needs

In 2023, the ONS ran a consultation on the future of population statistics in England and Wales. Users demonstrated a clear need for household estimates which were timelier than census data but more granular than survey data. Respondents stated a need for estimates of the number and size of households at various levels of geography, including local authority (LA) level. They also expressed a need for estimates of various kinds of living arrangements and family and household compositions, again for low-level geographies. This included information on specific types of households such as:

- single parents
- older people living alone
- multi-generational households
- co-residents
- blended families
- concealed households
- houses of multiple occupancy

Respondents also stated a need for information on relationships between household members. Respondents highlighted the importance of being able to use data on households in combination with information on personal characteristics and housing. Work to gather use cases and information about policy needs is ongoing including the 2031 Census topic consultation. This will continue to direct the development of future research.

[1] Statistics on household composition and families report on the structure or relationships between members of a household.

Definitions

In the census a household is defined as “one person living alone or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room or sitting room or dining area”. In administrative data, it is not possible to identify whether residents at an address share rooms and cooking facilities. Outputs produced directly from administrative data therefore use occupied addresses as the statistical unit rather than households. Using Census 2021 data, we have estimated that 0.31% of occupied addresses contain multiple households. The impact of the difference in statistical unit should be minimal. However, this does affect some population groups or geographic areas more than others. For example, occupied addresses with multiple households are more common in London than in other areas.

Our ambition is to produce statistics that meet user needs. However, we will need to balance what is possible from administrative data against what users require. Consultation respondents were generally negative about a potential move away from using households as the statistical unit. As part of this research, we have attempted to bridge the definitional divide between household and occupied addresses by incorporating survey data through statistical estimation. We will also conduct further engagement with users to direct the way in which we develop the estimates using definitions that would best meet user needs for statistics on how people live.

3. Work to date

Admin-based living arrangements dataset

The work to date has focused on creating a record-level linked administrative dataset called the admin-based living arrangements dataset (ABLAD). This is created as follows:

- The Address Index Matching Service (AIMS) is used to add a Unique Property Reference Number (UPRN) variable to all administrative data sources containing full address information. This is completed by ONS’s Data Engineering as part of the initial data processing.
- UPRNs from the Personal Demographic Service (PDS), English School Census (ESC), Individualised Learner Record (ILR) and Lifelong Learning Wales Record (LLWR) are joined to the statistical population dataset (SPD) using a unique identifier assigned by ONS as part of the reference data management framework (RDMF) (ONS, No Date). The SPD is a linked administrative dataset which uses a rules-based method to approximate the usually resident population of England and Wales. This is done by looking for specific types of activity within the data, sometimes referred to as “signs of life”.
- If an individual appears multiple times in a data source, we select one UPRN per data source. This is based on date information and the quality of the UPRN match.
- Then if an individual appears across multiple data sources, we use a dataset hierarchy to select one UPRN per person. This is broadly ESC, LLWR, ILR, PDS

but with some additional aged-based rules. This hierarchy was constructed based on findings from record-level comparisons with Census 2021 data.
- The ONS Address Frame is joined to SPD using the assigned UPRN. This adds geographical information, a variable that classifies UPRNs into households and communal establishments (address type), and an establishment type variable for communal establishments.
- Information from Higher Education Statistics Agency (HESA), Ministry of Justice (MoJ) and ESC data is then joined on. This is used to identify students living in halls of residence, prisoners and pupils boarding in English state schools. The address type and establishment type are then updated for these people.

The ABLAD is then split into four sub-datasets:

- Records with the address type of “household” are put into the admin-based occupied address dataset (ABOAD). In the latest version of the ABLAD (V5.1) for 2021 there were 53,507,145 records (95.12%) in the ABOAD.
    - The ABOAD contains 44,355 people living in implausibly large occupied address (31+ usual residents). In the future we plan to assess the impacts of these large occupied addresses on the ABHEs.
- If the address type is “communal establishment” these records are put into the admin-based communal establishment dataset (ABCED). In ABLAD v5.1 for 2021 914,345 records (1.63%) went into the ABCED.
- Some records have a UPRN, but the UPRN is not for a residential address. These records are put into a ‘non-address frame UPRN’ dataset. In ABLAD v5.1 for 2021 there were 1,521,695 records (2.71%) had a non-residential UPRN.
- If the record does not have a UPRN at all they are put it into a ‘no UPRN’ dataset. In ABLAD v5.1 for 2021 there were 305,920 records (0.54%) that did not have a UPRN.

The following outputs have been produced from the ABOAD:

- Number of occupied addresses
- Number of occupied addresses by size (number of usual residents)

These outputs have been produced directly from the record-level dataset. They do not have any adjustment for coverage issues, linkage errors or UPRN mis-recording.

4. Current Work

Work on the living arrangements topic is currently focused on producing admin-based estimates of the number and size of households for LAs in England and Wales. To produce admin-based household estimates (ABHEs) we propose using a multi-stage estimation approach:

1. To produce a national admin-based household population estimate by age and sex, by identifying and removing the communal establishment population from the admin-based population estimates (ABPEs).
2. To produce an admin-based number of households estimates for local authorities by converting the household population to the number of households.

3. To produce an estimate of the size of households by local authority.

Over the past 12 months the living arrangements estimation team have worked through an ambitious work plan. This was set out in our paper to MARP in May 2024 (Morgan *et al*., 2024).

During this work, we tested different methodologies for producing each stage of the ABHEs. The different methods at each stage were compared to the 2021 Census and LFS household estimates, both of which are accredited official statistics for households. This analysis was used to select the most successful method at each stage. At each stage of the process the living arrangements estimation team produced an internal paper and held a working group with technical experts. These allowed us to outline our results and propose a method to take forwards for that stage of the process. The chosen methods were then presented to ONS's senior leaders for sign off.

At the end of the methods selection process in July 2025, we produced a paper for the methodology and research assurance group (MaRAG). The report detailed the research that we had undertaken. It also proposed selected methods for producing estimates on the number and size of households in England and Wales at an LA level.

Stage one – producing an admin-based estimate of the household population of England and Wales

The admin-based population estimates (ABPEs) are an estimate of the total population of England and Wales. They are disaggregated by local authority (LA), sex, and birth cohort. To produce estimates of the household population which are coherent with the ABPEs we first need to remove the communal establishment (CE) population from the ABPEs. This is a similar requirement to removing CE residents from the official mid-year population estimates for producing survey weights.

We did this by calculating the proportion of those in the admin-based communal establishment dataset (ABCED) to the total of those in the ABCED and the admin-based occupied address dataset (ABOAD) combined. Then applying this proportion to the ABPEs to calculate the CE population ($\hat{C}_{A\beta}$). This was done at a national level disaggregated by age and sex.

$$\hat{C}_{A\beta} = \left(\frac{C_{A\beta}}{C_{A\beta} + O_{A\beta}}\right) \times \hat{P}_{A\beta}$$

where:

β currently is each combination of single year of age and sex
A is administrative data
$\hat{C}_{A\beta}$ is the admin-based (A) estimated CE population for subpopulation β
$C_{A\beta}$ is the CE population count from the admin-based communal establishment dataset (ABCED) for subpopulation β

$O_{A\beta}$ is the admin-based occupied address population count from the admin-based occupied address dataset (ABOAD) for subpopulation β
$\hat{P}_{A\beta}$ is the admin-based population estimate (ABPE) for subpopulation β

We then took the ABPEs disaggregated by age and sex and removed the previously calculated CE population ($\hat{C}_{A\beta}$). This was then summed to produce a national admin-based households population estimate (ABHPE, $\hat{h}_{A\beta}$).

$$\hat{h}_{A\beta} = \hat{P}_{A\beta} - \hat{C}_{A\beta}$$

$\hat{h}_{A\beta}$ is the admin-based household population estimate

Stage two – producing an admin-based estimate of the number of households for each local authority in England and Wales

The aim of stage two is to convert estimates of the household population (ABHPEs, $\hat{h}_{A\beta}$ into the admin-based number of households estimate (ABNHE) at a national ($\widehat{H}_A$) and LA ($\widehat{H}_{Al}$) level.

We started by calculating the household representative rate (HRR, $\rho_{S\beta}$) for England and Wales. This was produced from the household weighted annual population survey (APS) data[2]. HRR is the proportion of people in each demographic group, in this case single year of age and sex combination, that were the household reference person (HRP) for their household in the APS. The HRP is the person chosen to be the representative of a household. They are usually the oldest economically active person in the household (ONS, 2008).

$$\rho_{S\beta} = \frac{R_{S\beta}}{P_{S\beta}}$$

S is household weighted data from surveys
$\rho_{S\beta}$ is the household representative rate for subpopulation β from the APS
$R_{S\beta}$ is the number of household reference persons for subpopulation β from the APS
$P_{S\beta}$ is the estimated subpopulation β from the APS

This household representative rate ($\rho_{S\beta}$) was then applied to the ABHPE disaggregated by age and sex ($\hat{h}_{A\beta}$). This produced the national number of households for each age and sex combination ($\widehat{H}_{A\beta}$). This was then summed to produce an estimated number of households in England and Wales ($\widehat{H}_A$).

[2] For operational reasons APS was used in this stage over LFS but consideration is being given as to how to utilize LFS data in the future.

$$\widehat{H}_A = \sum_{\beta \in B} (\rho_{S\beta} \times \hat{h}_{A\beta})$$

$\widehat{H}_A$ is the admin-based number of households estimate

Using the ABOAD, we produced the distribution of occupied addresses across LAs. We did this by dividing the number of occupied addresses in each LA (O_{Al}) by the total number of occupied addresses in England and Wales (O_A). The national number of households for England and Wales ($\widehat{H}_A$) is then distributed across local authorities based on the proportion of occupied addresses in each LA. This produces the admin-based number of households estimate for local authorities ($\widehat{H}_{Al}$).

$$\widehat{H}_{Al} = \widehat{H}_A \times \frac{O_{Al}}{O_A}$$

$\widehat{H}_{Al}$ is the admin-based estimates number of households per local authority
O_{Al} is the admin-based occupied address population count for LA l;
O_A is the national admin-based occupied address population count.

Stage three – producing an admin-based estimate of the size of households for local authorities in England and Wales

The final stage estimates the number of households by household size for LAs in England and Wales. To do this we used a structure preserving estimator (SPREE) model. SPREE is a type of small area estimation technique which re-weights a cross tabulation to desired row and column totals. This is done using iterative proportional fitting (IPF). The method requires:

- A crosstabulation of the variable of interest known as a proxy structure. This is a detailed breakdown of the variable of interest that is either out of date or has a slightly different definition than the desired variable.
- True row and column benchmarks (allocation structure) which are known totals or sufficiently accurate estimates of the variable of interest. To adhere to statistical best practice, these benchmarks must sum to the same total.

The estimates are obtained by iteratively adjusting the proxy structure to fit the allocation structure. Initially the rows of the proxy table are scaled to match the row totals in the allocation structure. This will mean that the rows of the proxy table will sum to the row totals of the allocation structure. However, the columns will not match. Next the columns of the proxy table are scaled to match the column totals of the allocation structure. This will mean the column totals now match between the proxy and the allocation structure. However, the row totals will no longer match. These two steps are repeated until the model converges. This happens when both the rows and columns match between the allocation structure and the proxy data at the end of the second step.

In our model we used a cross tabulation of occupied address by size and LA from the ABOAD as our proxy data. This data is up to date but has a different definition (occupied address) to our desired variable (households).

For the allocation structure we used the number of households by LA ($\widehat{H}_{Al}$) produced from stage two of the estimation process as the row benchmarks. For the allocation structure column totals we used the LFS official statistics on household size. However, as this data would not match the row benchmark total, we calculated the proportion of each household size in the LFS estimates and applied these proportions to the total number of households in England and Wales ($\widehat{H}_{A}$) produced in stage two. Although it would be more typical to scale the rows up to the column totals as for this model the column totals are at a higher level of geography than the rows and therefore assumed to be of better quality. However, this was not possible in this project as we are using the benchmarks to constrain the model outputs to the admin-based household estimates.

Previous work was done to assess using SPREE to produce an admin-based estimate of household size (ONS, 2017).

Prior review

A paper outlining the research to date on this project has already been reviewed by MaRAG and methodologists working in the Methodology and Quality Directorate at ONS. Below is a summary of the feedback, suggested actions and progress.

	Feedback – suggested actions	Progress
1	Conduct analysis of the people whose UPRN is neither an occupied address nor a communal establishment i.e. those not assigned to a residential address (unassigned group). The analysis should assess if they are similar to those found in admin data i.e. those assigned a residential address (assigned group).	Completed and findings supplied to MaRAG. We found: • That the proportion of female CE residents in the unassigned group was higher than in the assigned group. • A higher proportion of those aged 19 to 21 years old lived in households in the unassigned group than in the assigned group. • For the oldest age group (90+) we found a higher proportion of CE residents in the unassigned group than in the assigned group.
2	Use 2021 Census data as the starting point to apply method two or method three. To identify where quality issues are introduced and asses if they are propagating through different estimation stages.	Completed for selected method in stage two. Added to the backlog of work post-publication for method three.
3	Complete a bias analysis of the survey data to be used in the methods.	Completed and reviewed by ONS methodologists.

4	Produce a short time series	Completed and analysis of the results is ongoing.
5	In stage two we use the household representative rates from the APS for each age and sex group. Concerns have been raised about the potential for differential response rate for different age and sex combinations. This may not be fully addressed by the survey weighting. Analysis should be done to investigate the bias of using this HRR. We could also trial using the HRR from the 2021 Census. This is what the ONS's household projections use.	This work is on our post-publication backlog.
6	In stage two we use the proportion of occupied addresses in each LA from the ABOAD. This is used to distribute the national number of households across LAs. Although the number of occupied addresses which have multiple households in them is small (estimated to be 0.31%) this does vary by LA. There are concerns that this could lead to an underestimation of the number of households in LAs where multi-household occupied addresses are more common.	We trialled a number of different methodologies for producing the admin-based number of households estimate. This included distributing the national number of households using the percentage of households in each LA according to survey data. The method outlined in this paper produced estimates which were closest to the 2021 Census estimates. We planned to test using dual system estimation to estimate the number of households missing from our admin data. This method has been trialled previously by ONS and show promise in this area (ONS, 2018). This was unable to be completed as we were unable to access the necessary data in the time scales of the project. We intend to trial using this method in the future.
7	In some LAs the household size distributions differs substantially from the national distribution. In these areas the SPREE model tends to provide poorer estimates.	This work will be picked up when the LFS is reweighted to the 2021 Census or when household weighted TLFS data is available.

	Trial running separate SPREE models by region when alternative data becomes available.	
8	One of the main assumptions we make in stage one is that people missing an address type in admin data have this data missing at random. Our analysis indicates this may not the case.	Work is ongoing in ONS to improve the record level admin data used in our estimates. Improvements to the assignment of people to address may reduce the size of the population not assigned to an address. This could be achieved by introducing new sources of UPRN data or improving the assignment methodology. This would reduce the impacts of this assumption. It may be possible to devise an imputation strategy to impute address type. However, this was outside of the scope of work detailed here.
9	Using a generalised structure preserving estimator (GSPREE) was not recommended for stage three. The estimates produced by GSPREE were similar to the SPREE results. However, the incorporation of unweighted APS data as a data source in the GSPREE model was hard to justify. This is due to the current quality limitations with the APS. ONS is currently working with the University of Southampton to develop an extension to the GSPREE method. This would allow for the use of weighted data as a data source.	Once this extended method is available to researchers, we plan to trial using it to produce the admin-based household size estimates (ABHSEs).

5. Future Work

The living arrangements estimation team intend to publish an update on our admin-based household estimates (ABHEs) for LAs in England and Wales in Winter 2025/26. We acknowledge that there is further work to be done to understand and improve the ABHEs.

Our aim in publishing this work is to show users our progress towards producing admin-based estimates and to gain feedback from users. Whilst there are clear user needs for more granular and timely statistics on households, we have limited understanding of users' needs around the accuracy and reliability of these statistics. For example, we are unclear whether uncertainty measures are required by users.

Therefore, additional engagement will need to be conducted with users to better understand their needs. Publishing these estimates will facilitate more meaningful conversations with users.

To support our research on admin-based household estimates (ABHE), ONS has been iteratively developing the admin-based living arrangements dataset (ABLAD). The ONS is now planning to conduct additional work which seeks to improve the record level data available on people's living arrangements. This work will be shaped by user needs (including data requirements for developing ABHEs), and the data available. It will also consider whether these developments will focus on improvement to the current record level data (ABLAD) or whether new design options could better meet needs of users. These developments are expected to improve the quality of data available for ABHEs. However, additional analysis will be needed to understand how changes to the underlying data affect the statistics produced by the admin-based household estimates.

The methods outlined in this paper require survey data. The Labour Force Survey is undergoing a period of transformation, to create the Transformed Labour Force Survey (TLFS) (ONS, 2025). There is currently a period of overlap, with both the LFS and TLFS running. We are currently in conversation with colleagues working on survey requirements and development. This will ensure that our needs continue to be met in the future. Once weighted household data is available from the TLFS, the living arrangements estimation team will need to conduct analysis to ensure that the TLFS is suitable for use in the methods proposed in this paper.

6. References

Morgan, A., Blake, A., and Mylles, S. (2024) *Producing admin-based household estimates – research to date and plans*. Available at EAP207-Producing-admin-based-household-estimates--research-to-date-and-plans.pdf (Accessed 21/10/2025)

Office for National Statistics (2008) *Labour Force Survey User Guide – Volume 8: Household and Family Data*. Available at: https://www.ons.gov.uk/file?uri=/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/labourforcesurveyuserguidance/lfsugvoltcm77238281.pdf (Accessed 17/06/2025)

Office for National Statistics (2017) *Research Outputs: Occupied address (household) estimates by size, 2011.* Available at: https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/censusanddatacollectiontransformationprogramme/futureofpopulationandsocialstatistics/administrativedatacensusresearchoutputs/householdsandfamilies/researchoutputsoccupiedaddresshouseholdestimatesbysize2011 (Accessed 17/06/2025)

Office for National Statistics (2018) *Research Outputs: An update on developing household statistics for an Administrative Data Census.* Available at: https://www.ons.gov.uk/aboutus/whatwedo/programmesandprojects/censusanddatacollectiontransformationprogramme/futureofpopulationandsocialstatistics/administrativedatacensusresearchoutputs/householdsandfamilies/researchoutputsanupdateonde

velopinghouseholdstatisticsforanadministrativedatacensus#household-statistics-for-2011 (Accessed 18/06/2025)

Office for National Statistics (No Date) *Reference Data Management Framework*. Available at Reference Data Management Framework - Office for National Statistics (Accessed 21/10/2025)

Annex 1: Definitions

Address	An address is a collection of information, usually in a fixed format, used to give the location of a place. A place can be any kind of building, or object that might not have a 'normal' address. This will include residential and business properties but may also include other structures such as a bus shelter or an electricity sub-station for example.
Occupied address	An occupied address is a private residential address identified by a UPRN that is found on the Address Frame which has been successfully linked to at least one individual in the Statistical Population Dataset. Most occupied address contain only one household. However, it is possible for multiple households to share an occupied address.
Household	One person living alone or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room or sitting room or dining area. A household can consist of a single family, more than one family or no families in the case of a group of unrelated people living together.
Communal Establishment (CE)	A place providing managed residential accommodation. 'Managed' here means full-time or part-time supervision of the accommodation, such as care homes, student halls of residence, hospitals, or prisons.
UPRN	A Unique Property Reference Number is a unique identifier for every addressable location in the UK. They provide every place with a consistent identifier throughout its lifecycle, from planning through to demolition.