

Article

Developing admin-based housing stock statistics for England and Wales: 2020

An overview of methods, data sources and data quality for the feasibility research on producing census-like housing stock statistics using administrative data.

Contact:
Andreea Butnaru, Sam Fisher,
Jamie Knowles, Shannon Mahon,
Emily Mason-Apps, Charlotte
Standeven, Stephan Tietz
admin.based.
characteristics@ons.gov.uk
+44 1329 444528

Release date:
5 December 2022

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [About our transformation research](#)
3. [Data used to produce the admin-based housing stock version 1.0 dataset for 2020](#)
4. [Comparing admin-based housing stock statistics with the 2011 Census](#)
5. [Comparing admin-based housing stock statistics with the Council Tax stock of properties](#)
6. [Future developments](#)
7. [Glossary](#)
8. [Data sources and quality](#)
9. [Related links](#)
10. [Cite this article](#)

1 . Main points

- To explore the feasibility of using administrative data to provide more regular census-like housing statistics for occupied residential addresses (down to small geographies) across England and Wales, we have developed an admin-based housing stock version 1.0 (ABHS V1.0) dataset for 2020.
- The ABHS was produced by linking a residential Address Frame to Valuation Office Agency (VOA) data to obtain property characteristic information, and to the admin-based household estimates (ABHE) datasets to identify occupied addresses.
- Because of undercoverage in the ABHE datasets, the ABHS identifies a lower proportion of occupied addresses (especially flats) than the 2011 Census.
- The number of people within addresses aligns more closely with the 2011 Census household size when using the ABHE version 3.0 (V3.0) to flag occupied addresses than when using the ABHE version 2.0 (V2.0).
- Comparisons with the 2011 Census showed that the ABHS reports a larger proportion of terraced addresses and smaller proportion of semi-detached addresses, in line with previous research that identified a preference of the 2011 Census respondents to define their property as semi-detached instead of end-terrace.
- Comparisons of the VOA's Council Tax stock of properties with all addresses on the ABHS showed that the number of bedrooms across all property types is very similar across all regions in England and Wales (differences of less than 0.8 percentage points); this suggests that the ABHS data are of good quality despite challenges identifying occupied addresses.
- Our associated statistical summary, [Admin-based housing stock profile for England and Wales: 2020](#), provides more detailed figures at a national and local authority level for all addresses in the ABHS and occupied addresses in the ABHS (using the ABHE V3.0 to flag occupied addresses).
- The Office for National Statistics (ONS) is exploring the feasibility of linking the ABHS dataset with an admin-based ethnicity dataset to produce housing by ethnicity statistics.

These are not official statistics and should not be used for policymaking or decision-making. They are published as feasibility research into a new method for producing census-like statistics on housing using administrative data. We advise caution when using the data.

2 . About our transformation research

At the Office for National Statistics (ONS), we are [exploring the feasibility of using administrative data to produce statistics on a range of housing topics](#), which might remove the need for us to collect data through a census or surveys and provide new analysis on housing topics not previously available. This may help housing planners and policymakers to better understand the characteristics of the dwelling stock in their areas and therefore better meet the future housing needs of local residents (see the [Census 2021 topic consultation](#)).

This article describes the method used to produce the admin-based housing stock version 1.0 (ABHS V1.0) dataset for 2020 that brings together data from several administrative sources to provide a new method for producing census-like statistics on the housing stock in England and Wales. The Valuation Office Agency (VOA) publish annual statistics on the [stock of properties by Council Tax band, providing information about property attributes](#) such as number of bedrooms, property type, and build age of properties in England and Wales. However, there are clear definitional differences between the Council Tax stock of properties published by the VOA and census housing statistics. The VOA produce statistics on residential (or domestic) properties, which are liable to pay Council Tax, while census or surveys produce statistics on households with at least one usual resident (occupied residential addresses).

It is not currently possible to distinguish between occupied and vacant residential properties using VOA data alone. Because of survey sample sizes, detailed and timely information for some household characteristics in the intercensal years can be sparse or limited when it comes to analysis for sub-regional geographies. We are therefore exploring the feasibility of creating the ABHS V1.0 dataset to produce more regular census-like housing statistics for occupied residential addresses down to small geographies across England and Wales.

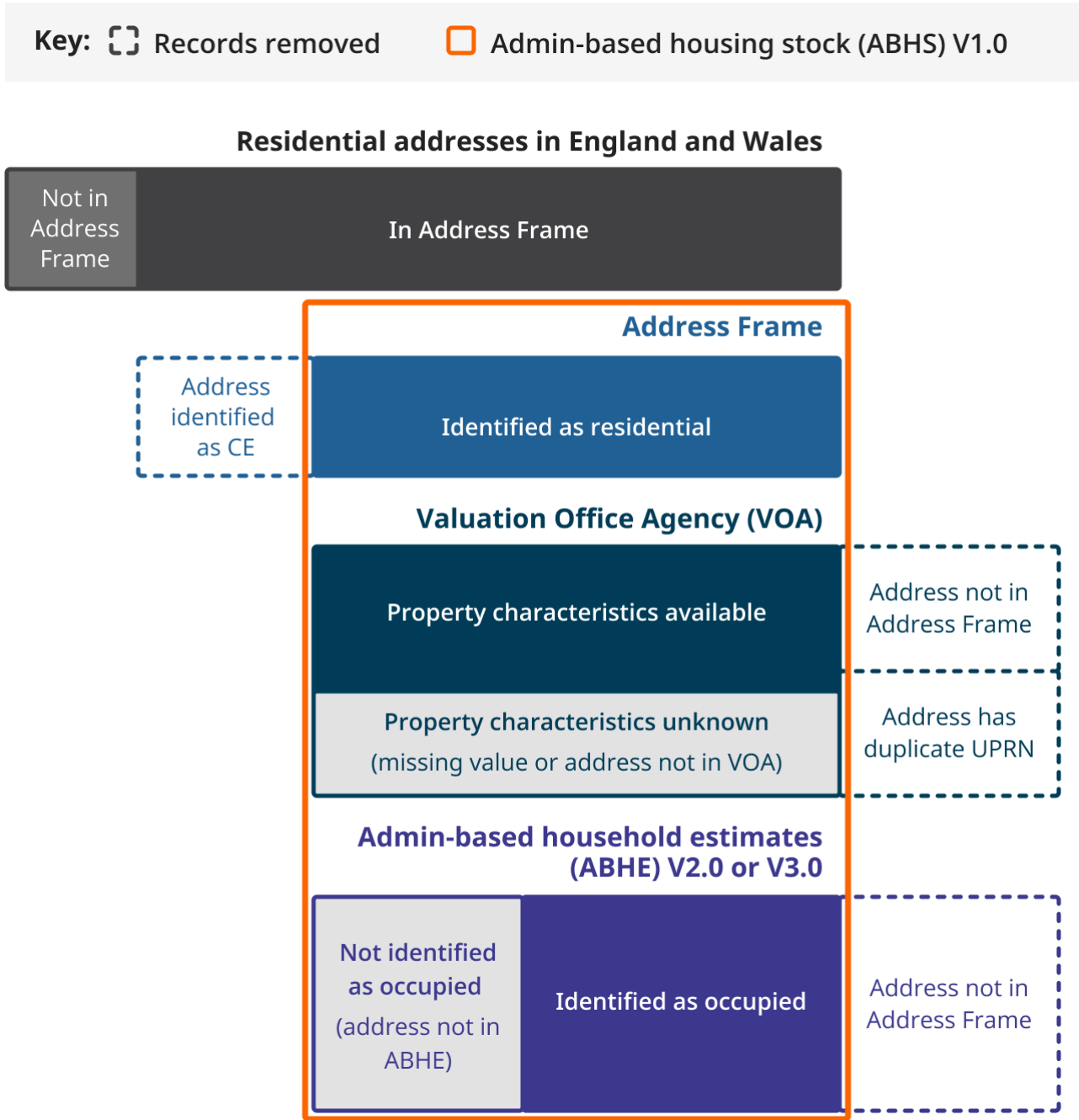
[Previous research](#) has demonstrated the viability of using linked VOA data to produce census-like statistics such as accommodation type and overcrowding, as well as providing detailed information about floor area. A recent application of this research is the removal of the number of rooms question from Census 2021, with this information coming from VOA data instead.

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 . Data used to produce the admin-based housing stock version 1.0 dataset for 2020

Figure 1 provides a visual representation of the steps taken to produce the admin-based housing stock version 1.0 (ABHS V1.0) dataset for 2020 by linking a residential Address Frame to Valuation Office Agency (VOA) data and the admin-based household estimates version 2.0 and version 3.0 (ABHE V2.0 and V3.0). Here we describe how we identified residential addresses, what property characteristics are available, and how we identified occupied addresses. More detailed information about the cleaning and linkage can be found in [Section 8](#).

Figure 1: Visual representation of the datasets used to identify occupied addresses and obtain property characteristics on the admin-based housing stock dataset



Source: Office for National Statistics

Notes:

1. Please note this diagram is not to scale.
2. 'CE' refers to communal establishment (see [Section 7](#)).

Identifying residential addresses in England and Wales

We used an Address Frame from June 2020 to identify residential addresses in England and Wales. The Address Frame was produced using data from 2020 with the same [methodology used to produce the Address Frame for Census 2021](#). We removed communal establishments (CEs) from the Address Frame to compare with the 2011 Census. This removed 0.1% of the Address Frame, resulting in approximately 26.0 million addresses on the residential Address Frame.

Available admin-based property characteristics

VOA property characteristics data from June 2020 were used to identify important housing characteristics for residential addresses. VOA data cover residential properties within England and Wales and include information on property attributes such as property type, number of rooms, and number of bedrooms. Of the 26.0 million addresses in the residential Address Frame, 99.2% of records linked to a VOA address.

Identifying occupied residential addresses

The ABHE V2.0 and V3.0 datasets for 2020 have been used to identify occupied addresses. The ABHEs are derived from the Statistical Population Datasets (SPDs), which provide estimates of the usually resident population of England and Wales. The ABHEs group usual residents from the SPDs into addresses to estimate the size and composition of occupied addresses for England and Wales. The ABHE V2.0 and V3.0 are derived from the SPD V2.0 and V3.0, respectively. The SPD V3.0 was designed with the explicit objective of reducing population overcount apparent in SPD V2.0, but subsequently it results in higher levels of undercoverage.

Addresses from the residential Address Frame were flagged as occupied if they successfully linked to at least one record in the respective ABHE dataset. We use the terms "ABHE V2.0 occupied flag" and "ABHE V3.0 occupied flag" to refer to those addresses that linked to at least one record on the ABHE V2.0 and the ABHE V3.0, respectively. The ABHE V2.0 occupied flag identifies 87.9% of records from the residential Address Frame as occupied, whereas the ABHE V3.0 occupied flag identifies 86.3% of addresses as occupied. Because the ABHE V2.0 and V3.0 are derived from different versions of the SPDs, linking to both datasets enables us to assess which underlying method may work better to help us identify occupied addresses. More information about the ABHE V2.0 and V3.0 datasets, and how they are derived from the SPDs, can be found in [Section 8](#).

Deriving census-like housing variables

The VOA accommodation type variable was derived from VOA data to resemble the seven 2011 Census accommodation types as closely as possible, while adding an additional eighth category for annexes. More information can be found in [Section 7](#).

We used the ABHE V2.0 and V3.0 datasets to derive a count of the number of people within each address to resemble the 2011 Census household size. It is worth noting that it is currently not possible to clearly identify multiple households at an address from administrative data alone.

To assess the quality of the ABHE datasets, we compared the number of people within an address according to the ABHS using the ABHE V2.0 and V3.0 with the 2011 Census household size (see Figure 2). We found that the ABHE V2.0 identifies fewer one-person and two-person addresses than the 2011 Census but more addresses with three or more people. For the ABHE V3.0, the distribution of number of people within addresses is more closely aligned with the 2011 Census household size, with the most notable differences being the ABHE V3.0 reporting a higher proportion of three-person addresses and a lower proportion of two-person addresses.

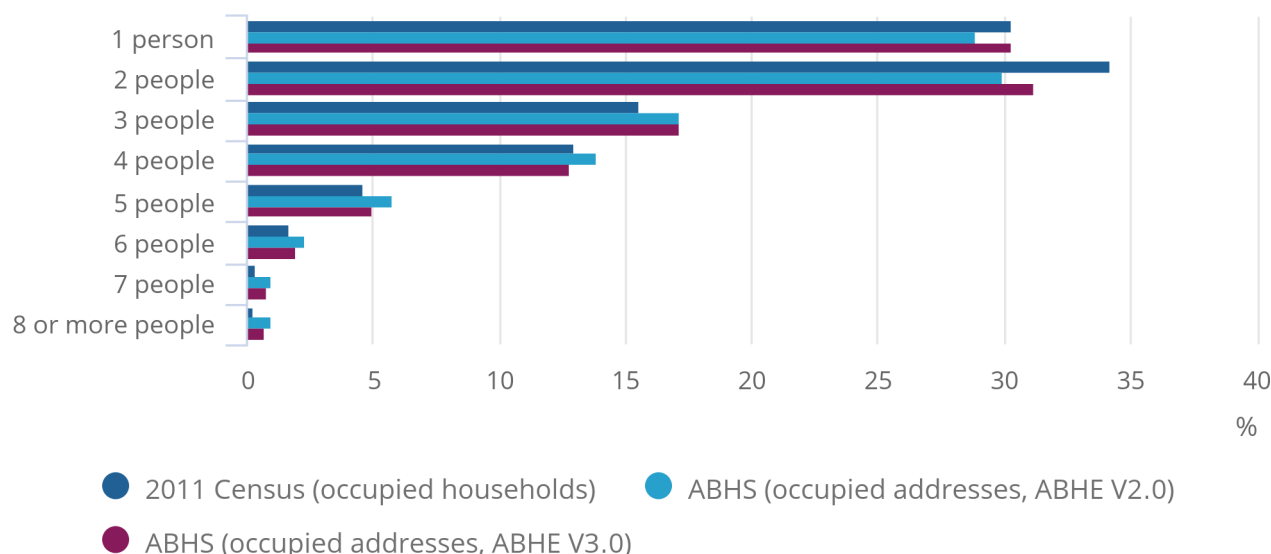
Overall, we tend to have more people per address according to the ABHS occupied flags than were reported by the 2011 Census. This can be explained by lags in the administrative data that are used to create the SPDs, for example the lag between moving to a new address and registering with a new GP or deregistering when people leave the country. Additionally, the census counts households while the ABHS counts addresses, some of which will contain multiple households.

Figure 2: Using the ABHE V2.0 occupied address flag identifies a smaller proportion of one- and two-person addresses compared with using the ABHE V3.0

Distribution of 2011 Census households by household size, alongside the distribution of occupied addresses in the ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) by number of people per address, for England and Wales

Figure 2: Using the ABHE V2.0 occupied address flag identifies a smaller proportion of one- and two-person addresses compared with using the ABHE V3.0

Distribution of 2011 Census households by household size, alongside the distribution of occupied addresses in the ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) by number of people per address, for England and Wales



Source: Office for National Statistics and Valuation Office Agency

Notes:

1. ABHS V1.0 refers to the admin-based housing stock version 1.0 dataset for 2020.
2. ABHE V2.0 and V3.0 refers to admin-based household estimates version 2.0 and version 3.0.

4 . Comparing admin-based housing stock statistics with the 2011 Census

The admin-based housing stock version 1.0 (ABHS V1.0) dataset for 2020 identifies 1.6 million more addresses in England and Wales compared with the 2011 Census. As there is a nine-year gap in data collection between the 2011 Census and the ABHS V1.0 for 2020, it is important to note that many new properties were built within that period, which would not have been present at the time of the 2011 Census. Figures from [the Department for Levelling Up, Housing and Communities](#) and [Welsh Government](#) estimate a net total of 1.7 million additional residential addresses across England and Wales between March 2011 and March 2020.

Occupied household spaces and addresses

While we cannot directly compare the absolute number of addresses between the 2011 Census and ABHS, the 2011 Census provides a comprehensive baseline to assess the quality of flagging occupied addresses in the ABHS.

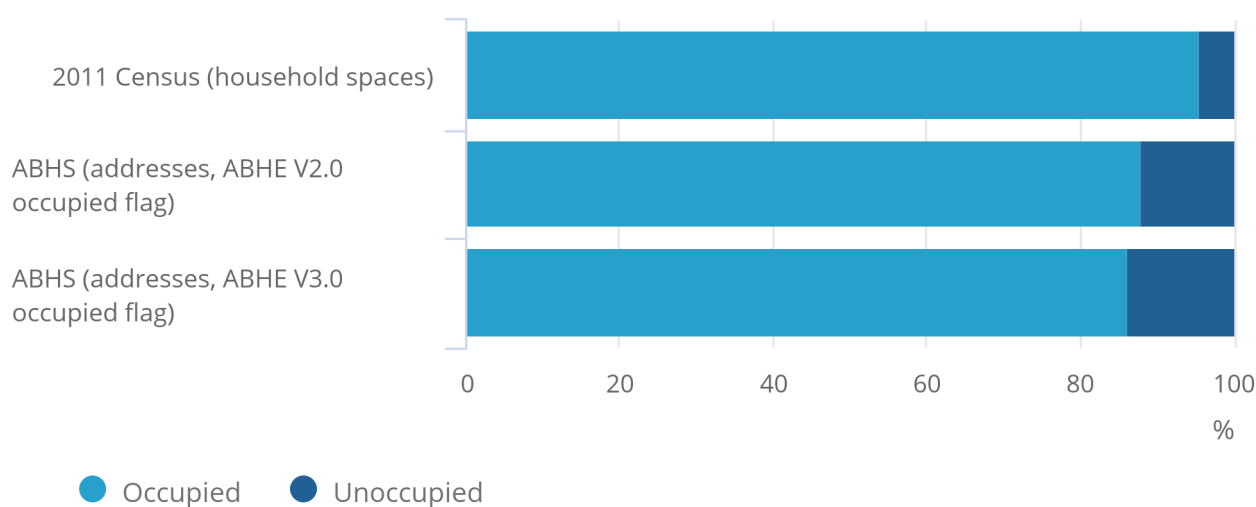
Figure 3 shows that the proportion of occupied household spaces in the 2011 Census is higher than the proportion of occupied addresses in the ABHS. Using the admin-based household estimates version 2.0 (ABHE V2.0) to flag occupied addresses, we observe an undercoverage of 7.8 percentage points, and using the admin-based household estimates version 3.0 (ABHE V3.0), the undercoverage is 9.4 percentage points. Most regions followed a similar pattern, with London showing the largest discrepancy between the proportion of occupied household spaces in the 2011 Census and occupied addresses in the ABHS (13.7 and 16.0 percentage points using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses, respectively).

Figure 3: The 2011 Census identifies a higher proportion of occupied household spaces compared with occupied addresses in the ABHS V1.0

Proportion of occupied and unoccupied households for the 2011 Census and addresses in ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) for England and Wales

Figure 3: The 2011 Census identifies a higher proportion of occupied household spaces compared with occupied addresses in the ABHS V1.0

Proportion of occupied and unoccupied households for the 2011 Census and addresses in ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) for England and Wales



Source: Office for National Statistics and Valuation Office Agency

Notes:

1. ABHS V1.0 refers to the admin-based housing stock version 1.0 dataset for 2020.
2. ABHE V2.0 and V3.0 refers to admin-based household estimates version 2.0 and version 3.0.

Occupied addresses by accommodation type

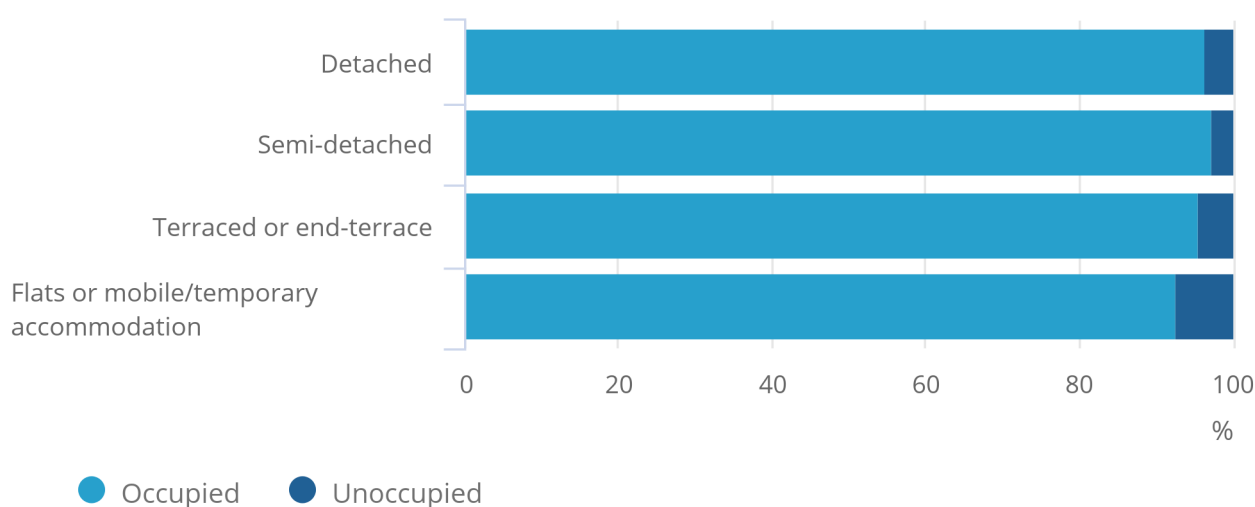
Figure 4, Figure 5, and Figure 6 show that the ABHS identifies a much higher proportion of unoccupied flats (17.5 and 19.6 percentage points using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses, respectively) than the 2011 Census. This means that the undercoverage of the ABHE has a more notable effect on the proportion of occupied flats compared with other accommodation types.

Figure 4: The 2011 Census identifies a higher proportion of unoccupied flats compared with other accommodation types

Proportion of occupied and unoccupied household spaces by accommodation type for the 2011 Census for England and Wales

Figure 4: The 2011 Census identifies a higher proportion of unoccupied flats compared with other accommodation types

Proportion of occupied and unoccupied household spaces by accommodation type for the 2011 Census for England and Wales



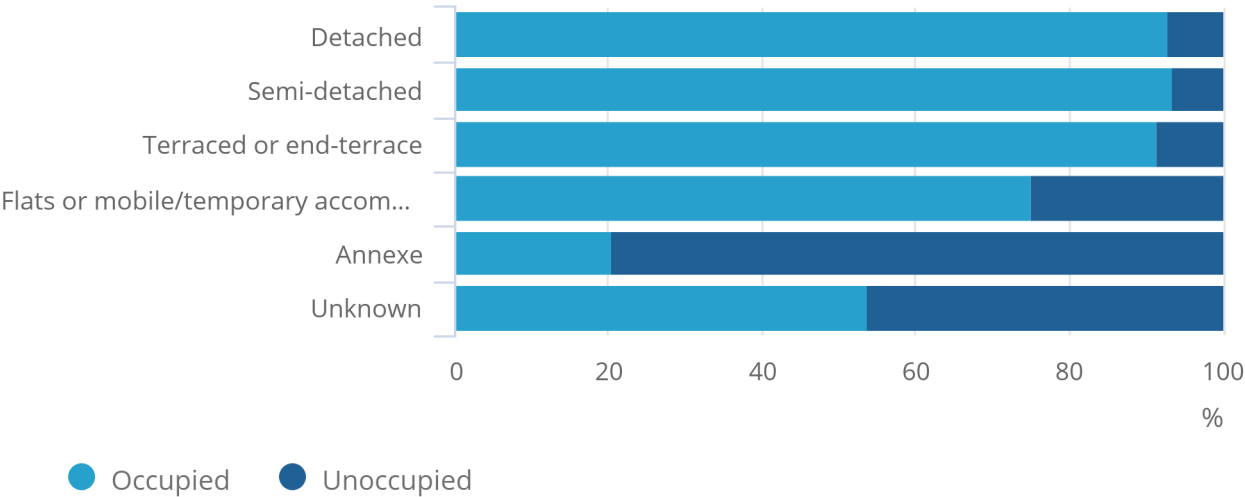
Source: Office for National Statistics

Figure 5: The ABHS V1.0 (using the ABHE V2.0 to flag occupied addresses) identifies a much higher proportion of unoccupied addresses that are flats, annexes or unknown compared with other accommodation types

Proportion of occupied and unoccupied addresses by accommodation type for the ABHS V1.0 for 2020 (using the ABHE V2.0 to flag occupied addresses) for England and Wales

Figure 5: The ABHS V1.0 (using the ABHE V2.0 to flag occupied addresses) identifies a much higher proportion of unoccupied addresses that are flats, annexes or unknown compared with other accommodation types

Proportion of occupied and unoccupied addresses by accommodation type for the ABHS V1.0 for 2020 (using the ABHE V2.0 to flag occupied addresses) for England and Wales



Source: Office for National Statistics and Valuation Office Agency

Notes:

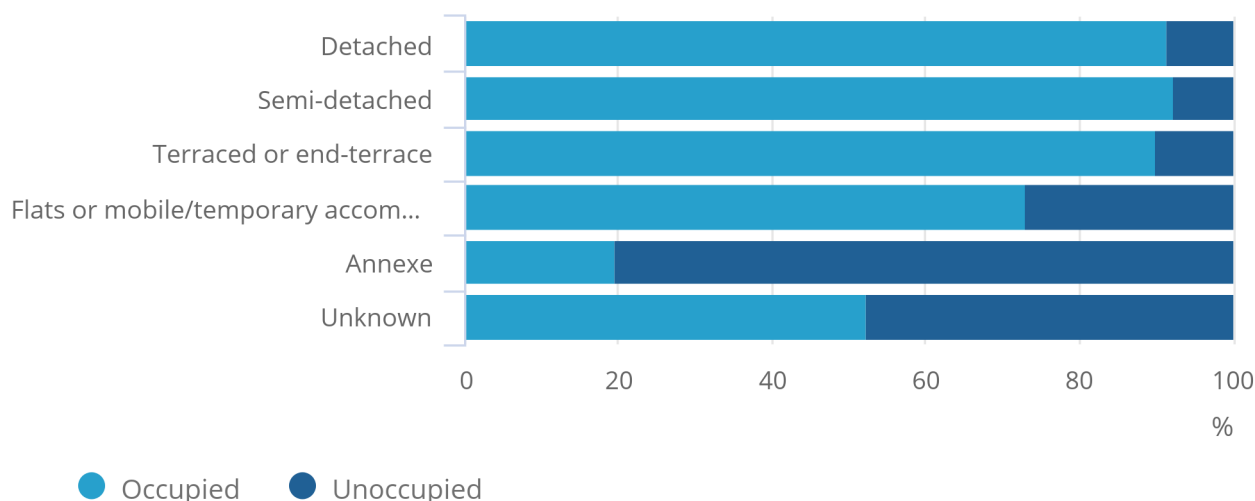
- 1. ABHS V1.0 refers to the admin-based housing stock version 1.0 dataset for 2020.
- 2. ABHE V2.0 and V3.0 refers to admin-based household estimates version 2.0 and version 3.0.

Figure 6: The ABHS V1.0 (using the ABHE V3.0 to flag occupied addresses) identifies a much higher proportion of unoccupied addresses that are flats, annexes or unknown compared with other accommodation types

Proportion of occupied and unoccupied addresses by accommodation type for the ABHS V1.0 for 2020 (using the ABHE V3.0 to flag occupied addresses) for England and Wales

Figure 6: The ABHS V1.0 (using the ABHE V3.0 to flag occupied addresses) identifies a much higher proportion of unoccupied addresses that are flats, annexes or unknown compared with other accommodation types

Proportion of occupied and unoccupied addresses by accommodation type for the ABHS V1.0 for 2020 (using the ABHE V3.0 to flag occupied addresses) for England and Wales



Source: Office for National Statistics and Valuation Office Agency

Notes:

1. ABHS V1.0 refers to the admin-based housing stock version 1.0 dataset for 2020.
2. ABHE V2.0 and V3.0 refers to admin-based household estimates version 2.0 and version 3.0.

We also looked at the distribution of occupied addresses by accommodation type. As we would expect from the ABHE undercoverage, Figure 7 shows a smaller proportion of occupied flats in the ABHS compared with the 2011 Census (1.7 and 1.9 percentage points using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses, respectively). However, the ABHS also has a larger proportion of occupied addresses with unknown characteristics (1.1% of occupied addresses in the ABHS). Unique Property Reference Numbers (UPRNs) are assigned to addresses, which are subsequently used for linking. The assignment of UPRNs is constantly improving. However, complex addresses can pose challenges as UPRN assignment can be perfect at a building level (such as a block of flats) but ambiguous at the unit level (such as an individual flat). This means that compared with other accommodation types, it is more difficult to link flats within a block of flats successfully and accurately. We will attempt to investigate this in future research.

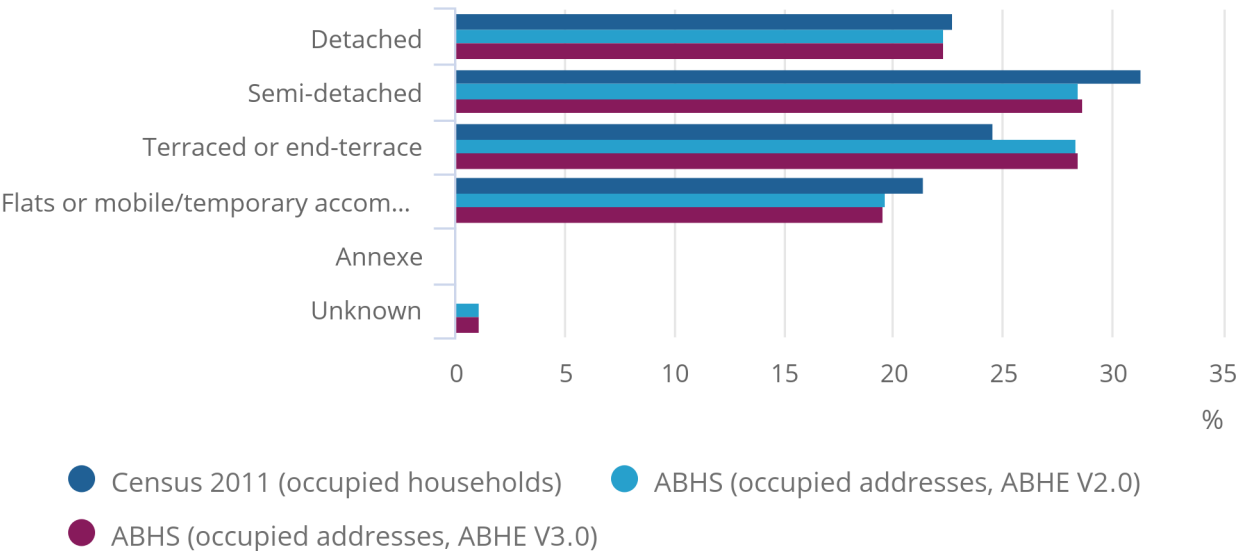
Figure 7 also shows that the ABHS reports a larger proportion of occupied terraced addresses (3.7 and 3.8 percentage points using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses, respectively) and smaller proportion of occupied semi-detached addresses (2.7 and 2.6 percentage points using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses, respectively) than the 2011 Census. [Our previous research comparing VOA accommodation type with the 2011 Census](#) observed a similar difference (negative 3.0 percentage point difference for semi-detached and 3.5 percentage point difference for terraced) because of a preference of the 2011 Census respondents to define their property as semi-detached instead of end-terrace.

Figure 7: The ABHS V1.0 identifies a larger proportion of occupied terraced addresses and a smaller proportion of occupied semi-detached addresses than the 2011 Census

Distribution of occupied households in the 2011 Census and occupied addresses in the ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) by VOA accommodation type, for England and Wales

Figure 7: The ABHS V1.0 identifies a larger proportion of occupied terraced addresses and a smaller proportion of occupied semi-detached addresses than the 2011 Census

Distribution of occupied households in the 2011 Census and occupied addresses in the ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) by VOA accommodation type, for England and Wales



Source: Office for National Statistics and Valuation Office Agency

Notes:

1. ABHS V1.0 refers to the admin-based housing stock version 1.0 dataset for 2020.
2. ABHE V2.0 and V3.0 refers to admin-based household estimates version 2.0 and version 3.0.

5 . Comparing admin-based housing stock statistics with the Council Tax stock of properties

The Valuation Office Agency (VOA) produce an annual release that provides statistics on the stock of domestic properties by council tax band and property attributes in England and Wales. To understand coverage and quality of the admin-based housing stock version 1.0 (ABHS V1.0) dataset for 2020, we compared proportions of different property types and number of bedrooms at national and regional level with the [VOA statistics produced for 2020](#).

Across England and Wales, the VOA report approximately 90,000 (or 0.3%) more addresses than the ABHS. This difference is because of the slightly different reference period (March 2020 for Council Tax stock of properties and June 2020 for ABHS) and some VOA addresses not being present on the residential Address Frame (for example, 20,000 addresses are flagged as a communal establishment).

When we compare the number of properties liable to pay Council Tax to the number of occupied addresses on the ABHS, we get 3.3 million fewer addresses using the admin-based household estimates version 2.0 (ABHE V2.0) to flag occupied addresses and 3.7 million fewer addresses using the admin-based household estimates version 3.0 (ABHE V3.0). We expect the number of properties that are liable to pay Council Tax to exceed the number of occupied addresses on the ABHS as we consider an address occupied when there is at least one usual resident, but properties may be liable to pay Council Tax even when they are unoccupied (see [Section 7](#)). However, we concluded in [Section 4](#) that the ABHS underestimates the number of occupied addresses and further work is required to identify all occupied addresses.

Property type

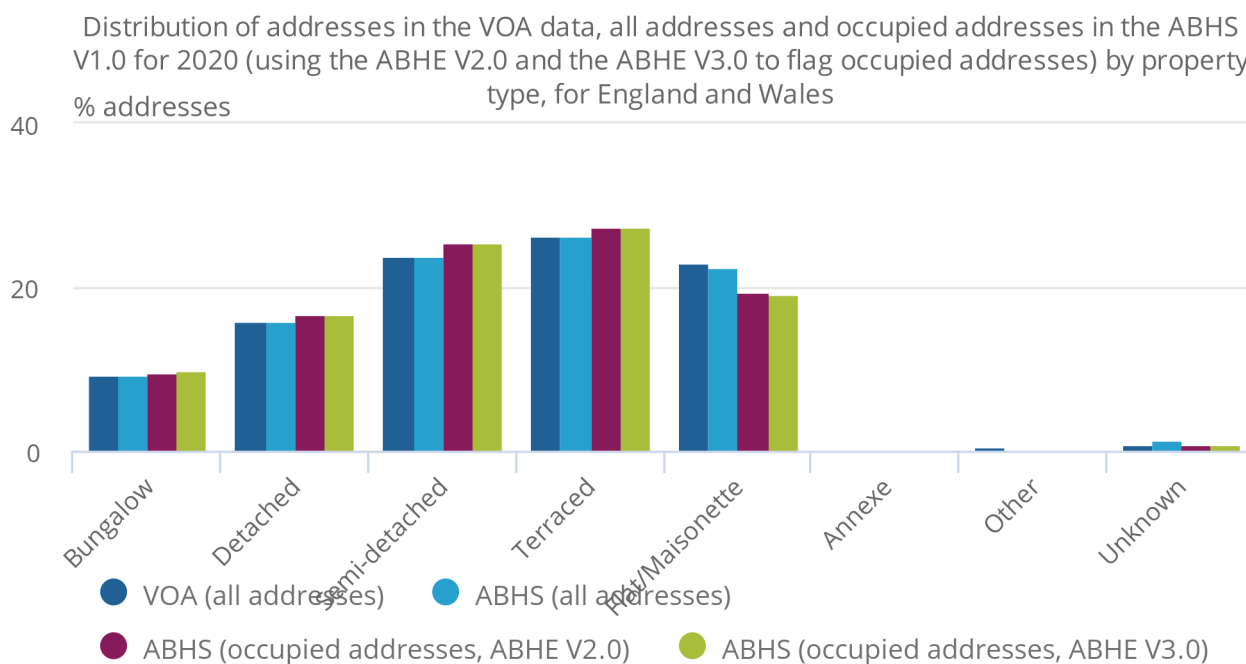
Figure 8 shows that the distribution of property types for all addresses within the VOA and the ABHS follow a very similar pattern. The ABHS has a larger proportion (0.8%) of unknown property types, which is mostly because of addresses on the Address Frame not having linked to a VOA address (see [Section 3](#)). For the remaining property types, the largest difference of 0.4 percentage points is observed for flats. This difference could be a result of the cleaning method we used on the VOA dataset before linking it, which is more likely to affect flats (see [Section 8](#)).

VOA data do not contain any information to identify addresses that are occupied by usual residents, and we therefore assessed the quality of the occupied flags in [Section 4](#) by comparing with the 2011 Census. Nonetheless, Figure 8 shows how the distribution changes for different property types. Using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses on the ABHS gives a lower proportion of flats and higher proportions of detached, semi-detached, and terraced properties compared with the VOA.

Figure 8: The distribution of all addresses in the VOA and ABHS V1.0 data by property type is similar, with terraced and semi-detached being the most common property types in England and Wales

Distribution of addresses in the VOA data, all addresses and occupied addresses in the ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) by property type, for England and Wales

Figure 8: The distribution of all addresses in the VOA and ABHS V1.0 data by property type is similar, with terraced and semi-detached being the most common property types in England and Wales



Source: Office for National Statistics and Valuation Office Agency

Notes:

1. ABHS V1.0 refers to the admin-based housing stock version 1.0 dataset for 2020.
2. ABHE V2.0 and V3.0 refers to admin-based household estimates version 2.0 and version 3.0.
3. VOA refers to Valuation Office Agency data.

Number of bedrooms

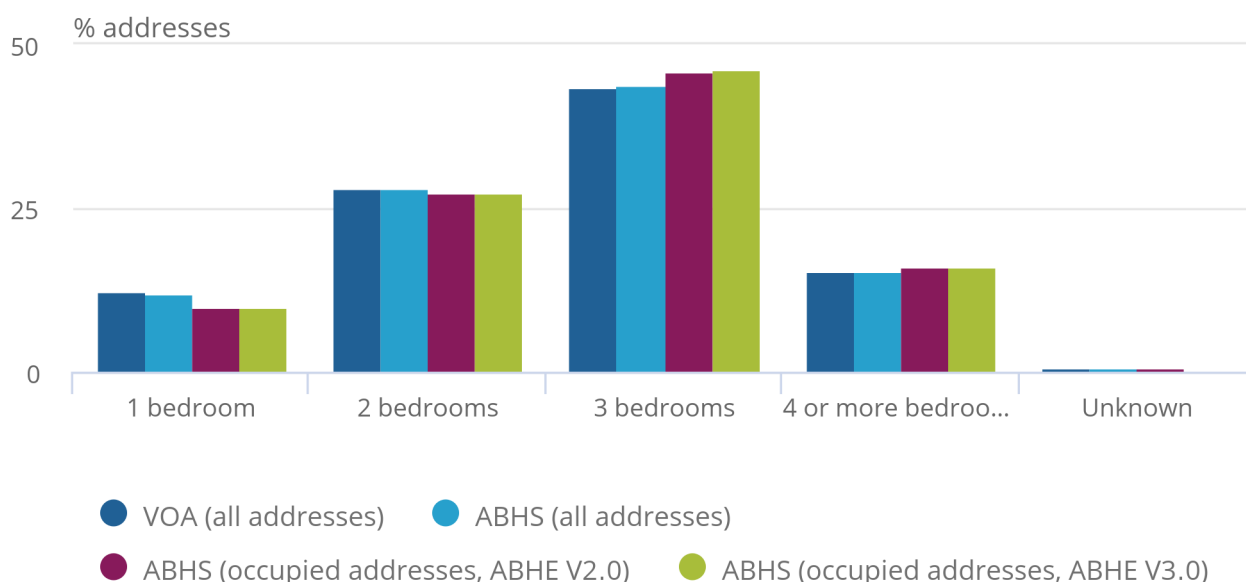
Figure 9 shows that, across England and Wales, all addresses within the VOA and the ABHS have a very similar distribution of number of bedrooms with all differences smaller than 0.2 percentage points. Occupied addresses in the ABHS had lower proportions of one- and two-bedroom properties (approximately 2.2 and 0.8 percentage points, respectively) and higher proportions of three-bedroom and four-or-more-bedroom properties (approximately 2.5 and 0.6 percentage points, respectively) compared with VOA data.

Figure 9: The distribution of all addresses in the VOA and ABHS V1.0 data by number of bedrooms is similar, with addresses most commonly having three bedrooms in England and Wales

Distribution of addresses in the VOA data, all addresses and occupied addresses in the ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) by number of bedrooms, for England and Wales

Figure 9: The distribution of all addresses in the VOA and ABHS V1.0 data by number of bedrooms is similar, with addresses most commonly having three bedrooms in England and Wales

Distribution of addresses in the VOA data, all addresses and occupied addresses in the ABHS V1.0 for 2020 (using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses) by number of bedrooms, for England and Wales



Source: Office for National Statistics and Valuation Office Agency

Notes:

1. ABHS V1.0 refers to the admin-based housing stock version 1.0 dataset for 2020.
2. ABHE V2.0 and V3.0 refers to admin-based household estimates version 2.0 and version 3.0.
3. VOA refers to Valuation Office Agency data.

In London, differences were more pronounced compared with other regions in England and Wales. As for England and Wales, the distribution of number of bedrooms was very similar for all addresses on the VOA and ABHS, with all differences smaller than 0.3 percentage points. However, using the ABHE V2.0 and the ABHE V3.0 to flag occupied addresses gave a lower proportion of one- and two-bedroom properties (approximately 4.4 and 0.2 percentage points, respectively) and a higher proportion of three-bedroom and four-or-more-bedroom properties (approximately 3.9 and 0.9 percentage points, respectively) compared with the VOA. These differences are driven by the ABHEs, which currently do not cover all population groups equally well.

We also looked at the number of bedrooms by property type. For all addresses within the VOA and the ABHS, the distribution of number of bedrooms across all property types is very similar, with all differences less than 0.8 percentage points across all regions in England and Wales. This suggests that the ABHS data are of good quality despite the discussed challenges for flagging occupied addresses (see [Section 4](#)).

6 . Future developments

We developed the admin-based housing stock version 1.0 (ABHS V1.0) dataset for 2020 and assessed its feasibility to provide more regular census-like housing statistics for occupied residential addresses (down to small geographies) across England and Wales. Figures at national and local authority level using the ABHS V1.0 for 2020 for all and occupied addresses using the admin-based household estimates version 3.0 (ABHE V3.0) can be found in [our statistical summary of the admin-based housing stock for 2020](#).

Our aim is to produce a dataset with a comprehensive set of property attributes and other housing information that can become the primary source for further analysis and statistics on housing characteristics for occupied addresses. Further work is required to assess and improve the quality of these statistics on housing characteristics for occupied addresses. Future work will include:

- using future versions of the ABHE datasets to assess whether this improves the ability to accurately identify occupied addresses on the ABHS
- testing other administrative data sources, such as utilities data to flag occupied addresses
- producing the ABHS for 2021 and conducting comparisons with Census 2021 at address level as well as to summary statistics once these are released to users (see [Census 2021 Housing analysis plans](#))
- exploring other variables available on the Valuation Office Agency (VOA) data where they meet user needs
- expanding the ABHS dataset by linking to surveys and other administrative data sources such as Energy Performance Certificate (EPC) (to include additional property attributes such as central heating and fuel type), Tenancy Deposit Protection Scheme (TDPS) and Zero Deposit data (to derive a tenure flag), as well as map data (to derive access to amenities)
- exploring the impact of different methods for imputing missing data
- exploring the feasibility of linking the ABHS dataset with an admin-based ethnicity dataset to produce housing by ethnicity statistics, which we plan to publish at the beginning of 2023

Feedback

We welcome feedback on the method used to produce admin-based housing stock statistics. We are very interested in understanding what uses housing stock statistics have and how likely they are to be of interest in the future to inform policies, target schemes, and monitor changes over time. This information will help us to ensure we meet user needs where possible. Please email your feedback to admin.based.characteristics@ons.gov.uk. Please include "admin-based housing stock" in the subject line of your response.

7 . Glossary

2011 Census accommodation type

The 2011 Census has seven different accommodation types. They are:

- a whole house or bungalow that is detached
- a whole house or bungalow that is semi-detached
- a whole house or bungalow that is terraced (including end-terrace)
- a flat, maisonette or apartment that is in a purpose-built block of flats or tenement
- a flat, maisonette or apartment that is part of a converted or shared house (including bedsits)
- a flat, maisonette or apartment that is in a commercial building (for example, in an office building, hotel, or over a shop)
- a caravan or other mobile or temporary structure

For the published 2011 Census accommodation type data used for [Section 4](#), the three flat, maisonette or apartment categories were grouped with the caravan or mobile or temporary structure category into a single "Flat, maisonette, apartment, or mobile or temporary" category.

VOA accommodation type (2011 Census definition)

The Valuation Office Agency (VOA) accommodation type variable is derived from VOA property type and VOA dwelling code, and it is used for comparison in [Section 4](#). For this research, the VOA accommodation types are derived in such a way to resemble the seven 2011 Census accommodation types as closely as possible, while adding an additional eighth category for annexes. "Annexe" is not a category in the 2011 Census accommodation type variable, but it is a new category we propose for the VOA property type of "annexe". The VOA describe an annexe as a building, or part of a building, which has been constructed or adapted for use as separate living accommodation.

We use shortened wordings for the accommodation type categories in this article. The VOA property types that could not be fully mapped to a 2011 Census accommodation type were marked as "unknown". This included "houses in a cluster", "bungalows in a cluster", "houses of unidentified type", "bungalows of unidentified type" or "flats of unidentified type". Full information on the category names and mapping method can be found in our [Admin-based accommodation type statistics for England and Wales, feasibility research: 2011 methodology](#).

VOA property type

The VOA have 29 different property types plus a further three categories for "unidentified" houses, flats, and bungalows. VOA property type is used together with VOA dwelling code to derive VOA accommodation type (2011 Census definition).

In contrast, the VOA's [Council Tax: stock of properties for 2020 publication](#) groups VOA property types into eight categories. We create an equivalent variable in the admin-based housing stock (ABHS) to allow us to directly compare these data (see [Section 5](#)).

Unknown property characteristics

For comparisons with the 2011 Census (see [Section 4](#)), property characteristics are marked as "unknown" when the residential Address Frame did not link to a VOA address, when the property attributes data was missing on the linked VOA data, or when the VOA data could not be mapped to a 2011 Census accommodation type group.

For all comparisons with published VOA data (see [Section 5](#)), property characteristics are marked as "unknown" when the residential Address Frame did not link to a VOA address or when the property attributes information was missing on the linked VOA data.

Occupied or unoccupied household spaces and addresses

The 2011 Census enumerates self-contained dwellings (or addresses). Dwellings are either unshared if they contain a single household space or shared if they are made up of two or more household spaces. The analysis presented as part of this research does not distinguish between shared and unshared dwellings as these cannot be identified from administrative data alone.

A household space is "the accommodation used or available for use by an individual household". A household is defined as one person or a group of people (not necessarily related) living in a household space and sharing cooking facilities, and a living room, sitting room or dining area. The census and other surveys such as the Labour Force Survey do not consider communal establishments to be households.

A household space or dwelling is considered occupied when there is at least one usual resident. For the 2011 Census purposes, a usual resident of the UK is anyone who, on Census Day, was in the UK and had stayed or intended to stay in the UK for a period of 12 months or more, or had a permanent UK address and was outside the UK and intended to be outside the UK for less than 12 months.

A household space or dwelling is considered unoccupied when there are no usual residents. A household space with no usual residents may be completely vacant. But it could also be used by short-term residents, by visitors present on census night, as a second home, or a combination of these.

Comparatively, addresses on the ABHS are also self-contained dwellings, but it is currently not possible to clearly identify multiple households at an address from administrative data alone. Addresses are flagged as occupied if they successfully linked to at least one record in the admin-based household estimates (ABHE) dataset and unoccupied if they do not link to any records on the ABHE dataset. We use the terms "ABHE V2.0 occupied flag" and "ABHE V3.0 occupied flag" to refer to those addresses that linked to at least one record on the ABHE version 2.0 (V2.0) and ABHE version 3.0 (V3.0) respectively.

Communal Establishment

Communal establishments are managed residential accommodation such as hotels, care homes, student halls of residence, and prisons.

Unique property reference number (UPRN)

A unique property reference number (UPRN) is a unique identifier for every address in Great Britain and is allocated by local government and Ordnance Survey (OS).

8 . Data sources and quality

Address Frame

The Address Frame has been developed to accurately identify all residential addresses in England and Wales. It is based on [AddressBase Premium \(ABP\)](#), which uses Local Land and Property Gazetteers (LLPGs) in conjunction with a range of address intelligence sources such as the [Valuation Office Agency \(VOA\)](#), [Royal Mail](#) and [Ordnance Survey](#). [Geoplace](#) have developed processes for understanding and assuring the quality of the address information. The Office for National Statistics (ONS) supplement ABP with additional administrative and commercial sources to create a frame of residential addresses, ensuring that it identifies missed or misclassified addresses in the ABP, validates addresses with a lower confidence in their accuracy, and removes invalid addresses from the final frame. The ONS applies quality checks and targets to minimise undercoverage and overcoverage of addresses in any given local authority. More detailed information on the design of Address Frame and its quality has been published for Census 2021 in [Section 2 of our Design for Census 2021 article](#) and [our Administrative data used in Census 2021, England and Wales methodology](#).

2011 Census data

The comparisons with the 2011 Census in [Section 4](#) are based on official output tables. No address-level linkage or comparisons with record-level 2011 Census data have been conducted for this research.

Valuation Office Agency (VOA) property characteristics data

The VOA capture data about properties for Council Tax banding purposes, meaning that VOA data should cover all properties in England and Wales that are liable to pay Council Tax.

The information that the VOA collect and hold about domestic and residential properties supports its statutory function of banding properties for Council Tax. For this research, we have used VOA property attributes data on property type, number of rooms, number of bedrooms, number of bathrooms, and build age (see [VOA's Property Details Guide](#) for a full list).

Unique property reference numbers (UPRNs) were mapped to the VOA's unique address reference number (UARN) for each address using the cross-reference table on AddressBase Premium. We removed 0.4% of addresses that had a duplicate UPRN, or where a UPRN could not be assigned.

It is worth noting that VOA data are not regularly updated until a property is sold, meaning that any modifications made to a property to change the number of rooms or bedrooms may not always be captured in the VOA data. Further information about VOA data and their quality can be found in our [Valuation Office Agency property attribute data: quality assurance of administrative data used in Census 2021 methodology](#).

Admin-based household estimates (ABHE)

The ABHEs are derived from the Statistical Population Datasets (SPDs), which provide estimates of the usually resident population of England and Wales. The ABHEs are created by taking all usual residents from the SPD that can be assigned a unique property reference number (UPRN) and grouping them into addresses to estimate the size and composition of occupied addresses. The ABHE version 2.0 and version 3.0 (ABHE V2.0 and V3.0) are derived from the SPD version 2.0 and version 3.0 (SPD V2.0 and 3.0), respectively.

The SPD V2.0 identifies usual residents based on their presence on two or more of the administrative data sources used to build it. More information about the SPD V2.0 dataset can be found in the [SPD V2.0 methodology report](#). To create the ABHE V2.0, the SPD V2.0 is linked to the Personal Demographic Service (PDS) data via NHS number and successfully assigns a UPRN to 99.8% of usual residents.

The SPD V3.0 identifies usual residents using activity-based rules, meaning that an individual is only included in the population dataset if there is evidence of them interacting with an administrative system within the 12 months prior to the reference date or if they appear in the same address and have a relationship with an active person. The SPD V3.0 was designed with the explicit objective of reducing the population overcount apparent in SPD V2.0, but subsequently it results in higher levels of undercoverage. More information about the SPD V3.0 can be found in the [SPD V3.0 methodology report](#). To create the ABHE V3.0, the SPD V3.0 successfully assigns a UPRN to 98.3% of usual residents directly from the PDS data.

Our ability to accurately identify occupied addresses depends on the [quality and coverage of SPD V2.0](#) and [SPD V3.0](#) as well as the quality of the UPRN assignment of the [Office for National Statistics' \(ONS'\) Address Index Matching Service](#) to address strings in the PDS data.

Admin-based housing stock (ABHS) data

The ABHS version 1.0 (V1.0) dataset was created by linking the June 2020 Address Frame with VOA data from June 2020, and with the ABHE V2.0 and V3.0 data for 2020.

Communal establishment (CE) addresses were removed from the Address Frame. This removed 0.2% of the Address Frame, resulting in approximately 26.0 million addresses.

VOA data were then linked to the Address Frame via UPRN, with 99.2% of addresses in the Address Frame linking to a VOA address. Only 0.8% of VOA addresses failed to link to an address in the Address Frame.

CEs and special population group (SPG) addresses were removed from the ABHE datasets. This removed 3.3% of records from the ABHE V2.0 and 3.5% from the ABHE V3.0. The ABHE V2.0 and the ABHE V3.0 were then aggregated by UPRN, meaning that we have exactly one record per address. The ABHE datasets were then both linked to the Address Frame via UPRN to flag occupied addresses. Of the 26.0 million records in the Address Frame, 87.9% linked to at least one record in the ABHE V2.0, and 86.3% linked to at least one record in the ABHE V3.0. For both the ABHE V2.0 and V3.0, only 0.9% of addresses did not link to an address in the Address Frame.

The ABHS V1.0 was then linked to the [National Statistics UPRN Lookup \(NSUL\)](#) to obtain additional geography variables. A small number of ABHS records that could not be linked to the NSUL was removed.

9 . Related links

[Admin-based housing stock profile for England and Wales: 2020](#)

Article | Released 5 December 2022

Summary statistics demonstrating the feasibility of producing census-like housing stock statistics using the admin-based housing stock 2020 V1.0 dataset.

[Admin-based accommodation type statistics for England and Wales, feasibility research: 2011](#)

Methodology | Last revised 10 March 2022

Further research demonstrating the improved potential of Valuation Office Agency (VOA) data to provide detailed information on accommodation type, and examining how VOA data compare with the 2011 Census data, and a request for feedback on the usefulness of these statistics.

[Housing Characteristics](#)

Webpage | Last updated 26 October 2022

These outputs are part of our ongoing research into the possibility of combining administrative data and surveys to produce statistics for census-type information about housing characteristics as part of an Administrative Data Census.

[Valuation Office Agency property attribute data: quality assurance of administrative data used in Census 2021](#)

Methodology | Last revised 31 July 2020

Summary of the quality assurance undertaken on administrative data for Valuation Office Agency property attribute data used in Census 2021.

[Valuation Office Agency data](#)

Webpage | Last revised 27 June 2017

Data source overviews intended to provide a high-level view of new data sources being evaluated by the Census 2021 Transformation Programme.

10 . Cite this article

Office for National Statistics (ONS), released 5 December 2022, ONS website, article, [Developing admin-based housing stock statistics for England and Wales: 2020](#)